2010

Fusion of Video and Doppler Radar for Traffic Surveillance

Arunesh Roy
Wright State University

Fusion of Video and Doppler Radar for Traffic Surveillance

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

By

ARUNESH ROY
B.S., University of Mumbai, 1998
M.S., Wright State University, 2002

2010
Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

July 7, 2010

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Arunesh Roy ENTITLED Fusion of Video and Doppler Radar for Traffic Surveillance BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Lang Hong, Ph.D.
Dissertation Director

Ramana V. Grandhi, Ph.D.
Director, Ph.D. in Engineering Progam

Jack Bantle, Ph.D.
Vice President for Research and Graduate
Studies and Interim Dean of Graduate Studies

Committee on Final Examination :

Lang Hong, Ph.D.

Fred Garber, Ph.D.

Arthur Goshtasby, Ph.D.

Michael Temple, Ph.D.

Kefu Xue, Ph.D.

ABSTRACT

Roy, Arunesh. Ph.D., Engineering Ph.D. Program Wright State University, 2010. Fusion of Video and Doppler Radar for Traffic Surveillance

Current Continuous Wave (CW) Doppler radar speed measurement systems lack the ability to distinguish multiple targets. Most systems can only identify the strongest (closest) target or the fastest target.

This dissertation is related to a fusion algorithm for a VIdeo-Doppler-radAR (Vidar) traffic surveillance system. The Vidar systems uses a robust matching algorithm which iteratively matches the information from a video camera and multiple Doppler radars corresponding to the same moving vehicle, and a stochastic algorithm which fuses the matched information from the video camera and Doppler radars to derive the vehicle velocity and angle information.

We use two heterogeneous sensors of very different modalities, the first a high resolution (1024x768 pixels) video camera operating at 30 Hz with a 1/3" sony CCD fitted with a narrow field-of-view ($\sim 17^\circ$) lens and the other a CW Doppler radar operating in the unlicensed Ka band (35 GHz) with a maximum detection range of 3000 ft. First, a high resolution Time-Frequency representation of the radar signal is obtained by employing the method of Time-Frequency reassignment. Then, the angle information obtained from the video camera is fused with the information from the Doppler radar to produce a velocity and angle track of the targets within the surveillance region.

Contents

List of Figures

ACKNOWLEDGEMENTS

This dissertation would not have been possible without my dissertation advisor, Prof. Lang Hong. I must acknowledge my colleagues in the sensor fusion lab, Mark Arlinghaus and Nicholas Gale, for their assistance on various parts of this work.

I am grateful to the faculty in the Electrical Engineering department for their support and advice. I would like to make a special reference to Prof. Kuldip Rattan, Prof. Scott Thomas and Prof. Prabhakar Mateti for all the encouragement I have received from them over the years. I would also like to acknowledge the staff of the Electrical Engineering department at Wright State, Vickie Slone, Marie Donahue and Barry Woods for their support.

I owe my deepest gratitude to my parents, Kamal and Debu, and family, Monika, Praveen, Uma and Rishi, for all their love and support. I must also acknowledge all my friends for their unwavering faith, friendship and unconditional love.

1

Preliminaries

1.1 Introduction

Most roadways are designed and built for safe operation at the posted speed limits. Exceeding those posted speed limits especially during adverse weather conditions cause many of the traffic accidents and fatalities that occur each year [1]. Implementing speed enforcement techniques in target areas can be a very effective way to reduce serious incidents on these roadways.

Traffic management on roadways is a challenging task which is increasingly being augmented with automated systems. Statistical data of the traffic flow, mean traffic speed and congestion are important metrics for system administrators which can be used to reduce congestion and ensure smooth flow of traffic in large metropolitan areas. Modern traffic information systems use a variety of sensors such as inductive loops, infrared cameras, radars, lasers, magnetometers, etc. to provide real-time information on the state of the traffic on a particular roadway.

Video cameras along with computer vision algorithms are used for surveillance of activities, in hostile (enemy) environments, disaster relief and monitoring (forest fires, floods, etc), traffic monitoring, etc to provide information regarding the scene. Automating some of these vision related tasks reduces operator workload and error.

Doppler radars are a mature technology for speed enforcement systems. Solutions using Doppler radars are typically employed in down-the-road or across-the-road configurations [2].

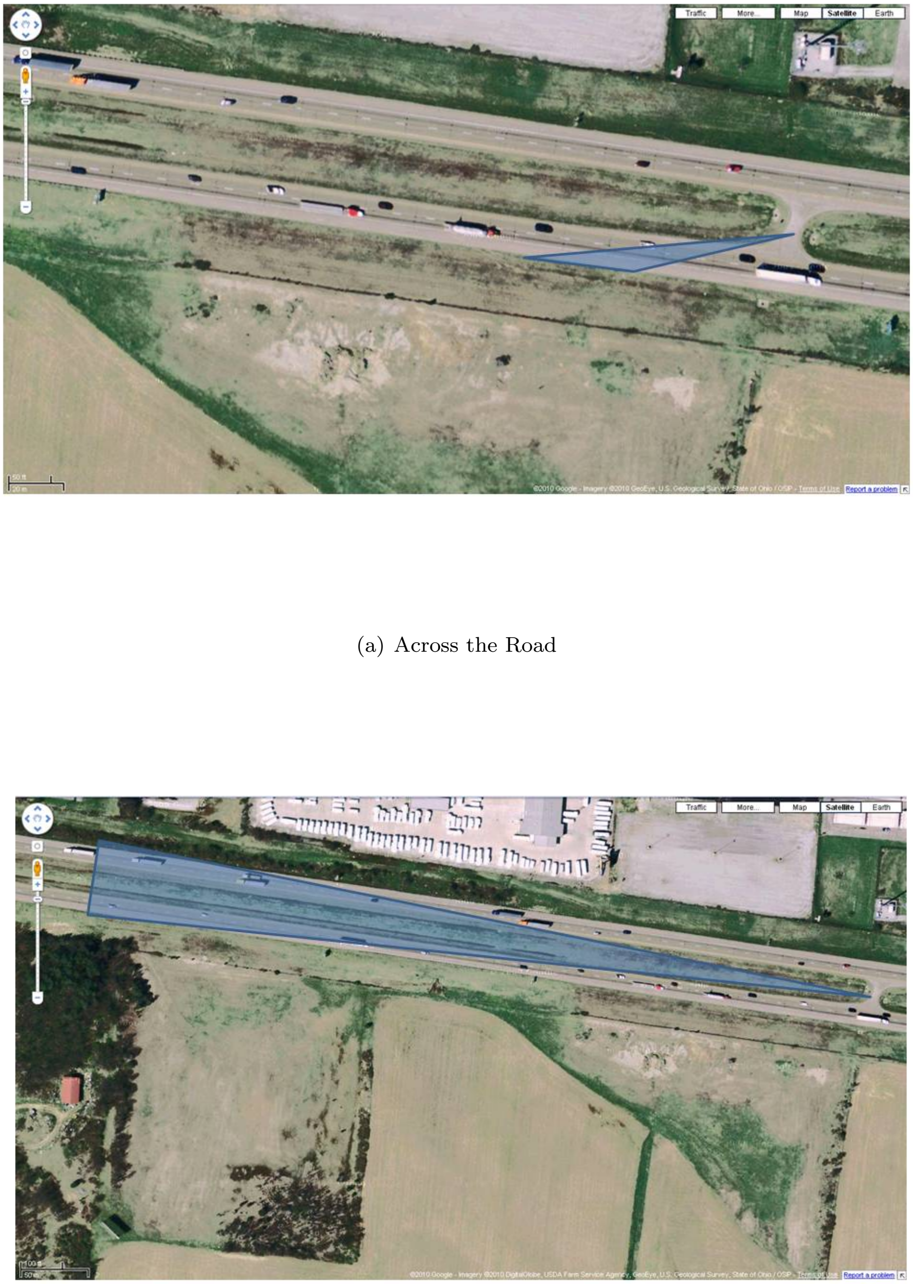
In the down-the-road scenario (Figure 1.1(a)), the axis of the Radar antenna is directed along the line of travel of traffic on the road. Since the microwave beam is almost parallel to the targets, the speed measured by the device is very close to the true target speed and the cosine effect is negligible. However, because of the large beamwidths ($5° - 15°$) and long range of the devices, identifying multiple targets is very difficult.

In the across-the-road configuration (Figure 1.1(b)), the antenna beam is focussed across the road thereby reducing the operational area and range of the coverage. This reduces the probability of multiple targets being illuminated by the microwave beam, but adds the complexity of the cosine effect into the measurement. The cosine effect can, however, be accounted for when the angle between the microwave beam and the direction of motion is known.

A traditional radar based traffic surveillance system uses a Doppler radar for vehicle speed monitoring which measures a vehicle speed along the line of sight (LOS). Although a Doppler radar based system has an advantage of a long detection range, there are several difficulties associated with the traditional radar based system, including :

- the Doppler radar beam angle is too large to precisely locate vehicles within the radar beam
- the angle between the vehicle's motion and the LOS, is unknown and therefore, needs to be small enough for a reasonable speed estimation accuracy
- since the Doppler radar does not report any angle information, vehicles with different speeds and angles can possibly record the same Doppler velocity

Most of these radar systems can track only one target at a time. Some models have the option to track and display two targets – the strongest (may be closest or largest) target (echo) and the fastest (or a faster) target in the beam. In either case, no precise target location information can be derived in a traditional Doppler Radar.


Traffic
More...
Map
Satellite
Earth
50 ft
20 m
©2010 Google - Imagery ©2010 GeoEye, U.S. Geological Survey, State of Ohio / OSIP - Terms of Use
Report a problem
Traffic
More...
Map
Satellite
Earth
100 ft
50 m
©2010 Google - Imagery ©2010 DigitalGlobe, USDA Farm Service Agency, GeoEye, U.S. Geological Survey, State of Ohio / OSIP - Terms of Use
Report a problem

(a) Across the Road

(b) Down the Road

Figure 1.1: Configurations for Doppler Radar Traffic Surveillance

In this dissertation, we will address this limitation and propose some methods to overcome this shortfall using a fusion scheme for data from Doppler radars with a video camera.

1.2 Problem Formulation

Our goal is to obtain accurate estimates of target velocities, as well as, target azimuth angles by the fusion of Doppler radar and video information. By processing the information from the two heterogeneous sensors, we will show a method to extract velocity and angle estimates of multiple targets. We will also show that the sensor fusion techniques gives us target association and improves the accuracy of the estimated quantities.

The proposed system makes some assumptions regarding the phenomenon being observed.

- The vehicles move with finite speed that has legal limits.
- The vehicle speeds are also bounded by physical limitations of current automobile technology.
- The vehicle movement is smooth and constant within the sampling interval
- The vehicle motion is constrained to the road plane.
- The road structure being observed is a relatively straight segment of road
- The sensor has a clear view of the vehicles at all times.
- The vehicles can be modeled as point targets for the Doppler radar.

The sensor package comprises of a Ka-Band Continuous Wave (CW) Doppler radar sensor which illuminates the region of interest and measures the relative velocity between the target and the sensor platform (range-rate). Simultaneously, a calibrated video camera, mounted with its its principal axis aligned with boresight of the Doppler radar, captures the video information from the scene. The video camera provides rich information regarding position, velocity and direction of the vehicles within its field of view. However, the information captured by a video camera is a 2D perspective

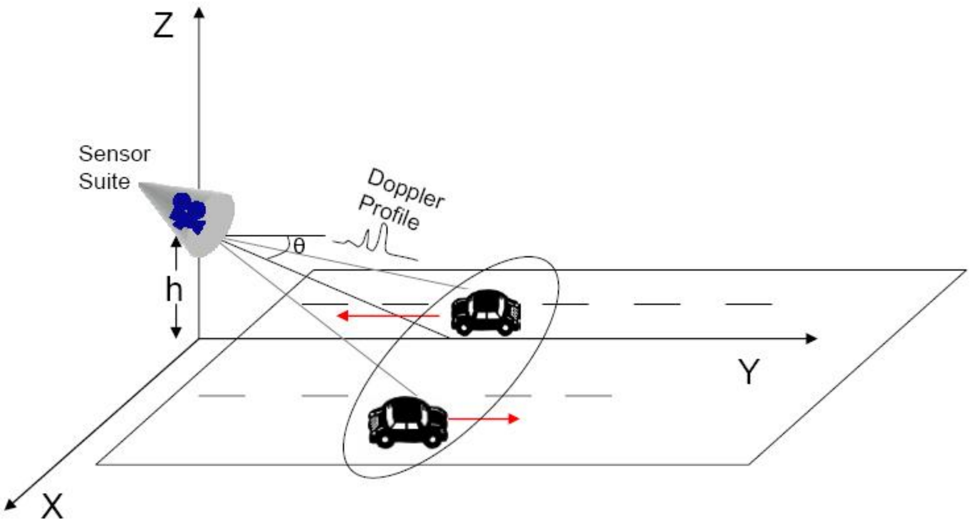


Figure 1.2: Automated Speed Enforcement Scenario

projection of the 3D scene. We have two different sensor suites that were designed. They are both included here for the sake of completeness and to show the evolution of the device.

1.2.1 Version 1.0

The sensor suite version 1.0 consists of one Ka band radar and one video camera. Figure 1.2 shows the proposed scenario. A Sensor package is rigidly mounted on a platform. The sensor platform is situated at height h above the region of interest and continuously monitors the region of interest.

Figure 1.3 shows the overview of the proposed system. The high resolution (1024x768 pixel) video camera captures the scene at video frame rate (30 Hz). Simultaneously, data is collected by the Doppler Radar at 96 KHz and is processed to extract the velocities of all the vehicles within the microwave beamwidth of the radar. The velocity histories are maintained over time and seperated from background clutter.

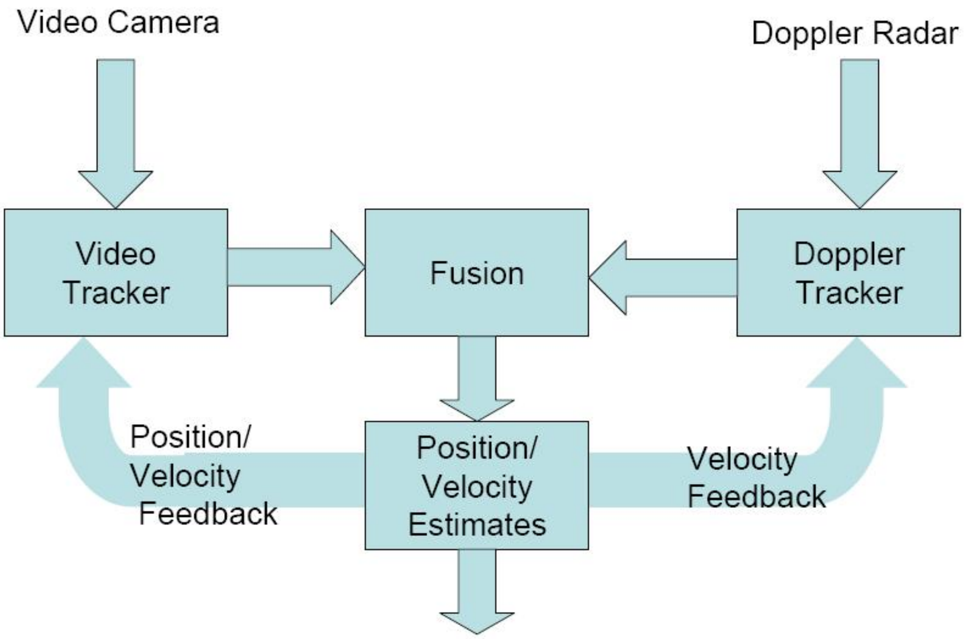


Figure 1.3: Overview of Fusion Scheme

1.2.2 Version 2.0

The sensor suite version 2.0 consists of three Ka band radars and one video camera. Of the three Radars, two have motion induced on them to create a pseudo-modulation scheme (discussed in further detail in the following chapters). The data from the Doppler radars is collected at sample rate of 150 KHz while the video camera runs at 30 Hz.

1.3 Literature Survey

Speed enforcement systems are commonly used by highway patrol but they are usually manually operated. Automated speed enforcement systems which are mounted on poles alongside roads are also commercially available. A survey of commercially available systems is given in [3]. Their range of detection is however limited to about 100m. For systems with longer ranges, the ability to discriminate between multiple targets does not exist and requires human intervention.

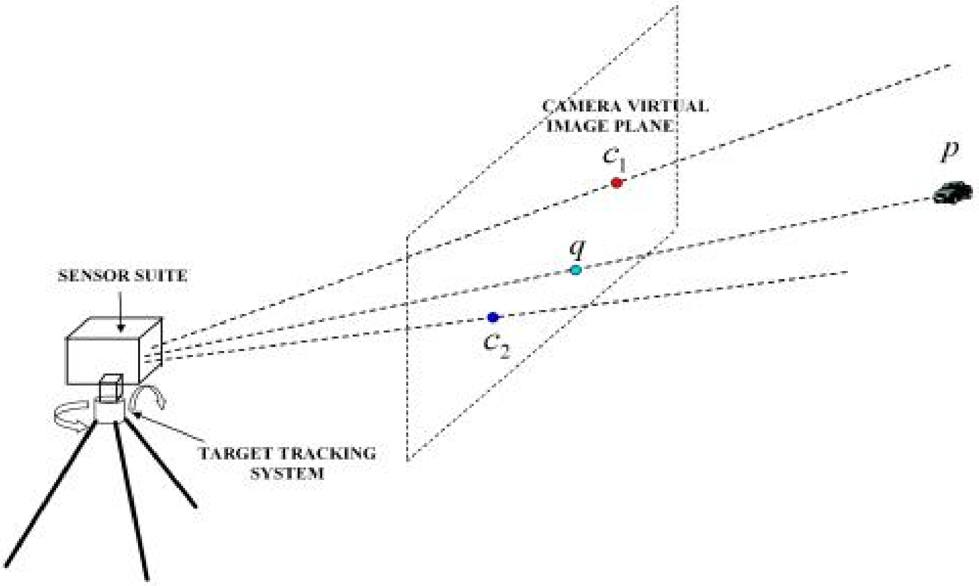


Figure 1.4: Vidar Device

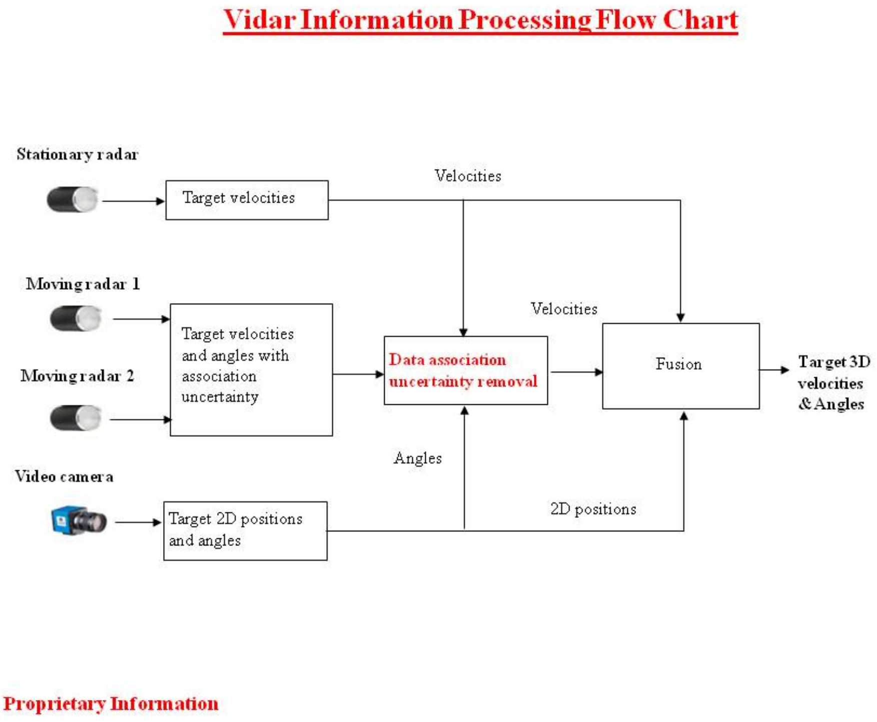

Vidar Information Processing Flow Chart
Stationary radar
Target velocities
Velocities
Moving radar 1
Velocities
Target velocities and angles with association uncertainty
Data association uncertainty removal
Fusion
Target 3D velocities & Angles
Moving radar 2
Angles
Video camera
2D positions
Target 2D positions and angles
Proprietary Information

Figure 1.5: Vidar Processing Flowchart

Tracking moving objects in video has been the focus of research in computer vision for a long time and finds applications in automatic segmentation of vehicles from background as well as other vehicles [4]. An important aspect of this task is the identification of unique tracks (multi-target tracking) and data association. The various techniques used to track vehicles in video sequences can be categorized into the following classes:

- Model-based Tracking : Assumes that a detailed model of the vehicle being tracked is available and can be matched to the model extracted from video.

- Blob Tracking : Using background subtraction algorithms, a background model is created for the scene and an incoming frame is compared to the background to determine moving pixels. The groups of moving pixels (connected components) are then processed to extract foreground blobs, which are then tracked.

- Contour Tracking : Similar to blob tracking except regions are represented by the bounding contour of the object and a Kalman filter is used to update the affine parameters of the contour itself.

- Feature Tracking : In this approach, feature points on moving objects are tracked and grouped into clusters which represent individual moving objects.

In [5], the mean traffic speed is estimated from traffic surveillance cameras. The camera calibration parameters are estimated from geometric primitives present in the image. The placement and pose of the camera is very important in these applications as noted by [6]. Most standard techniques which assume that the camera can be mounted with an almost nadir viewpoint, fail when dealing with an oblique viewpoint because of the perspective projective distortion induced. However, all of these systems are designed for traffic surveillance.

Several systems that fuse information from Doppler Radars and video cameras have been proposed in the automobile industry for collision sensing, adaptive cruise control and lane depature warning systems [7], [8], [9], [10], [11], [12], [13]. However, these systems use a frequency modulated CW Doppler radar to detect the range of the targets and fuse it with video to get better situational awareness. To the best of our knowledge, there has not been any attempt to fuse Doppler radar information with video for traffic surveillance and monitoring.

2

Background

In the following sections we will go over some of the fundamental algorithms that are used extensively in our work. We also review the sensors we propose to use and the techniques for extracting information from them.

2.1 Kalman Filter

A Kalman filter is an optimal recursive algorithm that estimates the state of a dynamic system from noisy measurements while minimizing the error in the least squares sense. It was introduced by R.E. Kalman in 1960 [14]. The Kalman filter assumes that the true state is corrupted by additive Gaussian white noise. Unlike batch processing techniques, which require the entire history of observations to be available at all times, the Kalman filter needs only the state at the previous time step and the current observation to compute the state of the system.

2.1.1 Linear Kalman Filter

If the phenomenon is modelled by a linear gaussian state space model, the Kalman filter recursion provides an exact analytical expression to compute the posterior distribution.

In other words, when the system model is given as a linear stochastic difference equation,

$$x_k = F_k x_{k-1} + w_{k-1} \qquad w_{k-1} \approx \mathcal{N}(0, Q_{k-1}) \tag{2.1}$$

$$y_k = H_k x_k + v_k \qquad v_k \approx \mathcal{N}(0, R_k)$$

where w_{k-1} and v_k are zero mean, independent, gaussian random variables with covariance matrices Q_{k-1} and R_k repectively, then the estimation solution is given by

$$p(x_k|x_{k-1}) = N\left(\hat{x}_{k-1|k-1}, P_{k-1|k-1}\right) \tag{2.2}$$

$$p(x_k|y_{k-1}) = N\left(\hat{x}_{k|k-1}, P_{k|k-1}\right) \tag{2.3}$$

$$p(x_k|y_k) = N\left(\hat{x}_{k|k}, P_{k|k}\right) \tag{2.4}$$

where the estimate and estimated covariance at each time is given using the prediction and update processes

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} \tag{2.5}$$

$$P_{k|k-1} = Q_{k-1} + F_k P_{k-1|k-1} F_k^T \tag{2.6}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(y_k - H_k \hat{x}_{k|k-1}\right) \tag{2.7}$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \tag{2.8}$$

where $N(\hat{x}, P)$ is a Gaussian distribution with mean $\hat{x}$ and covariance P and

$$S_k = H_k P_{k|k-1} H_k^T + R_k \tag{2.9}$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \tag{2.10}$$

Figure 2.1 shows the steps involved in estimating a linear process with an Kalman filter.

```
1. Initialization : set $x_0 = 0$ and $P_0 = I$
2. Prediction:
    • $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1}$
    • $P_{k|k-1} = Q_{k-1} + F_k P_{k-1|k-1} F_k^T$
3. Correction:
    • $S_k = H_k P_{k|k-1} H_k^T + R_k$
    • $K_k = P_{k|k-1} H_k^T S_k^{-1}$
    • $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left( y_k - H_k \hat{x}_{k|k-1} \right)$
    • $P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1}$
4. Set k=k+1 and return to Prediction step
```

Figure 2.1: Kalman filter pseudo-code

2.1.2 Extended Kalman Filters

For many problems, the assumptions of linearity and Gaussian noise do not hold true. When the model is non-linear, approximation schemes such as extended Kalman filters and unscented Kalman filters are used.

In an extended Kalman filter, the system is approximated using its Taylor series representation. The technique is effective when the non-linearities are relatively small. When large non-linearities are present or large model errors occur, the estimation error is prohibitively large and can even lead to divergence of the filter.

Consider

$$x_k = f(x_{k-1}) + w_{k-1} \tag{2.11}$$

$$y_k = h(x_k) + v_k$$

where w_{k-1} and v_k are zero mean, independent, gaussian random variables with covariance matrices Q_{k-1} and R_k repectively, then the estimation solution is shown in Figure 2.2.

A nice introduction to Kalman filters for estimation is provided in [15]. The subject is tacked in great detail in [16].

2.2 Tracking in Clutter

When the measurements acquired by a sensor are corrupted by noise and clutter, the accuracy of the estimator depends highly on the data association problem. There are many excellent resources on this topic from well established authors. See [17], [18], [19], [20], [21],[22].

Assume the measurements are acquired from a cluttered environment with probability of detection P_d and the gating probability is P_G, and that the true measurements at time $k+1$ conditioned on Z^k are normally distributed,

```
1. Prediction:
    • $\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1})$
    • $P_{k|k-1} = Q_{k-1} + J_x P_{k-1|k-1} J_x^T$
2. Correction:
    • $S_k = J_y P_{k|k-1} J_y^T + R_k$
    • $K_k = P_{k|k-1} J_y^T S_k^{-1}$
    • $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(y_k - h(\hat{x}_{k|k-1})\right)$
    • $P_{k|k} = P_{k|k-1} - K_k J_y P_{k|k-1}$
3. Set k=k+1 and return to Prediction step

note : $J_x$ and $J_y$ are the jacobians of $f(.)$ and $h(.)$ denoted by
    $J_x = \nabla f|_{\hat{x}_{k-1|k-1}}$
    $J_y = \nabla h|_{\hat{x}_{k-1|k-1}}$
```

Figure 2.2: Extended Kalman Filter Pseudo-code

$$p[z_{k+1}|Z^k] = \mathcal{N}[z_{k+1}; \hat{z}_{k+1|k}, S_{k+1}] \tag{2.12}$$

Then, we can define a region in the measurement space where the measurement will be found with high probability.

$$\tilde{V}_{k+1}(\gamma) = \left\{z : [z - \hat{z}_{k+1|k}]^T S_{k+1}^{-1} [z - \hat{z}_{k+1}] \leq \gamma\right\} \tag{2.13}$$

$$= \left\{\nu'_{k+1} S_{k+1}^{-1} \nu_{k+1} \leq \gamma\right\} \tag{2.14}$$

$$\tag{2.15}$$

where S_{k+1} is the covariance matrix of the innovation.

The set of validated measurements at time k is denoted as

$$Z(k) = \{z_i(k)\}_{i=1}^{m_k} \tag{2.16}$$

2.2.1 Nearest Neighbor Standard Filter

In the Nearest Neighbor Standard Filter (NNSF), the validated measurement closest to the predicted measurement is used to update the state of the filter. The distance measure to be minimized is the weighted norm of the innovation,

$$d^2(z) = [z - \hat{z}_{k+1|k}]^T S_{k+1}^{-1} [z - \hat{z}_{k+1}] \tag{2.17}$$

$$= \nu'_{k+1} S_{k+1}^{-1} \nu_{k+1} \tag{2.18}$$

The problem with choosing the nearest neighbor is that it has a small probability of not being the correct measurement. Therefore, the NNSF will sometimes lose track by using the wrong measurement to update the state of the filter.

2.2.2 Track Splitting Filter

The Track Splitting Filter splits the track into separate hypothesis at each time step, one for every measurement that falls within the validation gate. Then for each measurement, a separate updated state estimate and estimated covariance are computed using the Kalman filter equations. As the number of tracks into which the filter splits increase, the number of hypotheses grow exponentially. In order to reduce the computational requirement, the likelihood function of each split track is computed and unlikely tracks are discarded.

Denote the l^{th} sequence of measurements upto time k as

$$Z^{k,l} = \{z_{i_{1,l}}(1), z_{i_{2,l}}(2), \ldots, z_{i_{k,l}}(k)\}$$

where $z_i(j)$ is the i^{th} measurement at time j. The likelihood function of this sequence being a track originating from the same target

$$\Lambda(\Theta^{k,l}) = p[Z^{k,l}|\Theta^{k,l}] = p[z_{i_{1,l}}(1), z_{i_{2,l}}(2), \ldots, z_{i_{k,l}}(k)|\Theta^{k,l}] \tag{2.19}$$

Denoting by Z^k the complete set of measurements through time k, we can write

$$\Lambda(\Theta^{k,l}) = \prod_{j=1}^{k} p[z_{i_{j,l}}|Z^{j-1}, \Theta^{k,l}] \tag{2.20}$$

Then, under Gaussian assumptions,

$$p[(z(j)|Z^{j-1}\Theta^{k,l}] = N[z(j); \hat{z}(j|j-1), S(j)] = N(\nu(j); 0, S(j)] \tag{2.21}$$

$$\Lambda(\Theta^{k,l}) = c_k exp\left[-\frac{1}{2}\sum_{j=1}^{k} \nu'(j)S^{-1}(j)\nu(j)\right] \tag{2.22}$$

where

$$c_k = \prod_{i=1}^{k} |2\pi S(i)|^{-1/2}$$

The modified log-likelihood function can be written as

$$\lambda(k) = -2\log\left[\frac{\Lambda(\Theta^{k,l})}{c_k}\right] = \sum_{j=1}^{k} \nu'(j)S^{-1}(j)\nu(j) \tag{2.23}$$

which can be computed recursively as

$$\lambda(k) = \lambda(k-1) + \nu'(k)S^{-1}(k)\nu(k) \tag{2.24}$$

The log-likelihood function at time k has a chi-square density with kn_z degrees of freedom. The statistical test for accepting a track is that the log-likelihood function satisfies

$$\lambda(k) \leq a \tag{2.25}$$

where the threshold a is obtained from the chi-square table for kn_z degrees of freedom

$$P\{\chi^2_{kn_z} \geq a\} = \alpha \tag{2.26}$$

2.2.3 Probabilistic Data Association Filter

The Probabilistic Data Association Filter (PDAF) is a sub-optimal Bayesian algorithm. At each time step, a validation region is created and one of the measurements from the set inside the validation region is assumed to be from the target and the rest from cluttter modeled as random variables with uniform distribution.

The entire history of the PDAF is approximated as

$$p[x(k)|Z^{k-1}] = \mathcal{N}[x(k); \hat{x}(k|k-1), P(k|k-1)] \tag{2.27}$$

i.e, the state is assumed to be normally distributed (Gaussian) with mean the latest estimate and covariance given by the covariance of the estimate.

The set of validated measurement $Z(k)$ at time k are the measurements contained inside the elliptical validation region $\tilde{V}_k(\gamma)$.

Define the events

$$\theta_i(k) = \{z_i(k) \text{ is the target-orginated measurement}\}, i = 1, \ldots, m_k \tag{2.28}$$

$$\theta_0(k) = \{\text{none of the measurements at time k is target-originated}\} \tag{2.29}$$

with probabilities conditioned on Z^k given by

$$\beta_i(k) = P\{\theta_i(k)|Z^k\}, i = 1, \ldots, m_k \tag{2.30}$$

Since these events are mutually exclusive and exhaustive,

$$\sum_{i=0}^{m_k} \beta_i(k) = 1 \tag{2.31}$$

Using the total probability theorem with respect to the above events, the conditional mean of the state at time k can be written as

$$\hat{x}_{k|k} = E[x_k|Z^k] = \sum_{i=0}^{m_k} E[x_k|\theta_i(k), Z^k]P\{\theta_i(k)|Z^k\} \tag{2.32}$$

$$= \sum_{i=0}^{m_k} \hat{x}_{i_{k|k}} \beta_i(k) \tag{2.33}$$

where $x_{i_{k|k}}$ is the updated state estimate conditioned on the event $\theta_i(k)$ that the i^{th} validated meaasurement is correct. Then,

$$\hat{x}_{i_{k|k}} = \hat{x}_{k|k-1} + W(k)\nu_i(k) \tag{2.34}$$

where

$$\nu_i(k) = z_{i_k} - \hat{z}_{k|k-1}$$

For $i = 0$ i.e. when none of the measurements are correct, the estimate is

$$\hat{x}_{0_{k|k}} = \hat{x}_{k|k-1} \tag{2.35}$$

Then,

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + W(k)\nu(k) \tag{2.36}$$

where $\nu(k)$ known as the combined innovation is given as

$$\nu(k) = \sum_{i=1}^{m_k} \beta_i(k)\nu_i(k) \tag{2.37}$$

The error covariance associated with the updated state estimate is

$$P_{k|k} = \beta_0(k)P_{k|k-1} + [1-\beta_0]P^c_{k|k} + \tilde{P}_k \tag{2.38}$$

where

$$\tilde{P}_k = W(k)\left[\sum_{i=1}^{m_k} \beta_i(k)\nu_i(k)\nu_i'(k) - \nu(k)\nu'(k)\right]W'(k) \tag{2.39}$$

and

$$P^c_{k|k} = [I - W(k)H(k)]P_{k|k-1} \tag{2.40}$$

The evaluation of the association probabilities is done as follow :

$$\beta_i(k) = P\{\theta_i(k)|Z^k\} \tag{2.41}$$

$$= P\{\theta_i(k)|z_k, m_k, Z^{k-1}\} \tag{2.42}$$

$$= \frac{1}{c} p\left[z_k|\theta_i(k), m_k, Z^{k-1}\right] P\{\theta_i(k)|m_k, z^{k-1}\} \text{ (using Bayes' Rule)} \tag{2.43}$$

2.3 Robust Estimation

Under certain circumstances, the traditional assumption that data can be classified as targets and clutter fails. When number of data points in the data set is large, it becomes very hard to initialize and track the targets. In such cases, more robust methods are required for target tracking.

RANSAC (RANdom SAmple Consensus) [23] is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. The data is assumed

Objective : Robust fit of a model to a data set $\mathcal{S}$ which contains outliers

Algorithm

1. Randomly select a sample of s data points and estimate the model from this subset

2. Determine the set of data points S_i, which are within a distance threshold t.

   The set S_i is the consensus set of the sample and define the inliers of S.

3. If the size of S_i is greater than some threshold T, re-estimate the model using all the points in S_i and terminate.

4. If the size of S_i is less than T, select a new subset and go to 1.

5. After N trials, the largest consensus set S_i is selected and the model is re-estimated using all the points in the subset S_i.

Figure 2.3: RANSAC for Outlier Rejection

to consist of a set of inliers that can be defined by a parametric model and outliers that do not fit this model. In addition to this, the data can be subject to noise. RANSAC also assumes that there exists a procedure which can estimate the parameters of the model that optimally explains or fits this data.

RANSAC is composed of two main steps that are repeated in an iterative fashion.

- Hypothesize : First minimal sample sets are randomly selected from the input data set and the model parameters are computed using only the elements on the minimal sample set.

- Test : the entire dataset is checked for consistency with the model instantiated in the first step. The set of inliers to the model is called the consensus set.

Figure 2.3 shows the steps in a RANSAC algorithm for robustly estimating a model to fit data.

2.3.1 RANSAC for line-fitting

Consider the problem of fitting a straight line to a set of 2-D points. Assuming that the data set consists of both inliers and outliers, a least squares method for finding this line will in general produce a line with a bad fit to the inliers. So the problem is actually two-fold, a line fit to the data and a classification of the data into inliers and outliers.

Two points are selected at random from the data set; these two points define a candidate line fit. The inliers to this model are determined as the points that lie within a distance threshold. The process of random selection is repeated a number of times and the line with the largest consensus set is deemed the most robust fit.

> The RANSAC procedure is opposite to that of conventional smoothing techniques: Rather than using as much of the data as possible to obtain an initial solution and then attempting to eliminate the invalid data points, RANSAC uses as small an initial data set as feasible and enlarges this set with consistent data when possible [23].

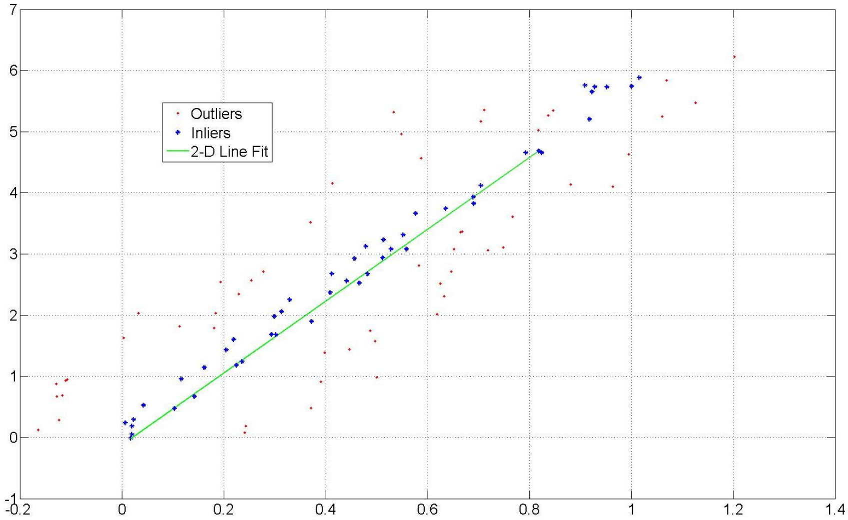
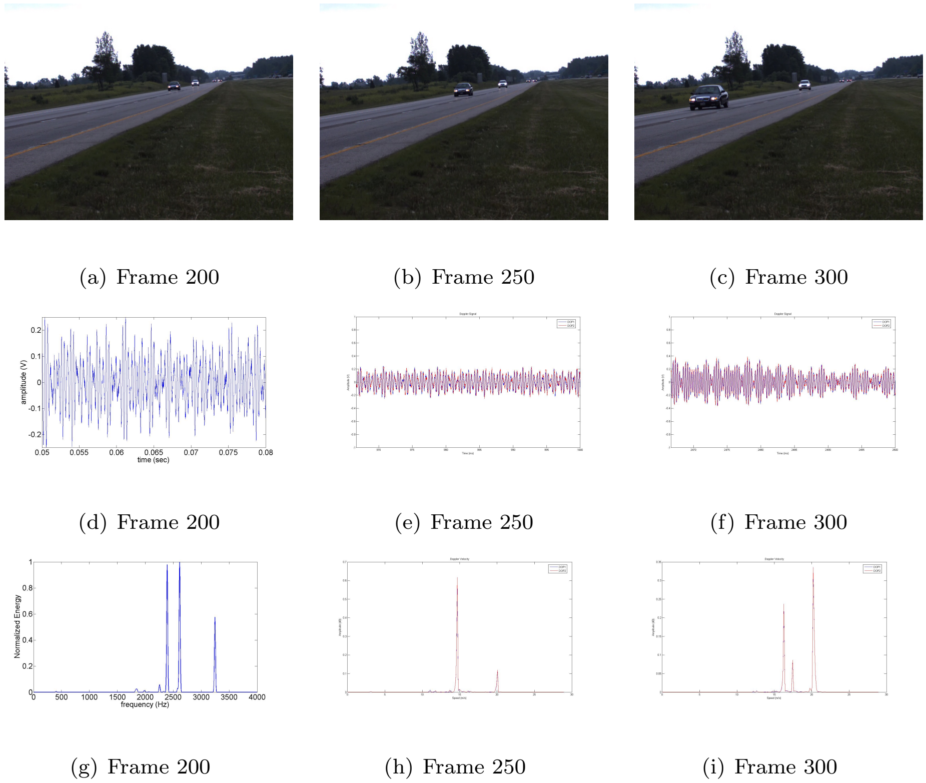


Figure 2.4: RANSAC for 2-D line fitting



(a) Frame 200 (b) Frame 250 (c) Frame 300

(d) Frame 200 (e) Frame 250 (f) Frame 300

(g) Frame 200 (h) Frame 250 (i) Frame 300

Figure 2.5: Sample Data : 2.5(a)-2.5(c) collected from a camera, 2.5(d)-2.5(f) collected from Doppler radar, 2.5(g)-2.5(i) Short-time FFT of Doppler radar data

2.4 Sensors

We use two different sensors in our work, a video camera and a CW Doppler radar. Figure 2.5 shows sample data from the two sensors which we discuss in detail in the following sections.

2.4.1 Video

The first sensor we use in our system is a high resolution (1024x768 pixels) video camera operating at 30 Hz. The camera has a 1/3” sony CCD and is fitted with a narrow field-of-view ($\sim 17°$) lens.

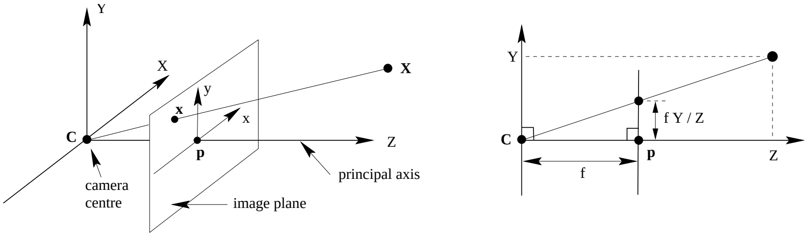


Figure 2.6: Pinhole Camera Model

2.4.1.1 Camera as a sensor

Video sensors (cameras) provide rich information about a scene and have a wide area of detecion. There are several excellent books available in computer vision [24],[25],[26],[27].We will briefly discuss some of the relevant details here.

Consider the basic pinhole model (Figure 2.6 [24]) of a camera which assumes that points in 3D are projected onto a plane at the camera. Let the center of projection be the origin of a Euclidean coordinate system with the projection plane at $Z = f$, which is called the image plane (focal plane).

In the pinhole camera model, a point $\underline{X} = (X, Y, Z)^T$ in the world coordinate system is mapped to a point $\underline{x} = (x, y)^T$ on the image plane where the line joining the center of projection to $\underline{X}$ meets the image plane.

If $\underline{X}$ is the world point represented in homogenous vector format ie. $\underline{X} = (X, Y, Z, 1)$ and $\underline{x}$ is the corresponding image point in homogenous format ie $\underline{x} = (x, y, 1)$, then

$$\underline{x} = P\underline{X} \tag{2.44}$$

where P is a 3 × 4 homogeneous matrix called the camera projection matrix. The generalized form of the perspective projection matrix is

$$P = K[R|t] \tag{2.45}$$

where K is a 3 × 3 upper triangular matrix called the camera calibration matrix,

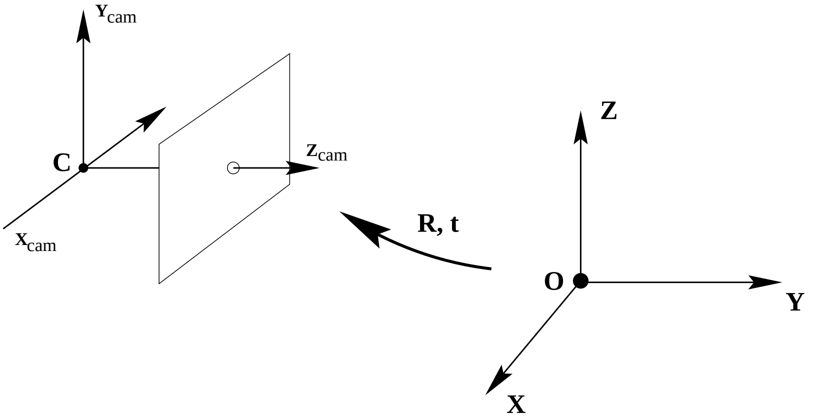


Figure 2.7: World Coordinates and Camera Coordinates are related via an Euclidean Transformation

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ & \alpha_y & p_y \\ & & 1 \end{bmatrix} \tag{2.46}$$

where

- α_x and α_y are the scaling parameters in the image x and y directions
- (p_x, p_y) is the principal point
- s is the skew (the angle between the image x and y axes)

The matrix K is also referred to as the camera intrinsics and can be determined using standard camera calibration algorithms [28]. The matrix R and the vector t, also called the camera extrinsics, define the relative orientation of the camera coordinate system relative to the world coordinate system (Figure 2.7[24]).

2.4.1.2 Camera as a Protractor

Any point $\underline{x}$ on the image plane lies on a ray $\underline{d}$, that passes through $\underline{x}$ and the camera center c. The calibration matrix relates the image point to the ray's direction. Any point in the world co-ordinates that lie on the ray can be defined as $\underline{\tilde{X}} = \lambda \underline{d}$ which map to the image plane as

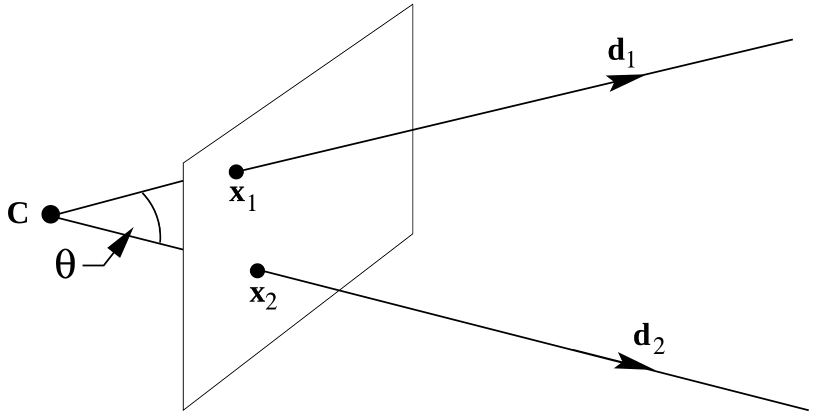


Figure 2.8: Angle between two rays

$\underline{x} = K[I|0](\lambda\underline{d}^T, 1) = K\underline{d}$. This means that the direction of any point on the image plane can be written as $\underline{d} = K^{-1}\underline{x}$.

The angle between two points $\underline{x}_1$ and $\underline{x}_2$ with directions $\underline{d}_1$ and $\underline{d}_2$ (Figure 2.8 [24]) can be obtained by the cosine formula for angle between two vectors in euclidean geometry

$$\cos\theta = \frac{\underline{d}_1^T\underline{d}_2}{\sqrt{\underline{d}_1^T\underline{d}_1}\sqrt{\underline{d}_2^T\underline{d}_2}} \tag{2.47}$$

$$= \frac{(K^{-1}\underline{x}_1)^T(K^{-1}\underline{x}_2)}{\sqrt{(K^{-1}\underline{x}_1)^T(K^{-1}\underline{x}_1)}\sqrt{(K^{-1}\underline{x}_2)^T(K^{-1}\underline{x}_2)}} \tag{2.48}$$

So if we know the calibration matrix, we can measure the direction of rays like a protractor.

2.4.1.3 Video Tracking

Identifying moving objects in a video sequence is a very important step in the automatic monitoring and analysis of traffic. In order to exploit the information in a traffic scene, we need to seperate the video into its components, the static background and the moving objects. When the background is stationary (or registered), this task can be accomplished by background subtraction, where each frame is compared to a reference (background model). Pixels in the current frame that are significantly different from the background are then labelled as foreground (moving) pixels. These foreground pixels can then be further processed for object initialization, localization and tracking.

The problems faced when modelling the background, are changes in illumination during different times of the day and weather conditions such as rain, fog or snow, shadows cast by moving objects, and the complex nature of traffic flow (especially at traffic lights or stop signs). An excellent review of background modelling as applicable to urban traffic video is presented in [29].

Once the background model has been created (Figure 2.9), pixels from the current frame are compared against pixels in the background image at the same location to check if they are significantly different from each other (Figure 2.10). A simple scheme may test the difference against a threshold. Alternately, the difference may be tested using the normalized statistics of the background image.

In our work, an open source implementation of the background subtraction and foreground object tracking provided in OpenCV [30] as part of their Video Surveillance application was modified to fit a typical traffic scenario (illumination, background, object sizes, object velocities, etc.). OpenCV is discussed in great detail in [31].

Once the pixels have been classified into foreground and background, the foreground labelled pixels are then morphologically enhanced using connected component analysis to segment regions of moving blobs (Figure 2.11). The scheme used for connected component analysis is described in [32].

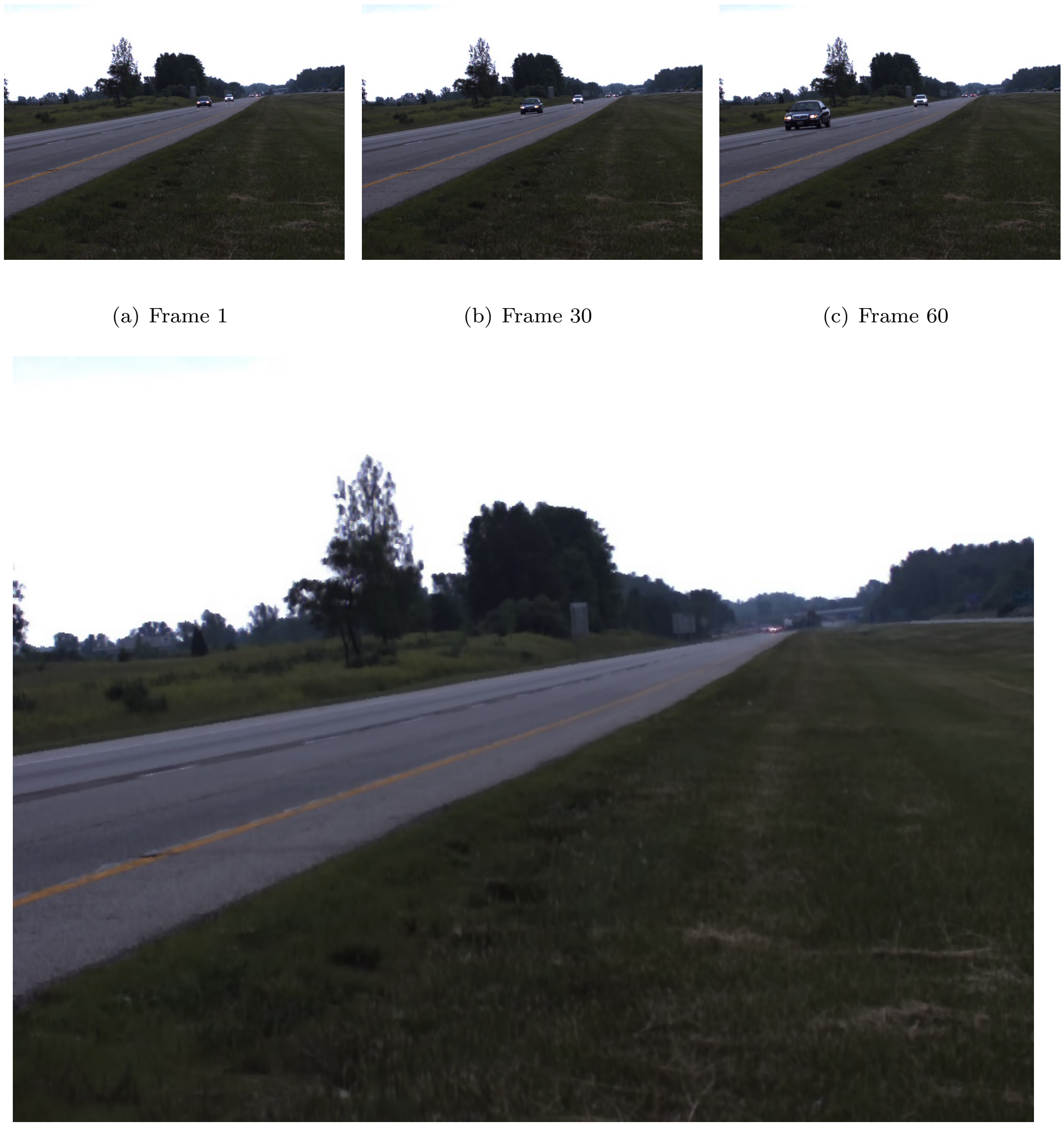
The analysis has the following components :

- identify connected components of the foreground mask as a blob
- track each blob by trying to find it in the current frame and the previous frame
- add a new blob into the tracked blobs list if it can be successfully tracked in multiple frames

Once the blobs with motion have been identified, a Kalman filter framework is used to predict the position of the blob in the next frame.

2.4.2 CW Doppler Radar

The second sensor used in our work is a CW Doppler radar operating in the unlicensed Ka band (35 GHz) with a maximum detection range of 3000 ft manufactured by Decatur Radar [33]. The



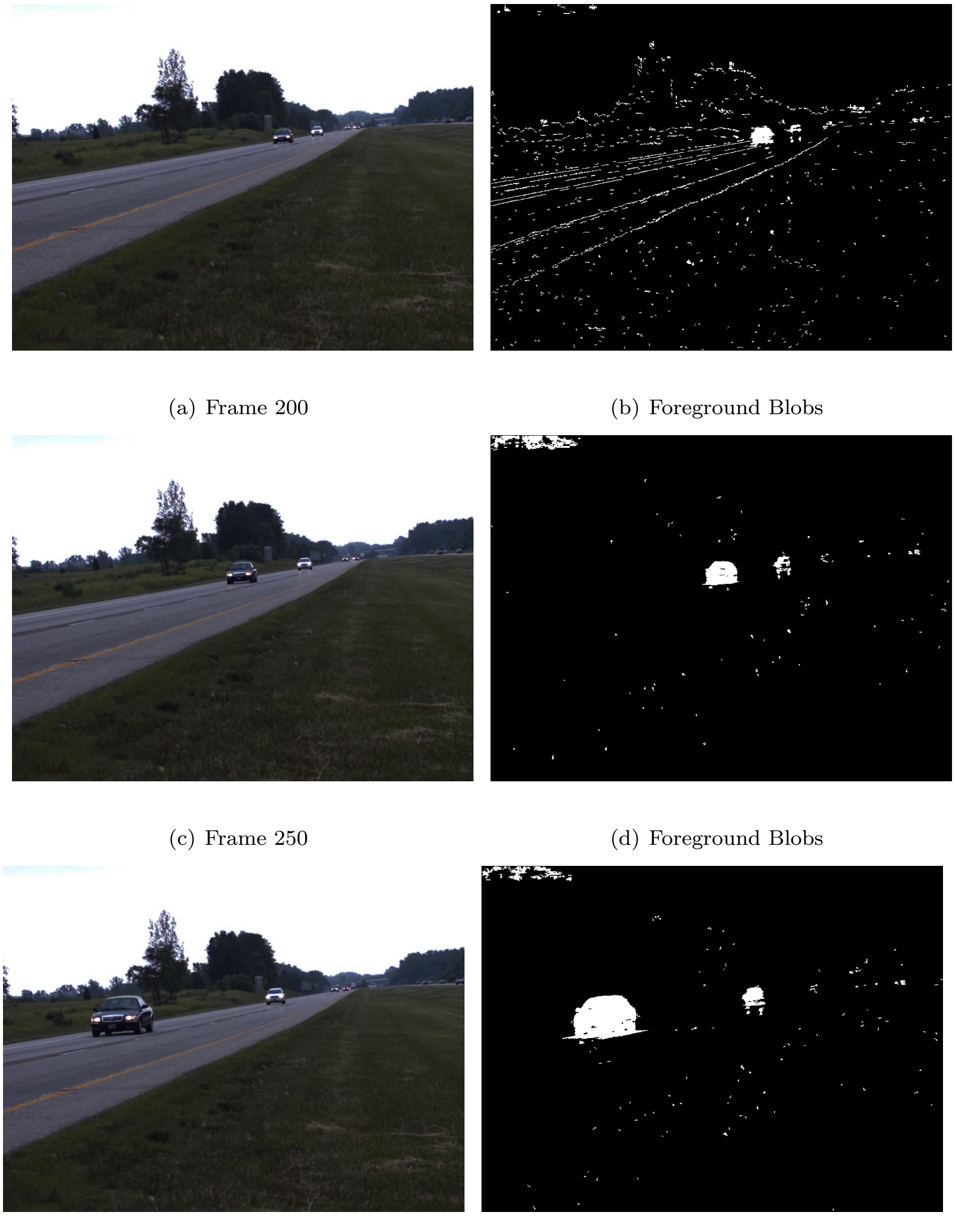
(a) Frame 1

(b) Frame 30

(c) Frame 60

(d) Background Model

Figure 2.9: Background Modeling



(a) Frame 200

(b) Foreground Blobs

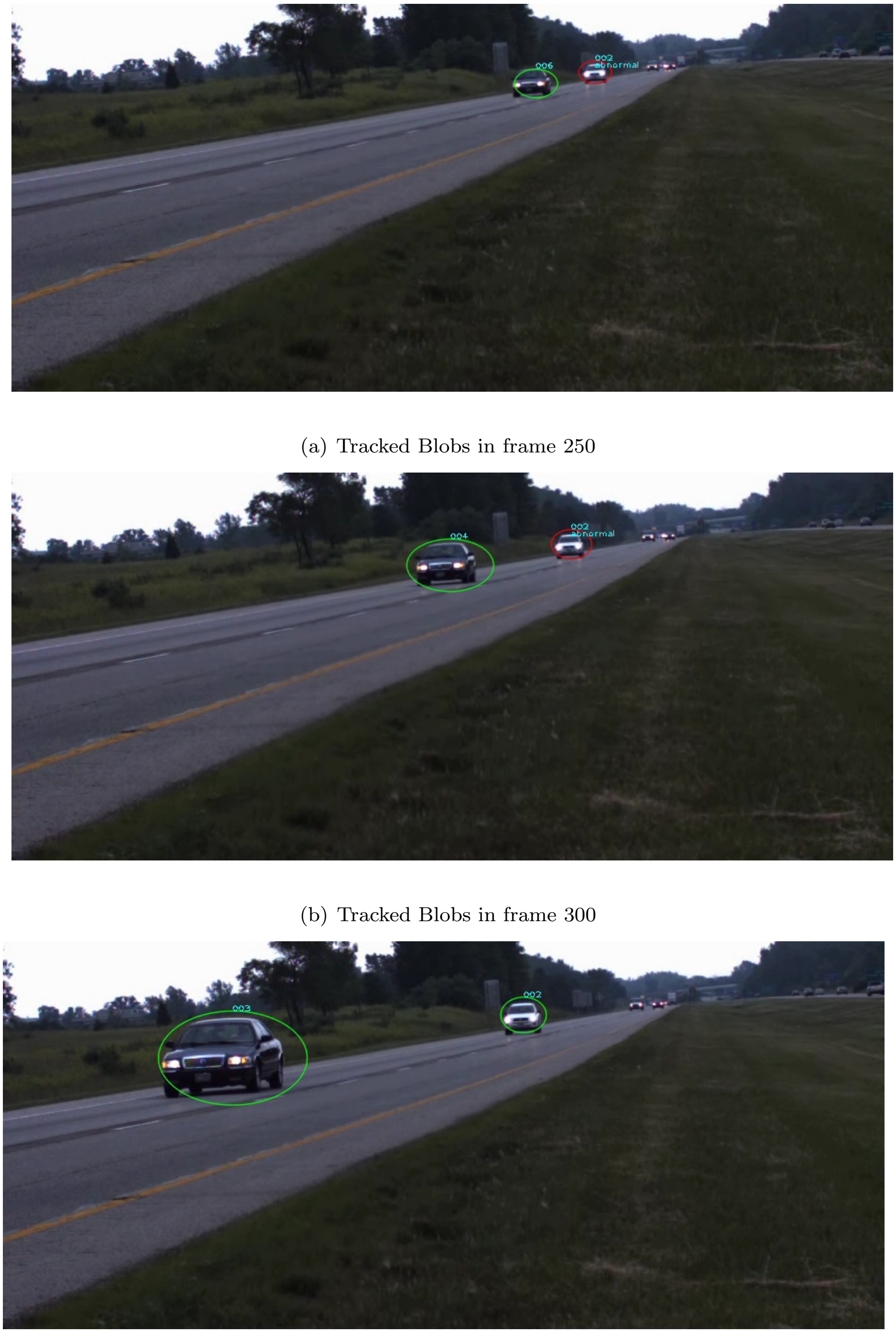
(c) Frame 250

(d) Foreground Blobs

(e) Frame 300

(f) Foreground Blobs

Figure 2.10: Foreground Segmentation



(a) Tracked Blobs in frame 250

(b) Tracked Blobs in frame 300

(c) Tracked Blobs in frame 350

Figure 2.11: Foreground Blobs Tracked

Doppler effect is the change in frequency and wavelength of a wave as perceived by an observer moving relative to the source of the wave. For electromagnetic waves, the Doppler shift in frequency (wavelength) depends only on the relative velocity between the observer and the source.

2.4.2.1 Theory of Operation

A Doppler radar emits a microwave frequency from its transmitting antenna, which is reflected off a moving target and received by the receiving antenna. The received signal has a shift in its frequency which is proportional to the target's velocity relative to the Doppler radar. The Doppler shift frequency (F_D)[34],[35]. is given by:

$$F_D = 2V\frac{F_0}{C}\cos\theta \tag{2.49}$$

where

F_0 = transmitter frequency in hertz

C = velocity of light$(3 \times 10^8\, m/s)$

V = velocity of the target(m/s)

θ = angle between microwave beam and target's path

If $\theta = 90°$ (i.e. the target is moving perpendicular to the microwave beam) there is no Doppler shift and $F_D = 0$, while if $\theta = 0$ (i.e. the target is moving parallel to the microwave beam) $F_D = 2VF_0/C$, which gives the maximum attainable Doppler shift.

Since no ranging is involved and no time-of-flight measurements are needed, the Doppler sensor is very simple, small and light weight. Figure 2.12 shows K-band and Ka-band Doppler sensors manufactured by Decatur Electronics.

Figure 2.5(d)-2.5(f) shows a short segment of the waveform received at the output of the Doppler radar. The speed information of the target is carried in the frequency of this received waveform. A simple way to determine the speed (frequency) of the target would be to perform a Fourier Transform (FFT) of the received signal Figure 2.5(g) - 2.5(i).

GENESIS II™

SELECT/DIRECTIONAL

Decatur electronics

ISO 9001 Certified

USE THE GII SELECT TO:

- Track traffic while moving or stationary
- Catch the faster target

OPTIONAL FEATURES

- K band antennas
- K directional antennas
- Ka band antennas
- Rear-view mirror display
- VIP - receive patrol speed from vehicle

www.DecaturRadar.com

Serving law enforcement for over 50 years
Radar • Trailers • Video

a division of Soncell A PUBLIC SAFETY COMPANY

[33]

Figure 2.12: CW Doppler Radar by Decatur Electronics

When multiple vehicles are within the field-of-view of the sensor, the Doppler radar receives multiple frequencies at its output with the received power corresponding to each vehicle given by (2.50).

$$\frac{P_E}{P_S} = \frac{G^2 \lambda^2 \sigma}{(4\pi)^3 D^4} \tag{2.50}$$

where

- P_E : received signal power
- P_S : transmitted signal power
- λ : wavelength of transmitted signal
- σ : radar cross section of target
- D : distance between radar and target
- G : gain of the radar antenna

It is important to note that the received signal's power is inversely proportional to the fourth power of the range to the target and directly proportional to the radar cross section of the target.

When there are two or more targets present, it is not possible to directly interpret the structure of the contributing signals from the FFT. Also, if the frequency of the signals is changing, then the FFT will not be able to detect the time-varying nature of the signal. In order to view this time-varying nature of the signal, some form of joint time-frequency processing,which can characterize the energy distribution of the signal in time and frequency, is neccessary.

2.4.2.2 Spectrogram and Reassigment

Spectrograms represent a one dimensional time-domain signal's energy as a two-dimensional function of time and frequency. The spectrogram of a signal is defined as the squared magnitude of the short-time Fourier transform

$$S(t,\omega) = \|X(t,\omega)\|^2 \tag{2.51}$$

where

$$X(t,\omega) = \int x(\tau)h(t-\tau)e^{-j\omega\tau}d\tau \tag{2.52}$$

$$= M(t,\omega)e^{j\phi(t,\omega)} \tag{2.53}$$

where $h(t)$ is a real-valued windowing function that limits the Fourier transform to the neighborhood of the current time t, $M(t,\omega)$ is the magnitude of the short-time fourier transform and $\phi(t,\omega)$ is its phase.

Instead of time-shifting the windowing function, the spectrogram can also be computed by shifting the time-varying signal,

$$X_t(\omega) = \int x(t+\tau)h(-\tau)e^{-j\omega\tau}d\tau \tag{2.54}$$

The resulting transform, called the Moving Window Transform, has the same magnitude but the phase differs by a linear term. This can be seen by substituting $t' = t+\tau$ in (2.52).

$$X_t(\omega) = \int x(t+\tau)h(-\tau)e^{-j\omega\tau}d\tau \tag{2.55}$$

$$= \int x(t')h(t-t')e^{-j\omega(t'-t)}dt' \tag{2.56}$$

$$= e^{j\omega t}\int x(t')h(t-t')e^{-j\omega t'}dt' \tag{2.57}$$

$$= e^{j\omega t}M(t,\omega)e^{j\phi(t,\omega)} \tag{2.58}$$

$$M_t(\omega) = M(t,\omega) \tag{2.59}$$

$$\phi_t(\omega) = \omega t + \phi(t,\omega) \tag{2.60}$$

For a sinusoid with frequency (ω_0), the phase of the short-time Fourier transform at $\phi(t,\omega_0)$ is constant through time. The phase of $\phi_t(\omega_0)$ however varies with frequency exactly ω_0.

In the moving window method, a time-domain signal $x(t)$ is decomposed into a set of coefficients, $\epsilon(t,\omega)$, based on a set of elementary signals $h_\omega(t)$

$$h_\omega(t) = h(t)e^{j\omega t} \tag{2.61}$$

where $h(t)$ is a kernel function (similar to the window function.

The coefficients of the decomposition are defined as

$$\begin{aligned} \epsilon(t,\omega) &= \int x(\tau)h(t-\tau)e^{-j[\omega t-\tau]}d\tau && (2.62)\\ &= e^{j\omega t}\int x(\tau)h(t-\tau)e^{-j\omega\tau}d\tau && (2.63)\\ &= e^{j\omega t}X(t,\omega) && (2.64)\\ &= X_t(\omega) && (2.65)\end{aligned}$$

For the moving window method, the time-frequency representation is constructed by the squared magnitude of the coefficients of the short-time Fourier transform i.e. the spectrogram.

The spectrogram as a time-frequency representation has relatively poor resolution. The choice of analysis window determines the time and frequency resolution and higher resolution in one domain leads to smearing in the other. Using a shorter length window leads to a higher resolution in time but because the bandwidth of the resulting spectrum is larger, the frequency resolution will be poorer.

The signal received from the Doppler radar poses additional challenges. Typically, it is assumed that all targets are point scatterers and hence the returned Doppler from the target is well localized in frequency. In practice, however this is not true. Each target has a finite Radar Cross Section (RCS) and may act as multiple scatterers. Since all scatterers are moving with the same velocity but have slightly different angles to the radar beam for a given vehicle, the received Doppler signal may be spread out over a small but finite range of frequencies.

Several key developments in the field of time frequency reassigment, first introduced by Kodera et al [36] and improved on by [37], are available to enhance the time and frequency resolution.

Reassignment can be interpreted as a two-step process, first the energy distribution is smoothed to reduce oscillatory influences and then the smoothed distribution is squeezed to focus individual contributions.

The signal $x(t)$ can be reconstructed from its moving window coefficients (Fourier coefficients) by

$$\begin{aligned} x(t) &= \int\int X_t(\omega) h^*_\omega(\tau - t) d\omega d\tau && (2.66) \\ &= \int\int X_t(\omega) h(\tau - t) e^{-j\omega[t-\tau]} d\omega d\tau && (2.67) \\ &= \int\int M_\tau(\omega) e^{j\phi_\tau(\omega)} h(\tau - t) e^{-j\omega[t-\tau]} d\omega d\tau && (2.68) \\ &= \int\int M_\tau(\omega) h(\tau - t) e^{-j[\phi_\tau(\omega)-\omega t+\omega\tau]} d\omega d\tau && (2.69) \end{aligned}$$

Under the assumption that the time variation of the signal is relatively slow compared to its phase variation, the maximum contribution to the reconstruction comes from the points in the neighborhood of t, ω such that,

$$\begin{aligned} \frac{\partial}{\partial\omega}\left[\phi_t(\omega) - \omega\tau + \omega t\right] = 0 && (2.70) \\ \frac{\partial}{\partial\tau}\left[\phi_t(\omega) - \omega\tau + \omega t\right] = 0 && (2.71) \end{aligned}$$

This is known as the principle of stationary phase. The points $\hat{t}, \hat{\omega}$ where the maximum contribution occurs are defined as,

$$\begin{aligned} \hat{t}(\tau, \omega) &= \tau - \frac{\partial\phi_\tau(\omega)}{\partial\omega} = -\frac{\partial\phi(\tau, \omega)}{\partial\omega} && (2.72) \\ \hat{\omega}(\tau, \omega) &= \frac{\partial\phi_\tau(\omega)}{\partial\tau} = \omega + \frac{\partial\phi(\tau, \omega)}{\partial\tau} && (2.73) \end{aligned}$$

In the Modified Moving Window Transform (reassigned spectrogram), the energy of the spectrogram within a window is assigned to the point of maximum contribution $\hat{t}, \hat{\omega}$ (center of gravity of the energy contributions), rather than the point t, ω.

The reassigned spectrogram can be efficiently computed using the discrete Fourier transform which is the most common method of representing time and frequency representations of signals in digital signal processing. The discrete Fourier transform $X(k)$ of a signal $x(n)$ in time domain is computed as

$$X(k) = \sum_{m=0}^{N-1} x(m)e^{\frac{-j2\pi km}{N}} \tag{2.74}$$

$X(k)$, the discrete Fourier tranform coefficients is equivalent to the sampled Fourier transform at frequencies $\omega_k = 2\pi k/N$. Typically, in a fast Fourier transform algorithm, N is chosen as a power of two to perform this computation very efficiently. It follows that the discrete short-time Fourier transform can be computed

$$\begin{aligned} X(n,k) &= \sum_{m=n-N+1}^{n} x(m)h(n-m)e^{\frac{-j2\pi km}{N}} && (2.75) \\ &= e^{\frac{j2\pi kn}{N}} \sum_{m=0}^{N-1} x(m+n)h(-m)e^{\frac{-j2\pi km}{N}} && (2.76) \\ &= e^{\frac{j2\pi kn}{N}} X_n(k) && (2.77) \end{aligned}$$

where $h(n)$ is a sampled windowing function of finite length with non-zero values only in the range $n = 0 \ldots N-1$. Then, $X_n(k)$ is the discrete Fourier transform of the windowed and time-shifted input signal.

The reassignment operations cannot be directly applied to the discrete short-time Fourier transform because partial derivatives cannot be computed on discretely sampled data but it is possible to approximate the partial derivatives using the method of finite differences.

$$\begin{aligned} \frac{\partial\phi(\tau,\omega)}{\partial t} &\approx \frac{1}{\Delta t}\left[\phi(t+\frac{\Delta t}{2},\omega) - \phi(t-\frac{\Delta t}{2},\omega)\right] && (2.78) \\ \frac{\partial\phi(\tau,\omega)}{\partial \omega} &\approx \frac{1}{\Delta \omega}\left[\phi(t,\omega+\frac{\Delta \omega}{2}) - \phi(t,\omega-\frac{\Delta \omega}{2})\right] && (2.79) \end{aligned}$$

For small values of Δt and $\Delta\omega$, this finite differences yield a good approximation of the partial derivatives of phase, which can be used in (2.72) and (2.73) to find the reassigned spectrogram.

For a more complete review of the method of reassignment, see [38], [39], [40].

2.4.2.3 Zoom FFT

The zoom FFT is a useful technique for focussing on a narrow frequency band in spectral analysis. When performing an FFT, the spectral resolution is determined by the total length of time spanned by the samples and not the sample rate. At higher sample rates, larger number of samples occupy a given time span. So for high resolution with high sample rate, where only a narrow band is of interest, FFT sizes are unneccessarily large and wasteful. In fact, for practical implementations on DSP hardware, very large FFT sizes may be impossible to realize.

The zoom FFT is a process where an input signal is mixed down to baseband and then decimated, prior to passing it into a standard FFT. The zoom FFT uses a reduction of sample rate to make a limited number of samples span a longer time; hence increasing the frequency resolution of the FFT.

The advantage is, for example, that if you have a sample rate of 10 MHz and require at least 10Hz resolution over a small frequency band, say 1 KHz, then you do not need a 2^{20} point FFT. Instead you just need decimate by a factor of 4096 (2^{12}) and use a 256 (2^8) point FFT which is obviously quicker. The zoom-FFT uses digital down conversion techniques to localise the standard FFT to a narrow band of frequencies that are centered on a higher frequency.

Advantages of the zoom FFT are :

- Increased frequency domain resolution
- Reduced hardware cost and complexity
- Wider spectral range

A continuous signal $x(t)$ is sampled by an analog-to-digital converter at a rate f_{s_1}. The resulting digitized N-point signal $x(n)$ has a spectral magnitude $|X(m)|$. The signal of interest is a narrow

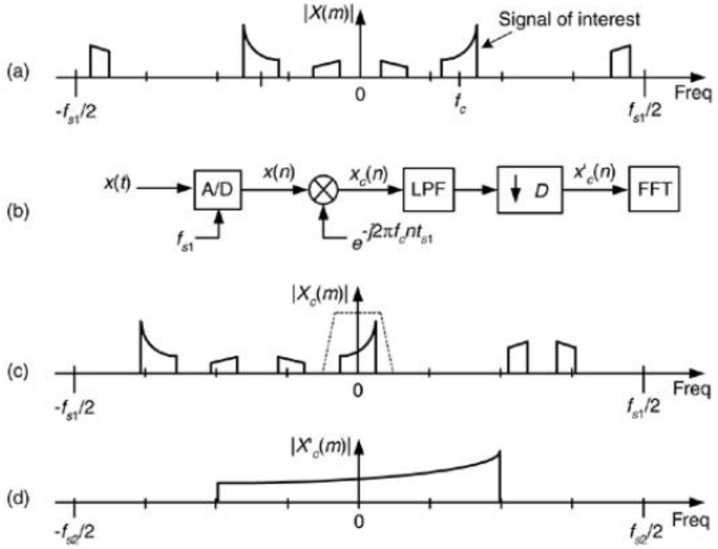


Figure 2.13: Zoom FFT

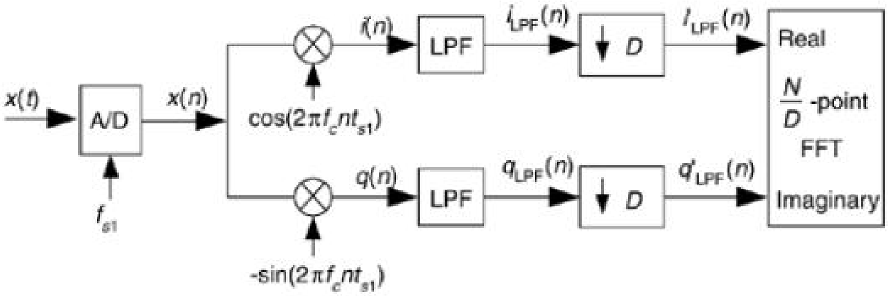


Figure 2.14: Zoom FFT Processing Diagram

band signal centered around frequency f_c. The zoom FFT technique dictates that we perform narrow band filtering and decimation in order to reduce the number of samples prior to the final FFT. First the signal is multiplied by a complex signal with frequency f_c and downconverted. The downconverted signal's spectrum, centered at 0 Hz, is $|X_c(m)|$. After low-pass filtering $x_c(n)$, the filter's output is decimated by an integer factor D yielding a time sequence $x'_c(n)$ with sample rate $f_{s_2} = f_{s_1}/D$. The length of $x'_c(n)$ is N/D allowing a reduced size FFT. The FFT is performed over the bandwidth of the decimated signal i.e. $-f_{s_2}/2$ to $f_{s_2}/2$ (figure 2.13 [41]).

The steps involved in zoom FFT shown in figure 2.14 [41] are :

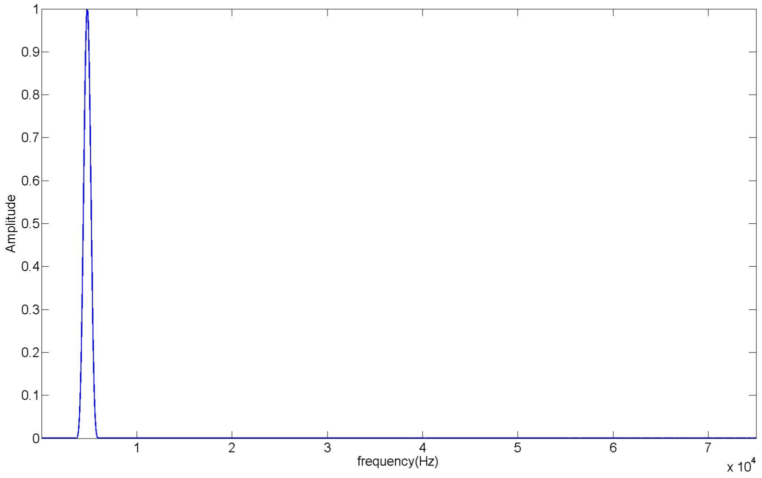
- frequency translation by complex downconversion
- low-pass filtering
- decimation
- FFT of smaller length

Figure 2.15 shows the FFT and zoom FFT of a constant frequency (f=4800 Hz) signal sampled at 150 KHz. The length of the FFT was 2^{14} which gives us a resolution of $\approx$1 Hz. The zoom FFT was performed with frequency translation centered around 5 KHz and a downsampling factor of 32. The FFT length for the downsampled signal is then only 2^9.

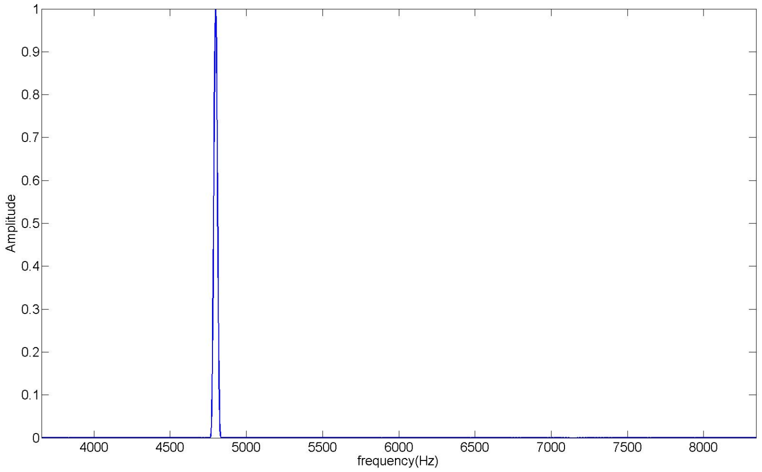
Figure 2.16 shows the spectrogram (short-time FFT) of the same signal but with a Hanning windowing function of length 2^{13} and overlap of 90%.

Combining the better frequency resolution of the reassigned spectrogram and the lower computation cost of the zoom FFT, we created a speeded up reassigned spectrogram in the spectral band of interest (Figures 2.17 and 2.18).

Figure 2.19 shows the improvement in the frequency distribution of the received Doppler signal when reassignment is used.

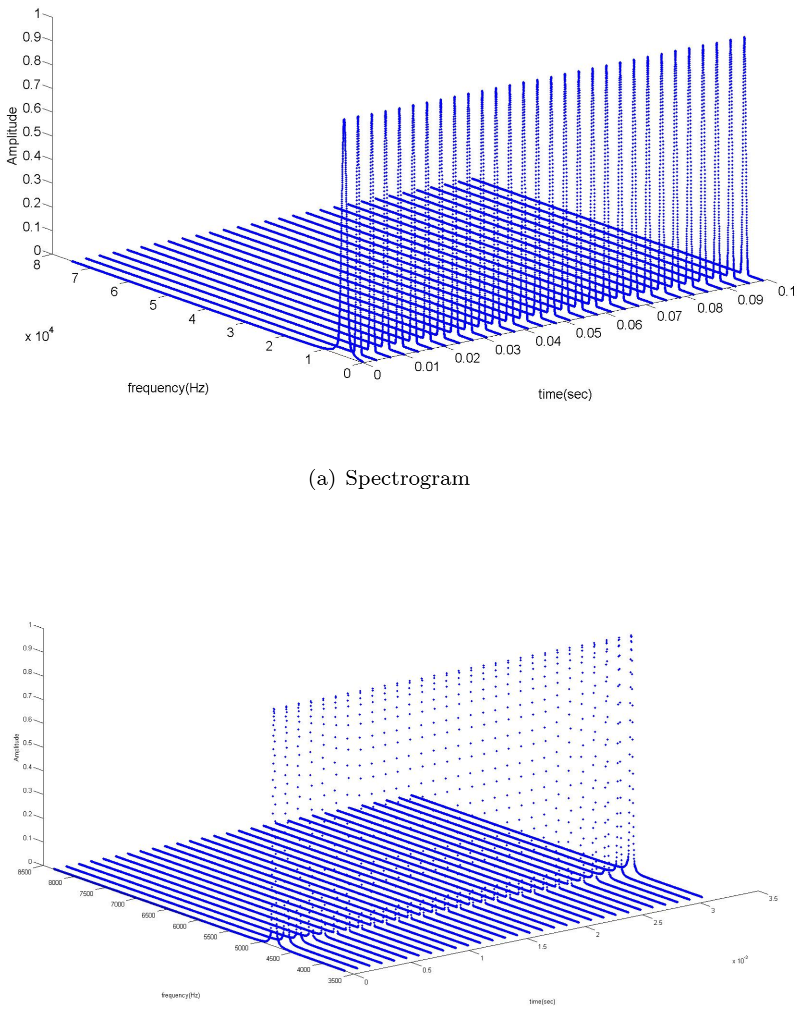


(a) Full FFT



(b) Zoom FFT

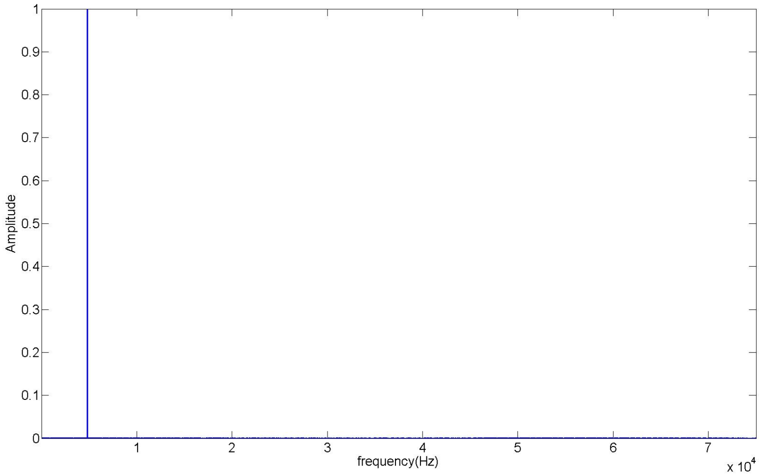
Figure 2.15: Comparison of FFT and Zoom FFT



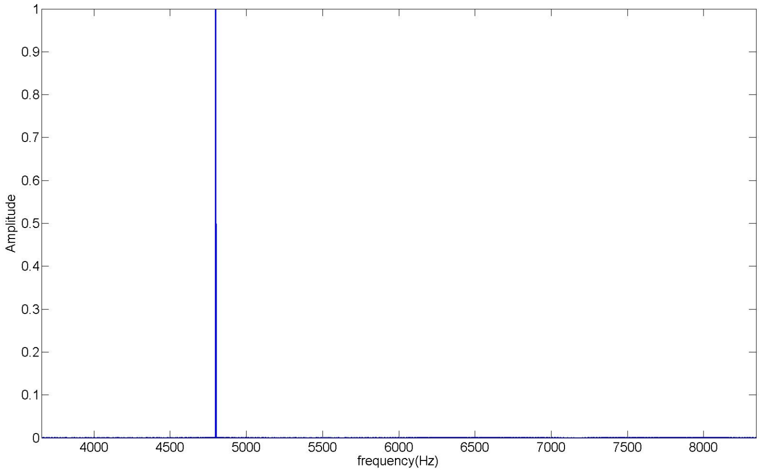
(a) Spectrogram

(b) Zoom Spectrogram

Figure 2.16: Comparison of Spectrogram and Zoom Spectrogram

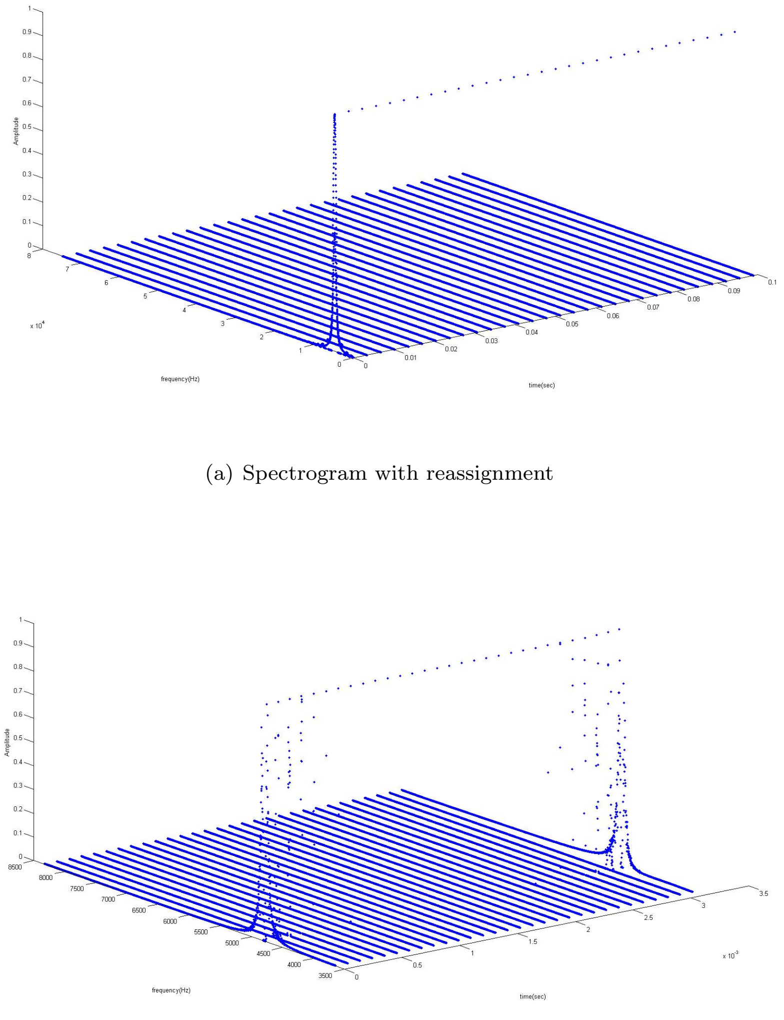


(a) Full FFT with reassignment



(b) Zoom FFT with reassignment

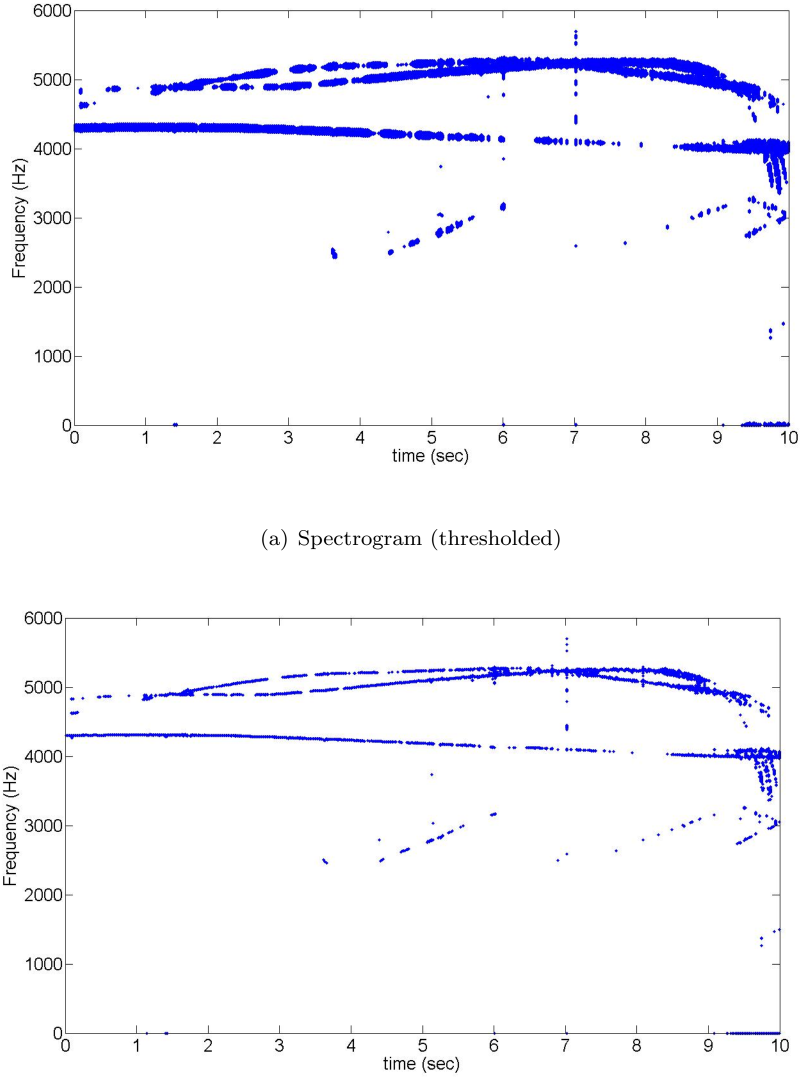
Figure 2.17: Comparison of FFT and Zoom FFT with reassignment



(a) Spectrogram with reassignment

(b) Zoom Spectrogram with reassignment

Figure 2.18: Comparison of Spectrogram and Zoom Spectrogram with reassigment



(a) Spectrogram (thresholded)

(b) Reassigned Spectrogram (thresholded)

Figure 2.19: Reassigned Spectrogram of Doppler Radar data

3

Video + Doppler

3.1 Setup

Figure 3.1 shows the proposed scenario [42]. A sensor package is rigidly mounted on a platform. The sensor platform is situated at height h above the region of interest and continuously monitors the region of interest.

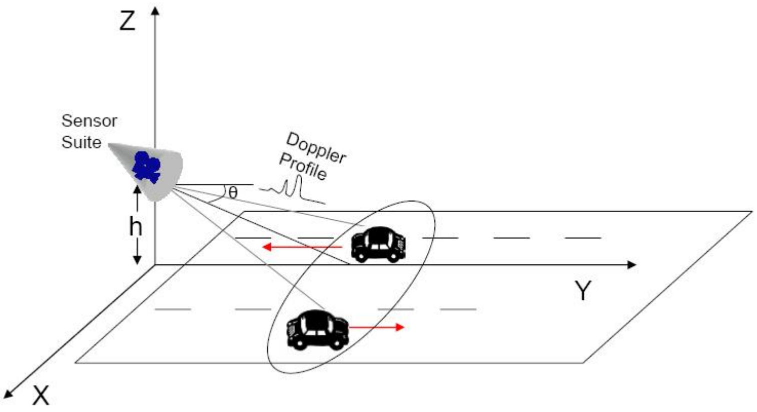


Figure 3.1: Automated Speed Enforcement Scenario

Figure 3.2 shows the overview of the proposed system. The high resolution (1024x768 pixel) video camera captures the scene at video frame rate (30 Hz). Simultaneously, data is collected by

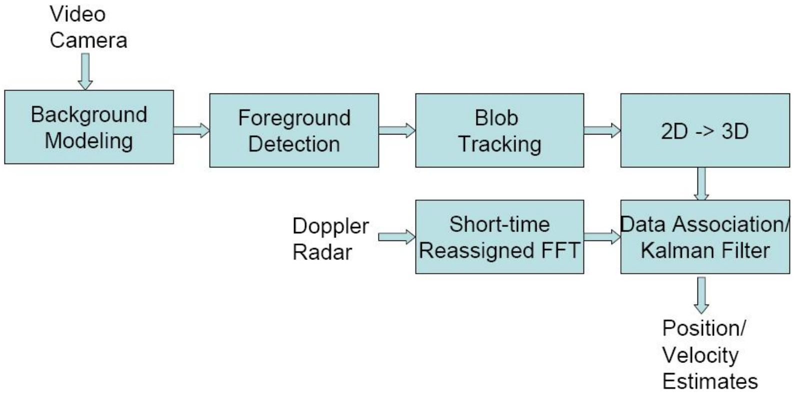


Figure 3.2: Elements of Fusion Scheme

the Doppler radar at 96 KHz and is processed to extract the velocities of all the vehicles within the microwave beamwidth of the radar. The velocity histories are maintained over time and seperated from background clutter.

Special Geometry

It is obvious that we have the freedom to choose the world coordinate frame while mapping a point from the world coordinate system to the image plane. Assume the scene being observed is a plane, then we can choose a coordinate frame such that the XY-plane of the world coordinate system corresponds to the plane being observed. Then, all points on the scene plane will have a zero Z-coordinate.

This leads to an interesting mapping between the scene plane and the image plane,

$$\underline{x} = P\underline{X} = \begin{bmatrix} p_{11} & p_{21} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} \tag{3.1}$$

$$= \begin{bmatrix} p_{11} & p_{21} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{3.2}$$

which is a 3 × 3 matrix representing a plane to plane projective transformation. From (3.2), it can be seen that if we can accurately determine the camera intrinsic and extrinsic parameters, we can recover the world coordinates of any point on the scene plane from its image plane coordinates.

$$x = \frac{p_{11}X + p_{12}Y + p_{14}}{p_{31}X + p_{32}Y + p_{34}} \tag{3.3}$$

$$y = \frac{p_{21}X + p_{22}Y + p_{24}}{p_{31}X + p_{32}Y + p_{34}} \tag{3.4}$$

$$p_{31}xX + p_{32}xY + p_{34}x = p_{11}X + p_{12}Y + p_{14} \tag{3.5}$$

$$p_{31}yX + p_{32}yY + p_{34}y = p_{21}X + p_{22}Y + p_{24} \tag{3.6}$$

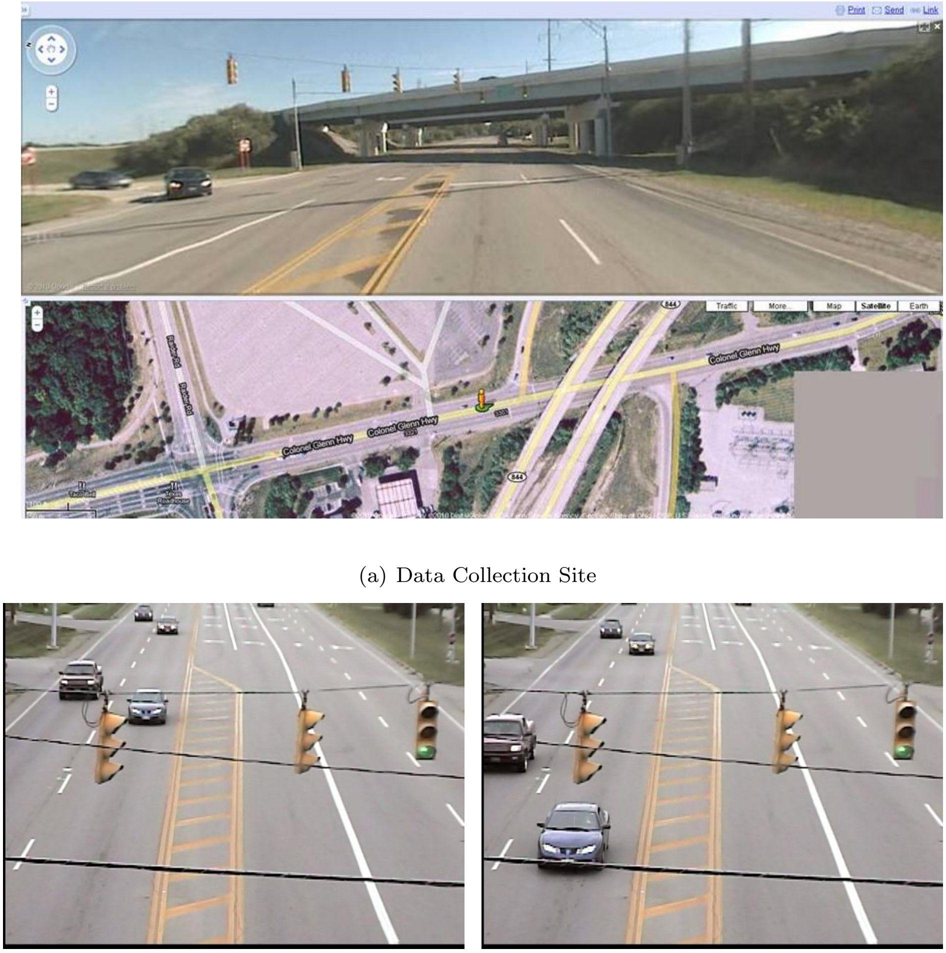
$$\begin{bmatrix} p_{31}x - p_{11} & p_{32}x - p_{12} \\ p_{31}y - p_{21} & p_{32}y - p_{22} \end{bmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} p_{14} - p_{34}x \\ p_{24} - p_{34}y \end{pmatrix} \tag{3.7}$$

or

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} p_{31}x - p_{11} & p_{32}x - p_{12} \\ p_{31}y - p_{21} & p_{32}y - p_{22} \end{bmatrix}^{-1} \begin{pmatrix} p_{14} - p_{34}x \\ p_{24} - p_{34}y \end{pmatrix} \tag{3.8}$$

3.2 Processing

We collected synchronized video and radar data from a bridge overlooking a busy four lane road (Figure 3.3). The height of the camera above the road and its pitch angle were measured manually. The sensor height was observed to be $8m$ above the road (plane) and at angle of $10°$ below the horizon.



(a) Data Collection Site

(b) Frame 45

(c) Frame 60

Figure 3.3: Collection site and sample data collected from the camera

$$P = K\left[R|t\right] = \left[R_x * R_y * R_z|t\right] \tag{3.9}$$

Assuming that there is no roll ($R_y = I$) and yaw (($R_z = I$), we can calculate the camera extrinsics. The camera intrinsics were found using Bouguet's toolbox [43] to be

$$K = \begin{bmatrix} 2093.12 & 0 & 320 \\ 0 & 2093.12 & 240 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.10}$$

Using the camera intrinsics and the camera extrinsics, we can find the location of any 2D point from image plane on the ground plane (XY-plane) using (3.8). Using background-foreground segmentation and connected component analysis, all blobs in the video are tracked. The 3D positions of the tracked targets are used as measurements in a Kalman filter to estimate the position and velocity of the vehicles.

The state vector for the Kalman filter is

$$\underline{X}_k = \begin{pmatrix} X \\ \dot{X} \\ Y \\ \dot{Y} \end{pmatrix} \tag{3.11}$$

The state model used for the Kalman filter is a simple constant velocity model,

$$\underline{X_k} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \underline{X}_{k-1} + w_{k-1} \tag{3.12}$$

with the observations being the positions recovered from the projective geometry,

$$\underline{Z}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underline{X}_k + v_k \tag{3.13}$$

The process noise covariance is given by

$$Q = \begin{bmatrix} T^4/4 & T^3/2 & 0 & 0 \\ T^3/2 & T^2 & 0 & 0 \\ 0 & 0 & T^4/4 & T^3/2 \\ 0 & 0 & T^3/2 & T^2 \end{bmatrix} \tag{3.14}$$

In order to improve the estimates of the speed and position of the tracked vehicles, we employ a sensor fusion scheme along with data association.

From the recovered position and velocity in world coordinates of each target, we can project the radial velocity in the direction of the Doppler radar. The Doppler radar will measure a lower velocity because of the cosine effect (2.49). Nevertheless, we can estimate the Doppler velocity from the position and velocity estimates obtained from the image plane.

From geometry in Figure 3.4, we can obtain the relationship between the magnitude of the radial velocity measured by the Doppler radar $|\underline{V}_d|$ and the true velocity $|\underline{V}|$, using the angles α and β,

$$|V_d| = |V| \cos\alpha \cos\beta$$

The angles α and β can be calculated using knowledge of the state of the target at time step $k-1$, by using the predicted position of the target $\underline{X}_{k|k-1} = F.\underline{X}_{k-1|k-1}$. If the predicted position $\underline{Y}_{k|k-1}$ and predicted velocity $\underline{V}$ of the targets are defined as,

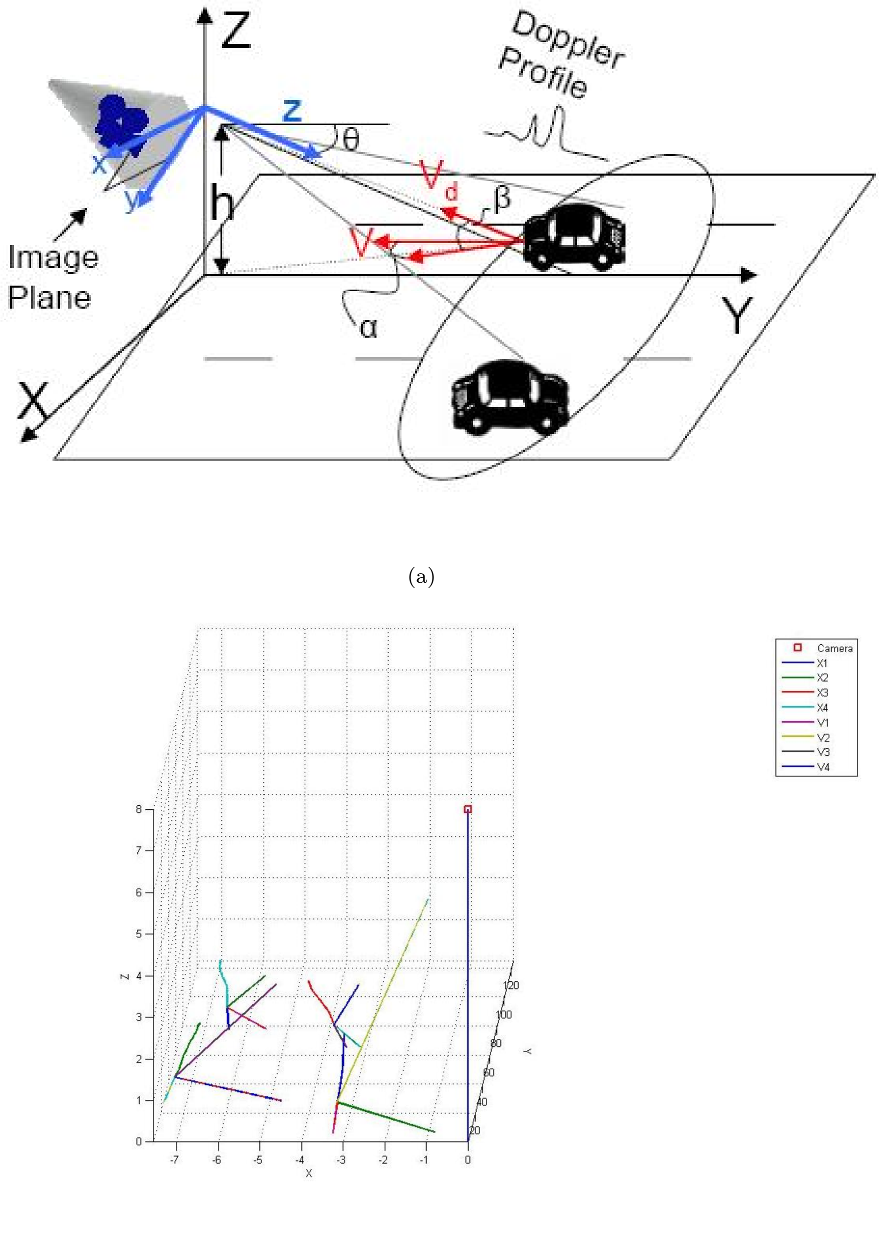
$$\underline{Y}_{k|k-1} = \begin{pmatrix} X_{k|k-1} \\ Y_{k|k-1} \end{pmatrix} \tag{3.15}$$

and

$$\underline{V}_{k|k-1} = \begin{pmatrix} \dot{X}_{k|k-1} \\ \dot{Y}_{k|k-1} \end{pmatrix} \tag{3.16}$$

and the camera center $\underline{C}$ is assumed to be known, then the angles α and β can be calculated as,

$$\cos\alpha = \frac{\underline{V}_{k|k-1} . \underline{Y}_{k|k-1}}{|\underline{V}_{k|k-1}||\underline{Y}_{k|k-1}|} \tag{3.17}$$



(a)

(b)

Figure 3.4: Target trajectories along with projection of radial velocities measured by Doppler

and

$$\cos\beta = \frac{\underline{V}^{XY}_{k|k-1} . \underline{C}_{Y_{k|k-1}}}{|\underline{V}^{XY}_{k|k-1}||\underline{C}_{Y_{k|k-1}}|} \tag{3.18}$$

where

$$\underline{V}^{XY}_{k|k-1} = |V_{k|k-1}| cos\alpha . \frac{\underline{Y}_{k|k-1}}{|\underline{Y}_{k|k-1}|} \tag{3.19}$$

and

$$\underline{C}_{Y_{k|k-1}} = \underline{Y}_{k|k-1} - \underline{C} \tag{3.20}$$

So we can calculate $|V_d|$ from the projections of the velocity in the image plane.

The Doppler radar is sampled at 96 kHz and converted into frequency measurements using short-time reassigned FFT (N=2^{14} point FFT) for the sequence of measurement collected in the time interval $(n\tau - \tau/2, n\tau + \tau/2)$, where $\tau = 1/f_s$; $f_s = 30$ Hz is the frame rate of the camera (Figure 3.5).

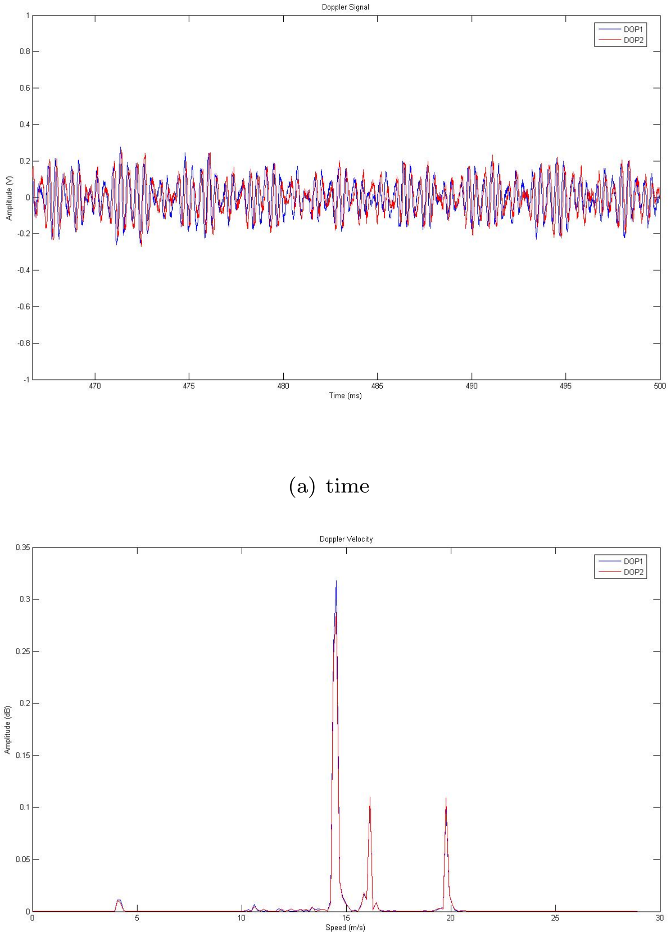
Using a sliding window along the time axis of the spectrogram, the frequency is fit to multiple lines using the 2D RANSAC from section 2.3.1. The algorithm used (Figure 3.7) models the frequency of each target as a line with restrictions on the change in slope between sliding windows. In this way, we can reject noise as well as cross interference from nearby frequencies.

Using the predicted Doppler velocity (Figure 3.4) obtained from the video, we form data association with the true measurement of Doppler velocity by the Doppler radar (Figure 3.6).

Once data from the Doppler radar and video has been associated, we use the true Doppler velocity and project it back to the 3D ground plane. This velocity is then used to update the state of the Kalman filter.

In the sensor fusion framework, state vector remains the same along with the process model but the observation is now changed to

$$h(\underline{X}) = \begin{bmatrix} X \\ Y \\ \sqrt{\dot{X}^2 + \dot{Y}^2} \end{bmatrix} \tag{3.21}$$



(a) time

(b) frequency

Figure 3.5: Doppler radar data in time and frequence domain

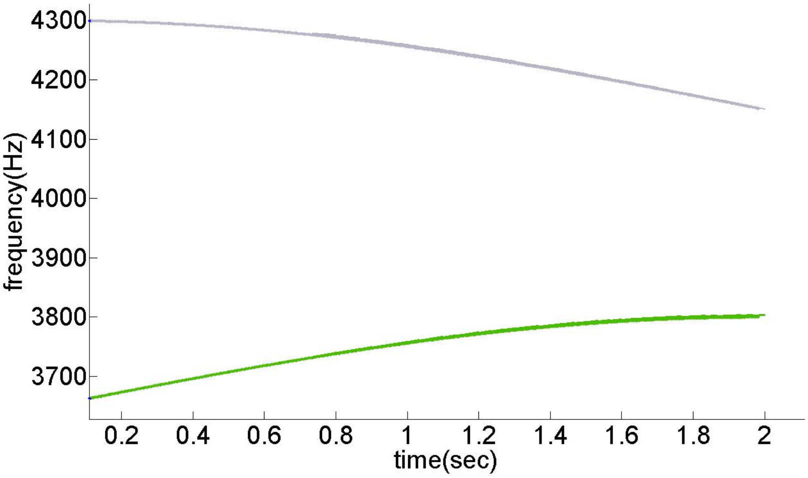


Figure 3.6: RANSAC for Frequency Tracking

```
Initialization : set number of target = 0

1. Perform the reassigned spectrogram of data using sliding windows
2. Get the set $\mathcal{S} = (t_i, f_i)$ of points which have amplitude greater than threshold
3. While k < number of targets
    (a) check for points $p_k$ in $\mathcal{S}$ that belong to the line $L_{k-1}$
    (b) use $p_k$ to define a line $L_k$ that approximates the frequency of target $k$
    (c) remove $p_k$ from the set $\mathcal{S}$
4. while $\mathcal{S} \neq \emptyset$ search for new targets in set $\mathcal{S}$
    (a) use RANSAC (figure 2.3) to search for lines in $\mathcal{S}$
    (b) if the set of inliers $p_k >$ threshold
        i. Set k=k+1
        ii. define a line $L_k$ from the set of inliers
        iii. remove $p_k$ from the set $\mathcal{S}$
```

Figure 3.7: RANSAC for Frequency Tracking Pseudo-code

Since the observation is now non-linear, an extended Kalman filter has to be used in place of the standard Kalman filter. In the time update step of the Kalman filter, the Doppler radar measurements are used to correct the error in the velocity (and indirectly position) obtained from the video. The Jacobian of the measurement model is given by

$$\hat{H}_k = \frac{\partial h(\underline{X}_{k|k-1})}{\underline{X}_{k|k-1}} = \left.\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{\dot{X}}{\sqrt{\dot{X}^2+\dot{Y}^2}} & \frac{\dot{Y}}{\sqrt{\dot{X}^2+\dot{Y}^2}} \end{bmatrix}\right|_{\underline{X}_{k|k-1}} \tag{3.22}$$

In order to place more emphasis on the measurements from the Doppler radar, the measurement noise covariance matrix R is chosen as,

$$R = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_1^2 & \\ & & \sigma_2^2 \end{bmatrix} \tag{3.23}$$

where $\sigma_1 > \sigma_2$, which places more weight on the Doppler radar measurements.

3.3 Results

Figure 3.8 shows the result of tracking the targets using only the image plane estimates obtained by background subtraction and blob tracking.

The estimate of the velocity from just the video camera is lower than the posted speed limits (which are a good indicator of vehicle speeds). The speed of the fastest vehicle is about 16 m/s (35 mph) which is significantly lower than the posted speed of 45 mph.

Figure 3.9 shows the effect of adding the sensor feedback from Doppler radar to the estimation process. The estimates of the X and Y positions are updated by the velocity measured by the Doppler radar. The effect can be seen in Figure 3.10 where the estimates of the velocities in the image plane begin to converge to the measured Doppler velocity.

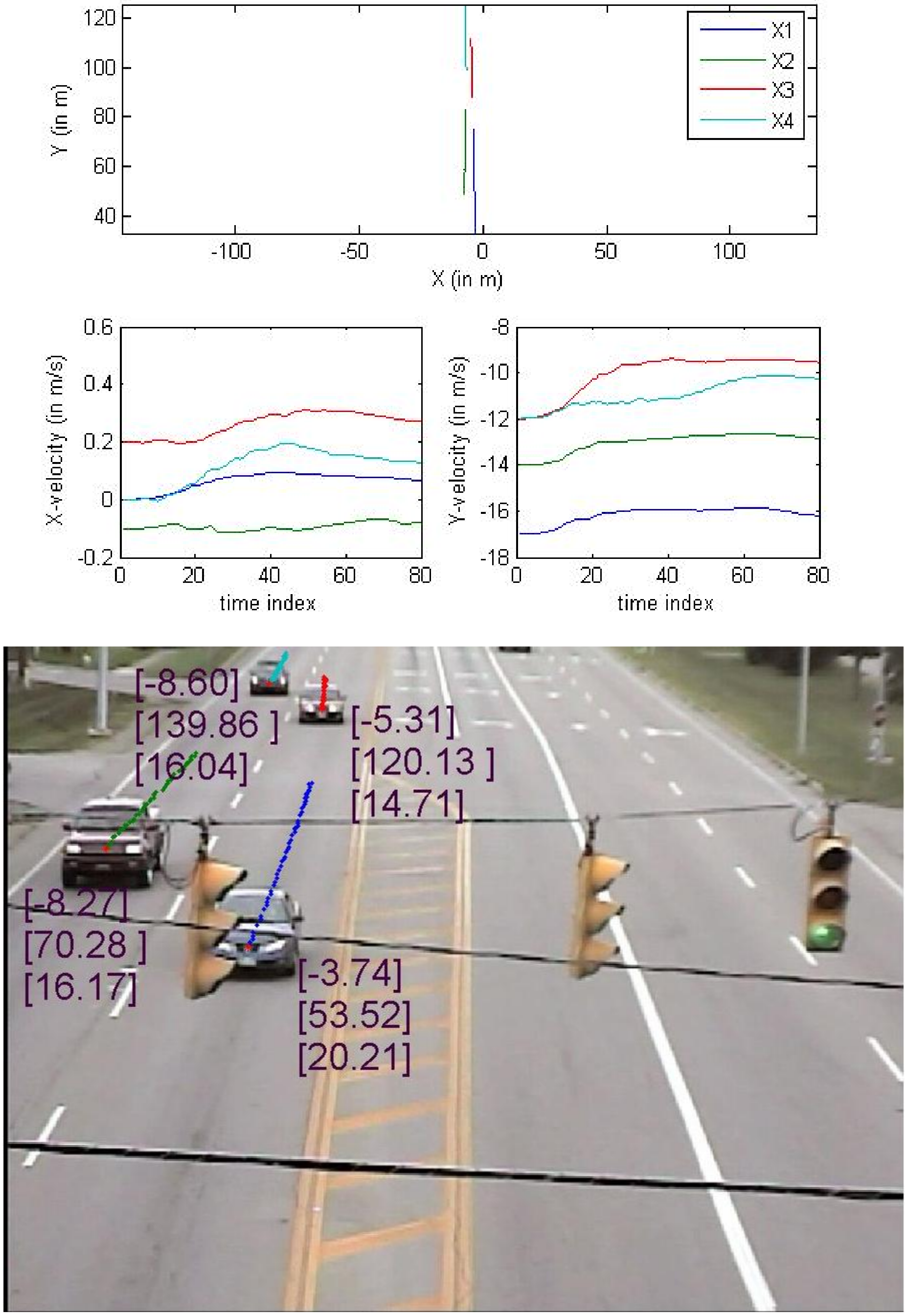


Figure 3.8: Estimated position using camera projection matrix

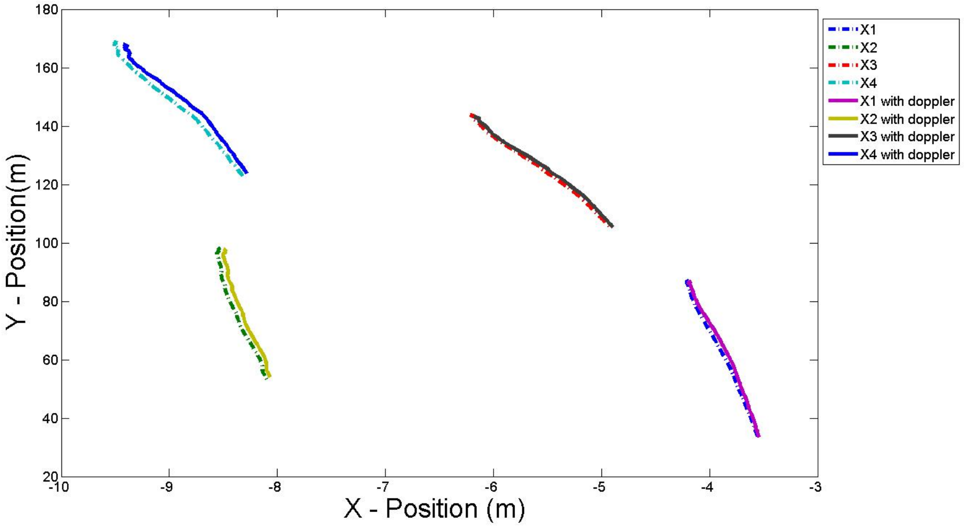


Figure 3.9: Comparison of estimated position from camera only and with camera and Doppler sensor fusion

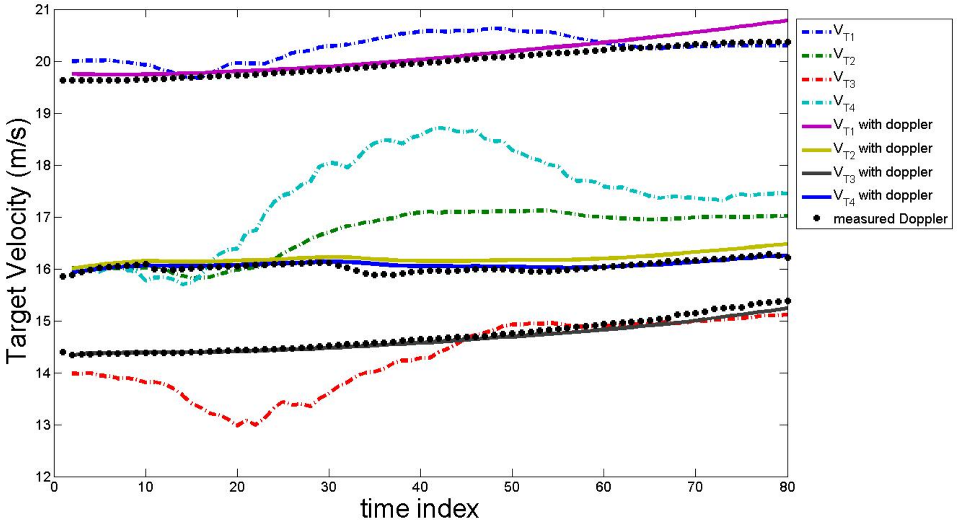


Figure 3.10: Comparison of estimated velocity from camera only and with camera and Doppler sensor fusion

We do not have any ground truth data but cursory observation of the velocity of the cars reveals a maximum speed of about 20 m/s (45 mph), which is the posted speed limit at the site of the data collection.

4

Vidar : Video-Doppler-Radar

Since the Doppler radar does not give us any angle or range information, there is ambiguity when associating the measurements from the radar to the video which cannot be resolved directly. Multiple hypothesis testing is one method that could be used when the data association information is not readily available. The video to radar matching problem is classified as a one-to-many problem with no unique solution for a single iteration. As the number of targets increases, the problem becomes more severe. To resolve the ambiguity, it would be very helpful if we were able to obtain an angle measurement as well from the radar.

Our solution to the problem is unique in that we employ a physical method to pseudo-modulate the frequency of the CW Radar [44],[45],[46]. We effect this modulation by continuously moving the two Radars in a causal fashion. In the following section, we provide a description of the device and its functioning.

4.1 Device Description

A device was prototyped comprising of the sensors along with data processing as well as data recording instruments in a sensor suite package dubbed the Vidar (VIdeo-Doppler-RadAR) The internal layout of the Vidar is shown in Figure 4.1. It is has a first moving Doppler radar, a second

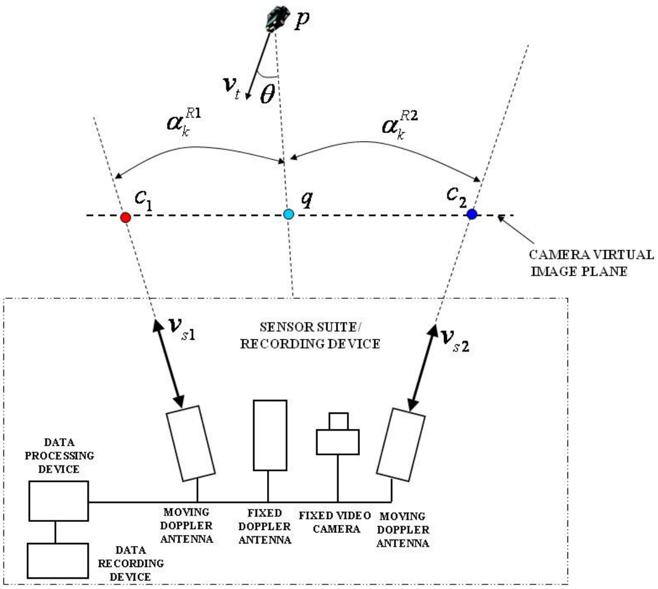


Figure 4.1: Vidar Device Motion Geometry

moving Doppler radar, a fixed or stationary Doppler radar, a fixed video camera, a data recording and processing device (a computer with a data acquisition card and a hard drive).

Figure 4.1 also indicates the sensing geometry where the camera virtual image plane of the video camera, the first moving Doppler radar motion ray, the second moving Doppler radar motion ray, the radar direction ray connecting the Vidar to a moving vehicle and the intersection of the first Doppler radar motion ray with the virtual image plane and the intersection of the second Doppler radar motion ray with the virtual image plane are all shown.

The first and second Doppler radars in the Vidar apparatus are extended or retracted (moved side to side as illustrated in Figure 4.1 on sliding tracks (not shown) by a stepper motor (not shown). An optical encoder is mounted on each of the tracks so that the sliding speeds of the Doppler radars (v_{s_1} and v_{s_2}) and the relative offsets to each other can be predetermined. The sliding track orientation angles (ψ_{s_1} and ψ_{s_2}) are also predetermined. Using a calibration method, the intersections first and second motion rays with the virtual image planes are predetermined as well.

4.2 Charecterization of the Sensor's Motion

To provide constant (causal) motion to two independent Doppler radars, we designed a system of tracks moving in different directions driven by stepper motor spinning at a constant velocity. The motion of the radars is then analogous to the reciprocating motion of a piston connected to a crank through a connecting rod in an internal combustion engine.

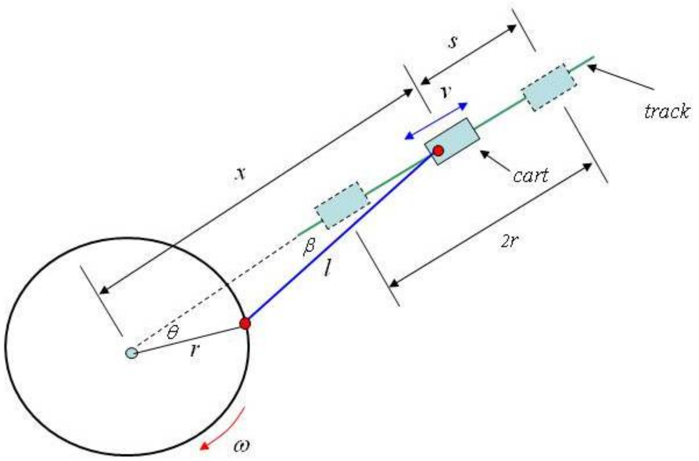


Figure 4.2: Sensor Motion Apparatus

Figure 4.2 shows the setup of a piston-crank rod system. Let us define the quantities of interest as

- l = rod length (distance between piston pin and crank pin)
- r = crank radius (distance between crank pin and crank center, i.e. half stroke)
- θ = crank angle (from cylinder bore centerline at TDC)
- x = piston pin position (upward from crank center along cylinder bore centerline)
- v = piston pin velocity (upward from crank center along cylinder bore centerline)
- ω = crank angular velocity in rad/s

The crankshaft angular velocity ω is related to the revolutions per minute (rpm) made by the crank.

$$\omega = \frac{2\pi rpm}{60} \tag{4.1}$$

We note the relationship (using the law of cosines)

$$l^2 = r^2 + x^2 - 2rx\cos\theta \tag{4.2}$$

Rearranging the terms, we get

$$l^2 - r^2 = x^2 - 2rx\cos\theta \tag{4.3}$$

$$l^2 - r^2 = x^2 - 2rx\cos\theta + r^2\left((\cos^2\theta + \sin^2\theta) - 1\right) \tag{4.4}$$

$$l^2 - r^2 + r^2 - r^2\sin^2\theta = x^2 - 2rx\cos\theta + r^2\cos^2\theta \tag{4.5}$$

$$l^2 - r^2\sin^2\theta = (x - r\cos\theta)^2 \tag{4.6}$$

$$x - r\cos\theta = \sqrt{l^2 - r^2\sin^2\theta} \tag{4.7}$$

$$x = r\cos\theta + \sqrt{l^2 - r^2\sin^2\theta} \tag{4.8}$$

In order to get the relation of velocity with respect to time we differentiate (4.8)

$$\begin{aligned}
v &= \frac{dx}{dt} && (4.9)\\
&= \frac{dx}{d\theta}\frac{d\theta}{dt} \qquad \text{(chain rule)} && (4.10)\\
&= \left(\frac{d}{d\theta}(r\cos\theta + \sqrt{l^2 - r^2\sin^2\theta})\right).(\omega) \qquad \text{since } \theta = \omega t,\ d\theta/dt = \omega && (4.11)\\
&= \omega\left(-r\sin\theta - \frac{(\frac{1}{2})(-2)r^2\sin\theta\cos\theta}{\sqrt{l^2 - r^2\sin^2\theta}}\right) && (4.12)\\
&= -\omega r\sin\theta - \frac{\omega r^2\sin\theta\cos\theta}{\sqrt{l^2 - r^2\sin^2\theta}} && (4.13)
\end{aligned}$$

The analytical closed form equation of the velocity was compared to the velocity measured by the linear optical encoded fitted on the tracks and found to be in close agreement (Figure 4.3).

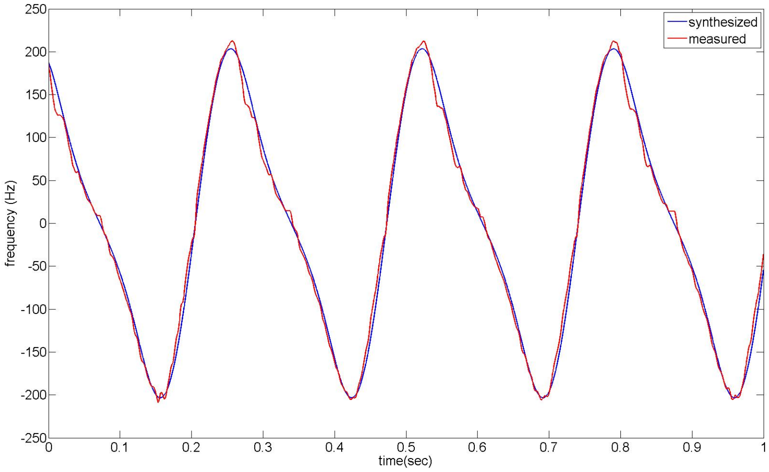


Figure 4.3: Sensor Motion Comparison

4.3 Doppler Radar Processing and Tracking

Figure 4.4 shows the reassigned spectrogram of data from a typical data collection on the Vidar device.

The Doppler frequencies of a moving vehicle k measured by the stationary radar f_D^3 according to (2.49) is

$$f_{D_k}^3 = K_a V_{T_k} \cos\theta_k \tag{4.14}$$

The frequencies of targets in the stationary radar are tracked (Figure 4.5) using the alogorithm listed in Figure 3.7.

The purpose of adding a motion to the moving radars is to extract the azimuth angle from the modulation index. In the following section, we show how the azimuth angle can be recovered from the information from the three radars only.

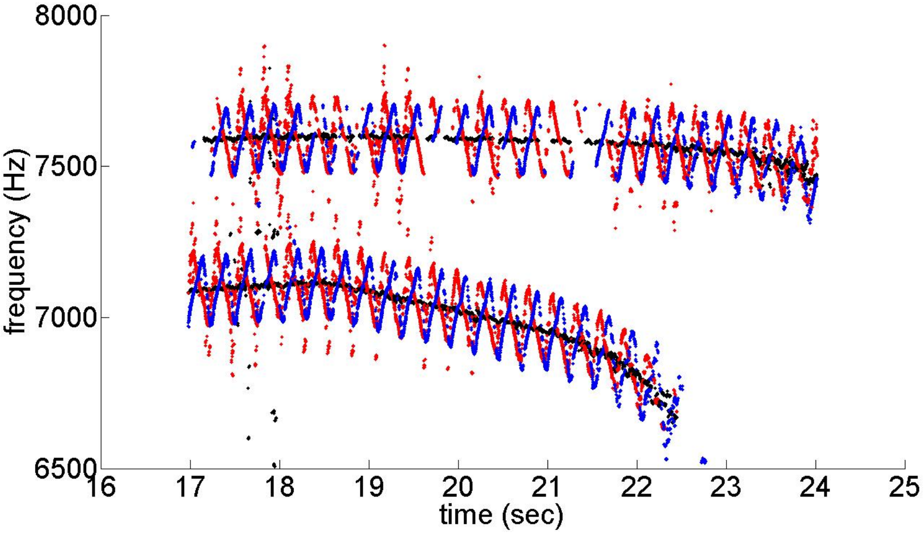


Figure 4.4: Reassigned Spectrogram of Sample Radar Data

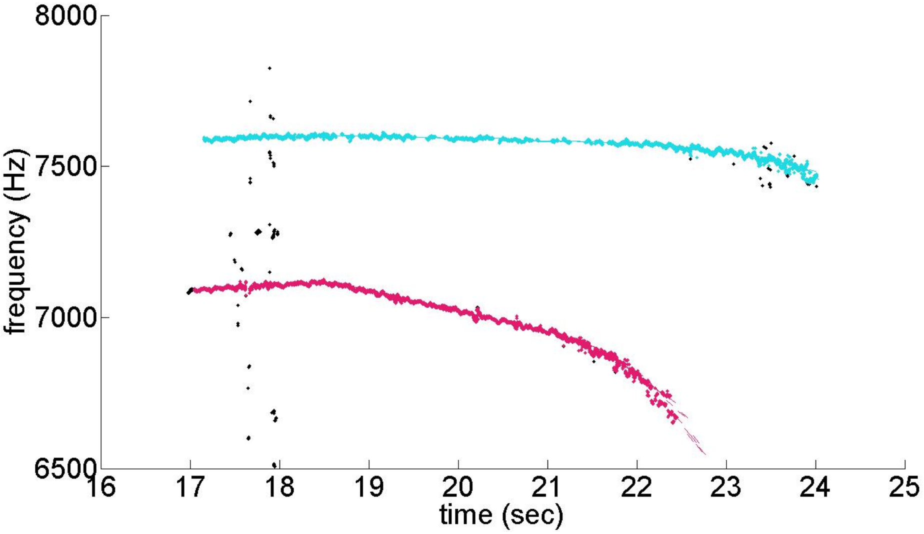


Figure 4.5: Stationary Radar frequency tracking using 2D RANSAC

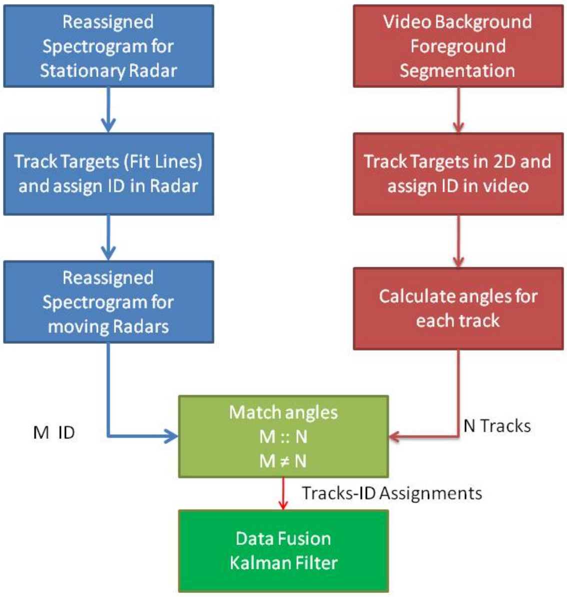

Reassigned Spectrogram for Stationary Radar
Track Targets (Fit Lines) and assign ID in Radar
Reassigned Spectrogram for moving Radars
Video Background Foreground Segmentation
Track Targets in 2D and assign ID in video
Calculate angles for each track
M ID
N Tracks
Match angles M :: N M ≠ N
Tracks-ID Assignments
Data Fusion Kalman Filter

Figure 4.6: Vidar Processing Block Diagram

4.4 Radar only Tracking

When both the sensor and the target are in motion, the measured Doppler frequency is the summation of the individual Doppler frequencies. Therefore, the Dopppler frequency measured by the moving radars is

$$f^1_{D_k} = K_a V_{T_k} \cos\theta_k + K_a V_{s_1} \cos\alpha^{R_1}_k \tag{4.15}$$

$$f^2_{D_k} = K_a V_{T_k} \cos\theta_k + K_a V_{s_2} \cos\alpha^{R_2}_k \tag{4.16}$$

The measurement from the three radars contains information about the target velocity as well as its azimuth angle. We first look at how we can extract this information by processing only the Doppler radar data.

Algorithm Development

The two quantities of interest are the target true velocity V_{T_k} and azimuth angle η_k. We address each of these in following sections.

Recovering the azimuth angle

From (4.15) and (4.16), we can rearrange to write,

$$\cos\alpha_k^{R_1} = \frac{f_{D_k}^1 - K_a V_{T_k}\cos\theta_k}{K_a V_{s_1}} \tag{4.17}$$

$$\cos\alpha_k^{R_2} = \frac{f_{D_k}^1 - K_a V_{T_k}\cos\theta_k}{K_a V_{s_2}} \tag{4.18}$$

From (4.14),

$$\cos\alpha_k^{R_1} = \frac{f_{D_k}^1 - f_{D_k}^3}{K_a V_{s_1}} \tag{4.19}$$

$$\cos\alpha_k^{R_2} = \frac{f_{D_k}^1 - f_{D_k}^3}{K_a V_{s_2}} \tag{4.20}$$

The denominator can be replaced by the analytical closed form from (4.13).

$$\Delta f_{D_k}^{13} = \cos\alpha_k^{R_1} = \frac{f_{D_k}^1 - f_{D_k}^3}{f_{s_1}} \tag{4.21}$$

$$\Delta f_{D_k}^{23} = \cos\alpha_k^{R_2} = \frac{f_{D_k}^1 - f_{D_k}^3}{f_{s_2}} \tag{4.22}$$

From (4.21) and (4.22), we can use a least squares scheme to solve for V_T and δ. Solving for δ for multiple targets gives us a good solution which can be used as a calibration constant for all further cases. With δ known, we can solve for η_k and V_T^k for any target directly.

4.5 Video and Radar Fusion and Tracking

Processing the information from the radars alone gives us an estimate (weak) of the azimuth angle and an estimate (strong) of the velocity of the targets. The video camera is very good sensor for measuring angles but cannot give us accurate information regarding the velocity while the Doppler radar is excellent for measuring velocity but cannot resolve angles accurately. We can apply a fusion scheme to aid in the improvement of both the estimated velocity and the estimated angles.

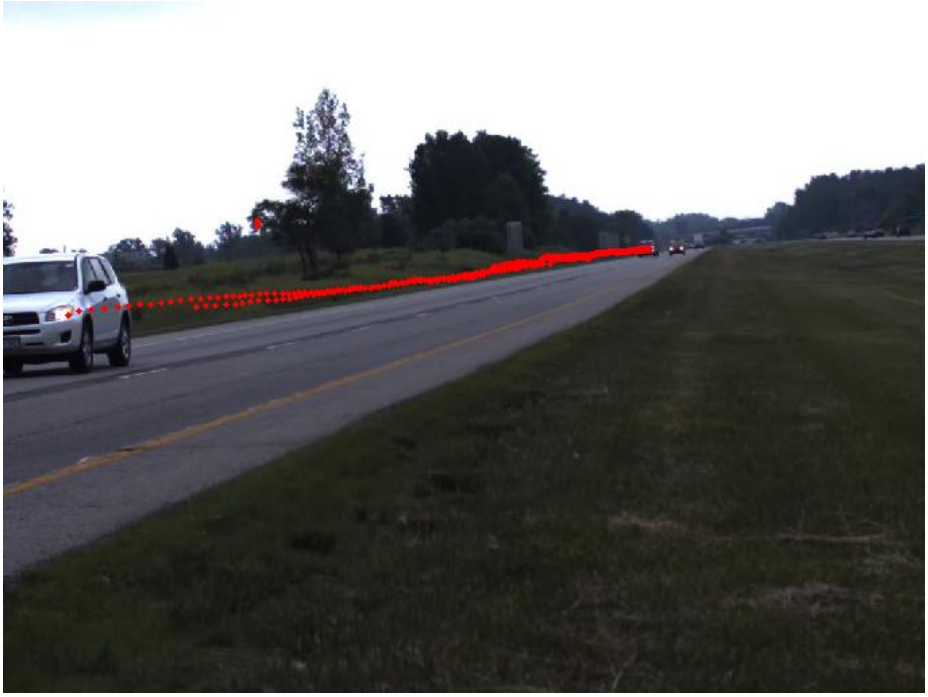


Figure 4.7: Video Tracks

The video is processed using the sequence of foreground segmentation as well as connected component analysis to give us tracks of foreground objects (Figure 4.7). By performing tracking on the video side as well, we are able to reject spurious noise and artifacts from the foreground segmentation process.

Since camera instrinsics are known by the calibration procedure, we can convert the (x, y) positions to angles quite easily (Figure 4.8).

Then, by using the angles obtained from the radar as well as video as a cue for data association, we can fuse the information to get a better estimate of the target velocity and azimuth angle. We can use the information in a Kalman filter with the azimuth angle from video as an extra measurement when available.

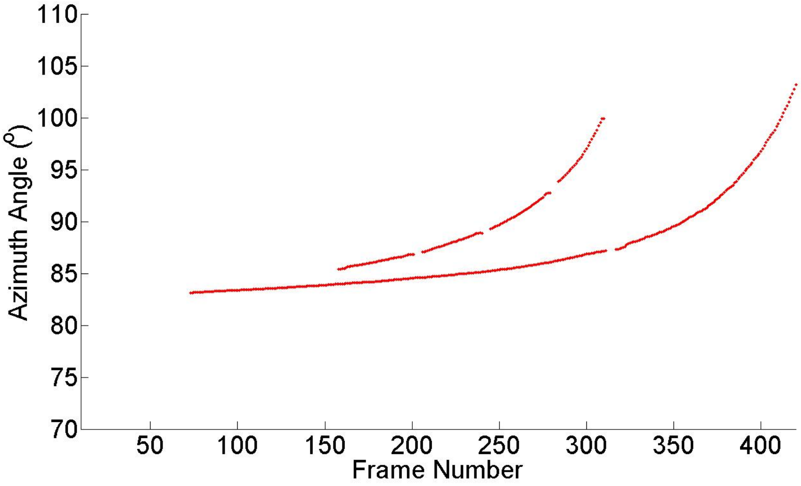


Figure 4.8: Video Azimuth Angles

Algorithm Development

Since the azimuth angle can be measured by two different sensors, it becomes important to appropriately weight the measurements from each sensor in order to improve the overall accuracy of the estimation process.

The state vector for the Kalman filter is

$$\underline{X}_k = \begin{pmatrix} V_{T_k} \\ \theta_k \\ \dot{\theta_k} \end{pmatrix} \tag{4.23}$$

The velocity is assumed to be constant so a simple markov model is used to propagate it while the azimuth angle is modelled by a constant velocity model to allow for a smooth change.

The state model used for the Kalman filter is

$$\underline{X_k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \underline{X}_{k-1} + w_{k-1} \tag{4.24}$$

with the observations $\underline{Z}$ being

$$\underline{Z}_k = h(\underline{X}, \underline{U}) + v_k \tag{4.25}$$

and the inputs $\underline{U}$

$$\underline{U}_k = \begin{bmatrix} p_k \\ q_k \\ \delta \end{bmatrix} \tag{4.26}$$

The process noise is assumed to have a Gaussian distribution with zero mean and covariance Q given by

$$Q = \begin{bmatrix} T^2 & 0 & 0 \\ 0 & T^4/4 & T^3/2 \\ 0 & T^3/2 & T^2 \end{bmatrix} \tag{4.27}$$

$$h(\underline{X}, \underline{U}) = \begin{bmatrix} f^3_{D_k} \\ 0 \\ \theta_k \end{bmatrix} = \begin{bmatrix} K_a V_{T_k} \cos\theta_k \\ p_k \cos(\theta_k + \delta) + q_k \cos(\theta_k + \delta) \\ \theta_k \end{bmatrix} \tag{4.28}$$

Since the observation is now non-linear, an Extended Kalman filter has to be used. The jacobian of the measurement model is given by

$$\hat{H}_k = \frac{\partial h(\underline{X}_{k|k-1})}{\underline{X}_{k|k-1}} = \left. \begin{bmatrix} \cos\theta_k & -V_{T_k} \sin\theta_k & 0 \\ p_k \cos(\theta_k + \delta) & q_k \sin(\theta_k + \delta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \right|_{\underline{X}_{k|k-1}} \tag{4.29}$$

In order to place more emphasis on the measurements from the Doppler radar, the measurement noise covariance matrix R is chosen as,

$$R = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{bmatrix} \tag{4.30}$$

where $\sigma_3 \ll \sigma_2$, which places more weight on the angle measurements from video.

5

Vidar Results

In this chapter, we show simulation results on data generated in MATLAB for the Vidar. We apply the alogorithms developed in the previous sections to estimate the target velocity and azimuth angle. We also tested ground truth data collected with the estimated velocities and azimuth angles. Finally, the feasibility of the algorithms for multi-target tracking is shown.

5.1 Simulations

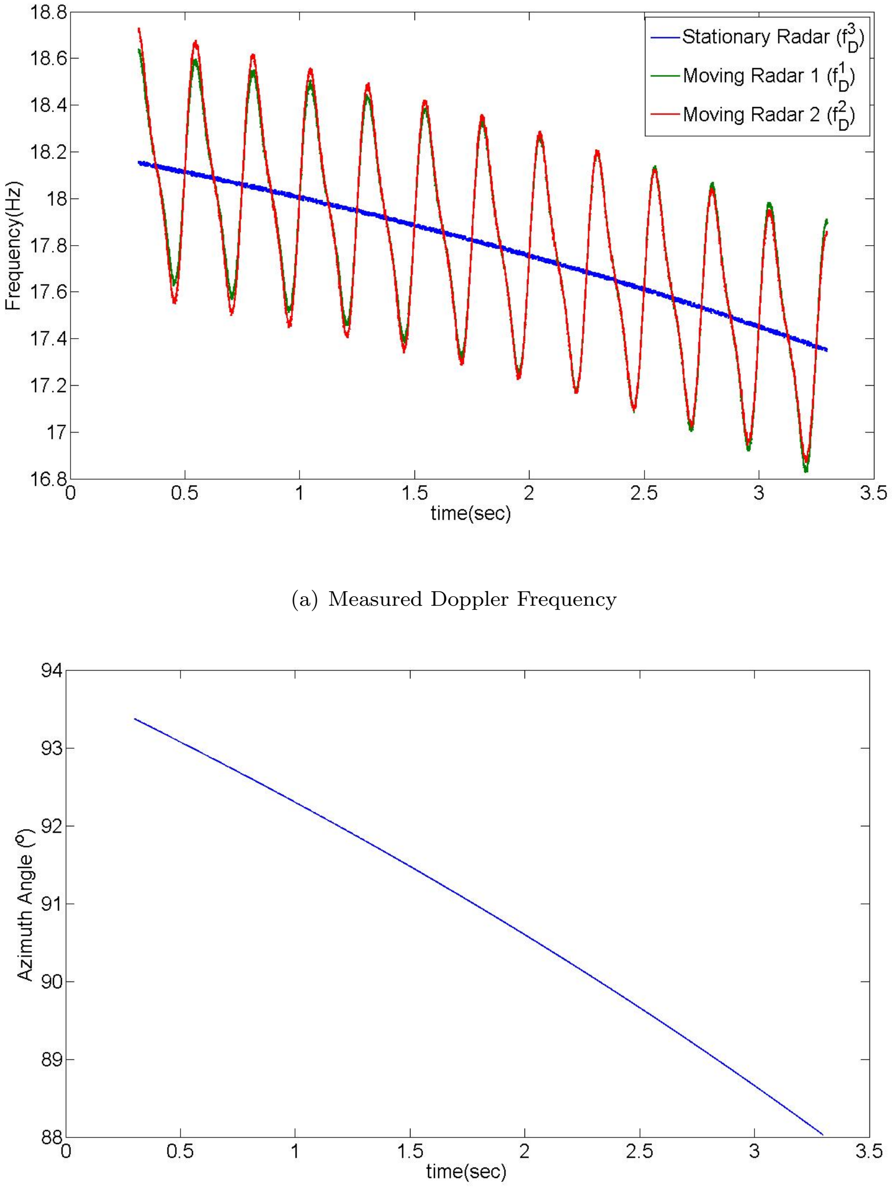
A target with constant velocity of $\approx$ 19 m/s at a range of $\approx$ 300 m was generated. The target observations were then generated for the Vidar. Sensor noise was taken into a account.

In Figure 5.1(a), the Doppler measured by all three Radars is shown while Figure 5.1(b) shows the azimuth angle as seen in the video.

The generated data was then used as an input for both Radar only Tracking as well as Sensor Fusion Tracking.

We can see that the estimate of the velocity (Figure 5.2(a)) is fairly good but the azimuth angle (Figure 5.2(b)) is very noisy.

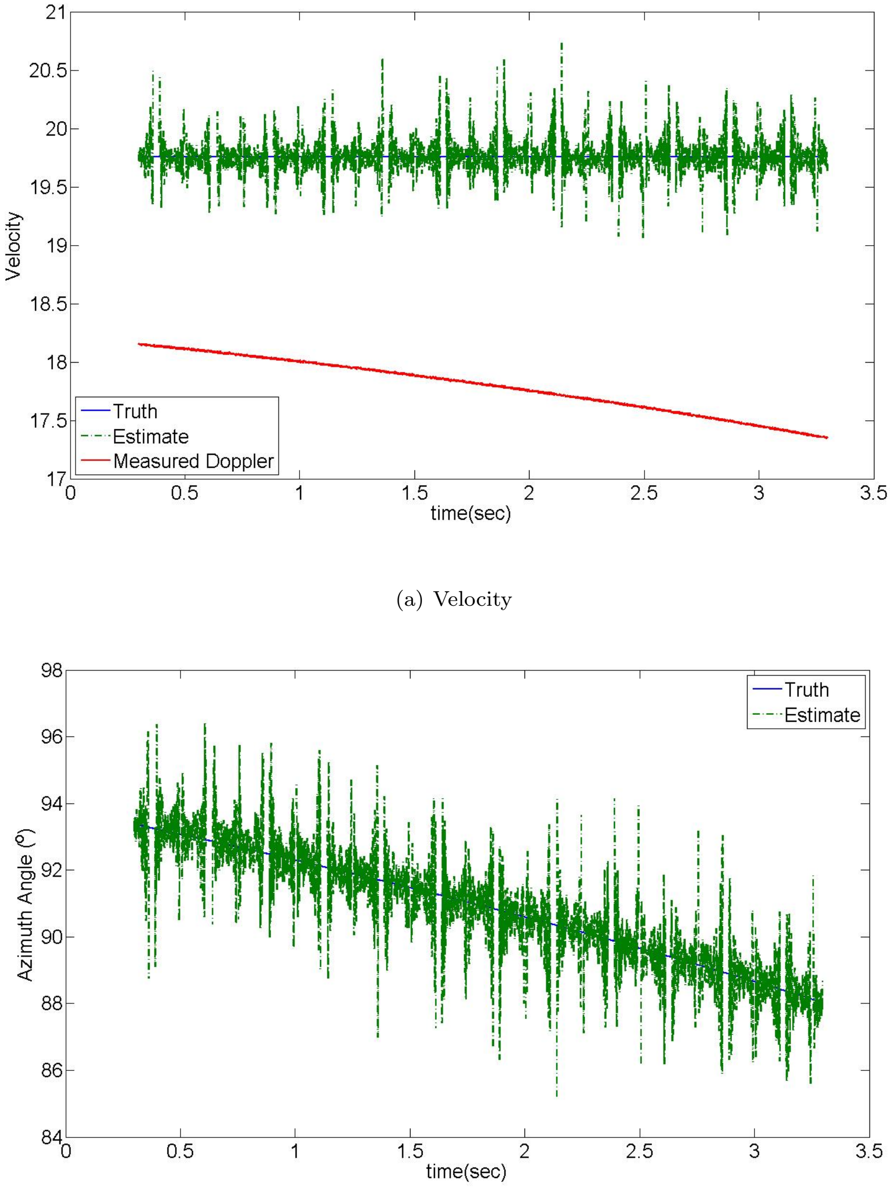
The error in velocity (Figure 5.3(a)) was observed to be within 0.5 m/s while the azimuth angle error (Figure 5.3(b)) was sometimes greater than 4°.



(a) Measured Doppler Frequency

(b) Measured Azimuth Angle

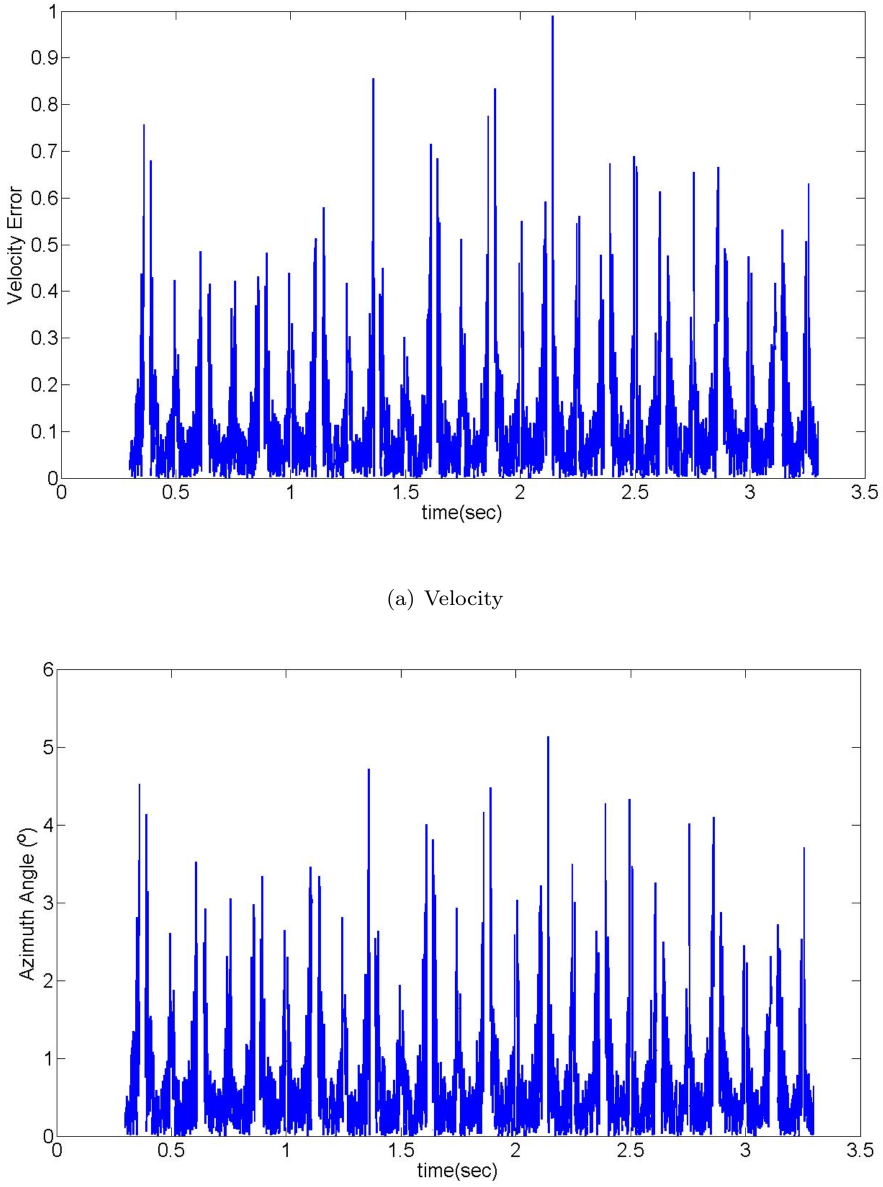
Figure 5.1: Simulation data



(a) Velocity

(b) Azimuth Angle

Figure 5.2: Radar only estimate



(a) Velocity

(b) Azimuth Angle

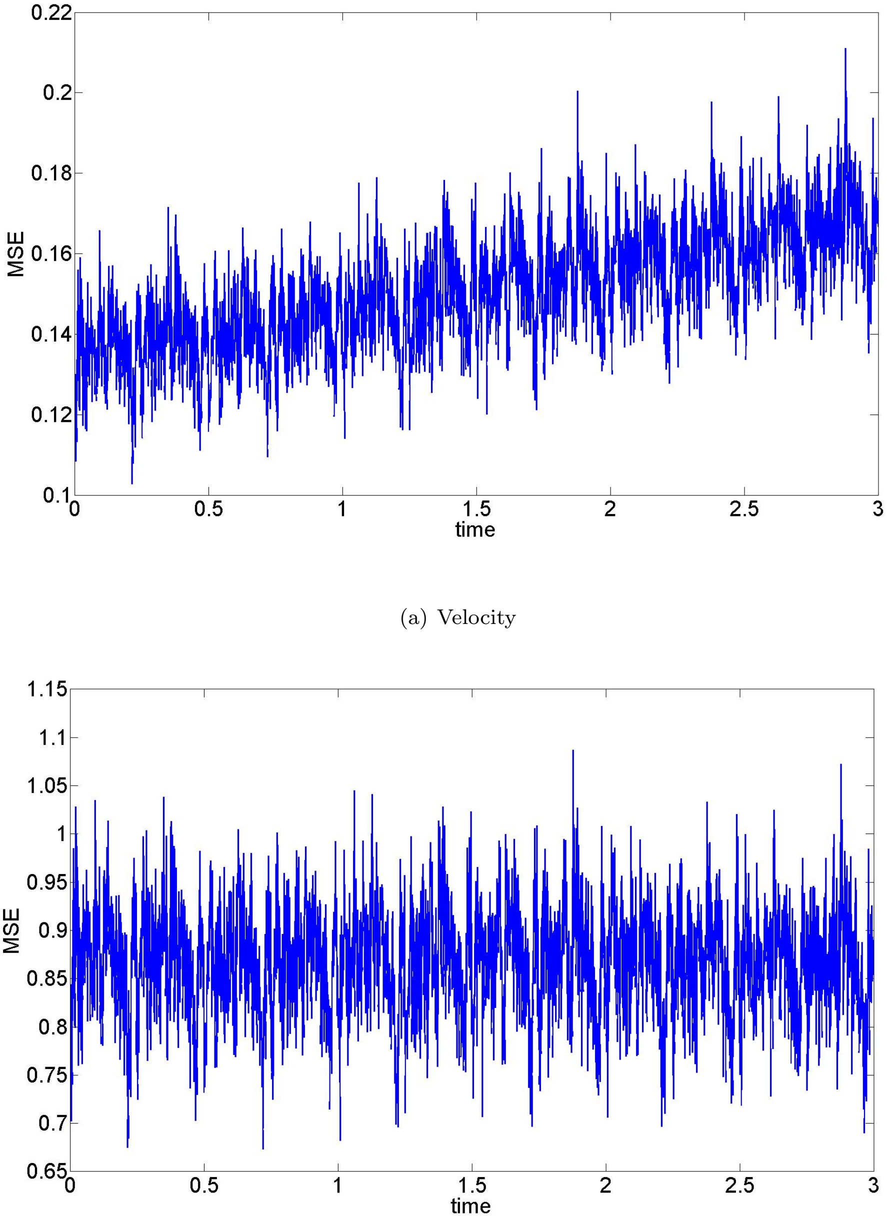
Figure 5.3: Radar only estimate error

Figure 5.4(a) and 5.4(b) show the mean square error for 500 Monte Carlo runs of the simulation.

The effect of using the additional information from the video is obvious right away. The velocity estimates are better (Figure 5.5(a)) but the major improvement comes in the estimate of the azimuth angle (Figure 5.5(b)).

The advantage of using sensor fusion becomes even more apparent when looking at the errors in velocity (Figure 5.6(a)) and azimuth (Figure 5.6(b)).

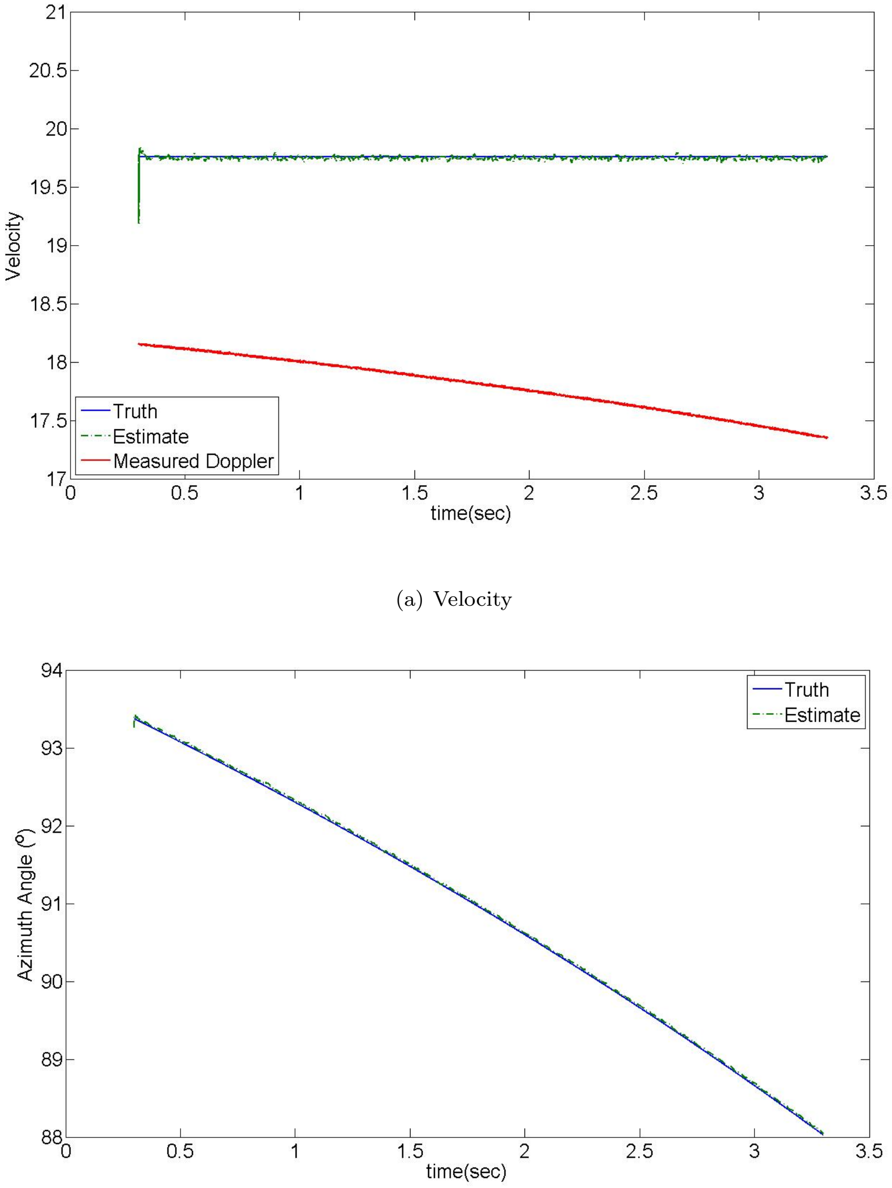
Figure 5.7(a) and 5.7(b) show the mean square error for 500 Monte Carlo runs of the simulation.



(a) Velocity

(b) Azimuth Angle

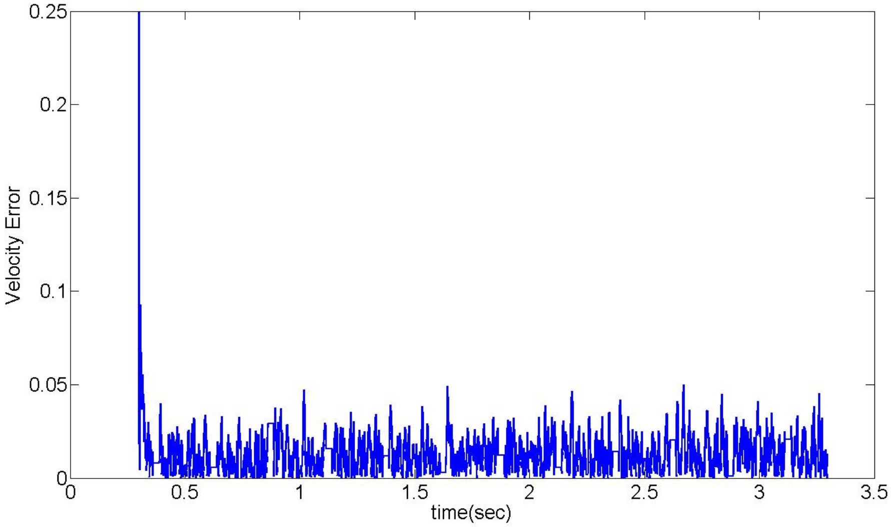
Figure 5.4: Radar only Monte Carlo error



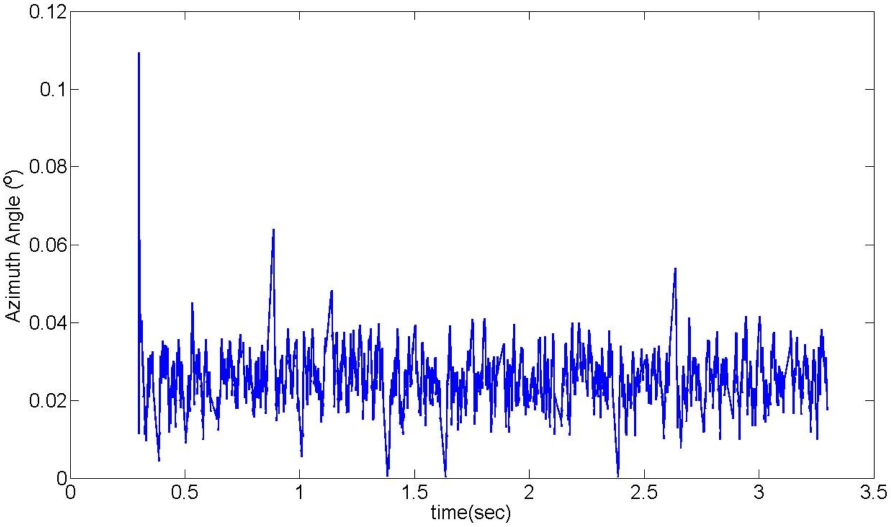
(a) Velocity

(b) Azimuth Angle

Figure 5.5: Sensor fusion estimate

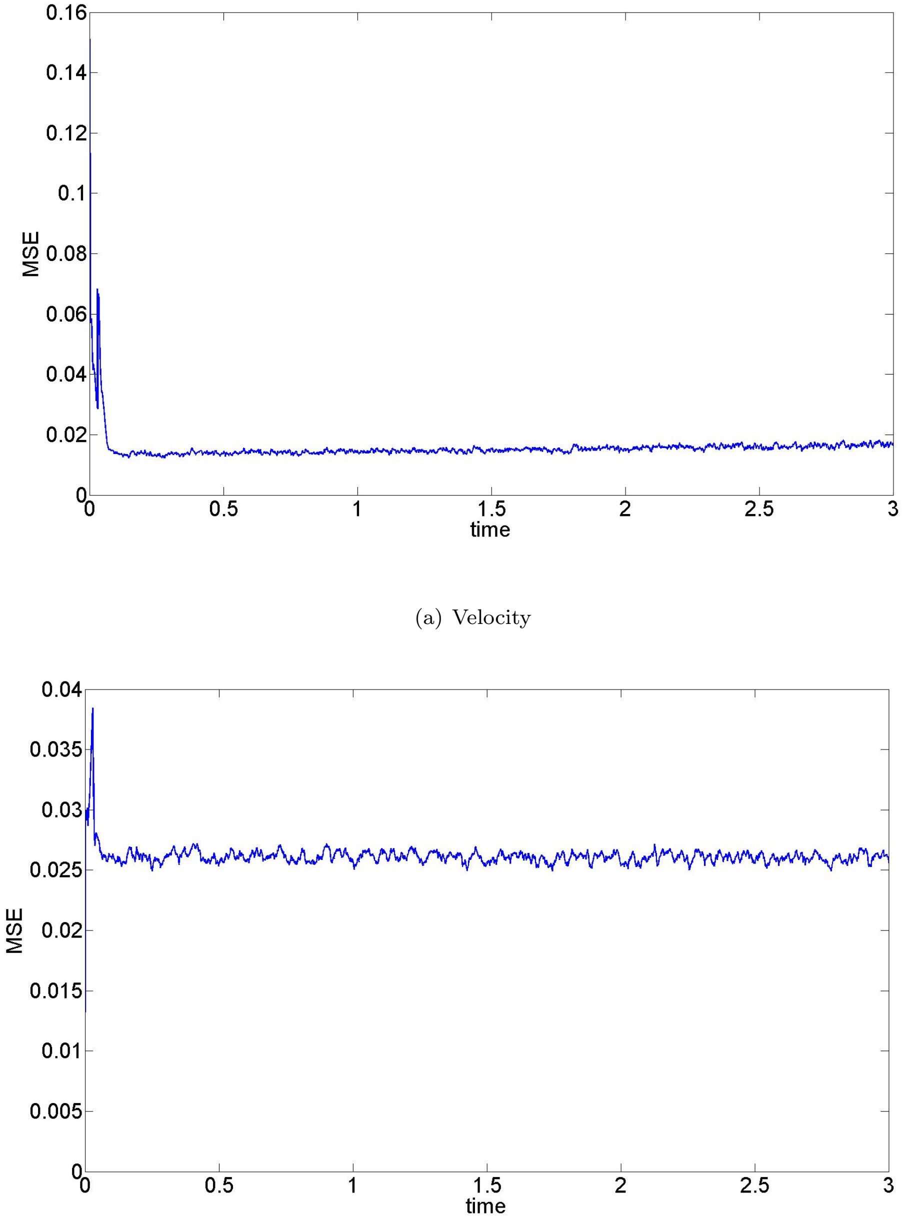


(a) Velocity



(b) Azimuth Angle

Figure 5.6: Sensor fusion estimate error



(a) Velocity

(b) Azimuth Angle

Figure 5.7: Sensor fusion Monte Carlo Error

5.2 Ground Truth

In order to validate the estimates from the Vidar, we collected data at two sites (Figure 5.8). To validate the velocity estimates, we collected ground truth data for the velocity by using a GPS device on board a target vehicle that was observed by the Vidar. The GPS records speed measurements with an precision of 1 mph. This measurement was used as a validation tool for the Vidar.

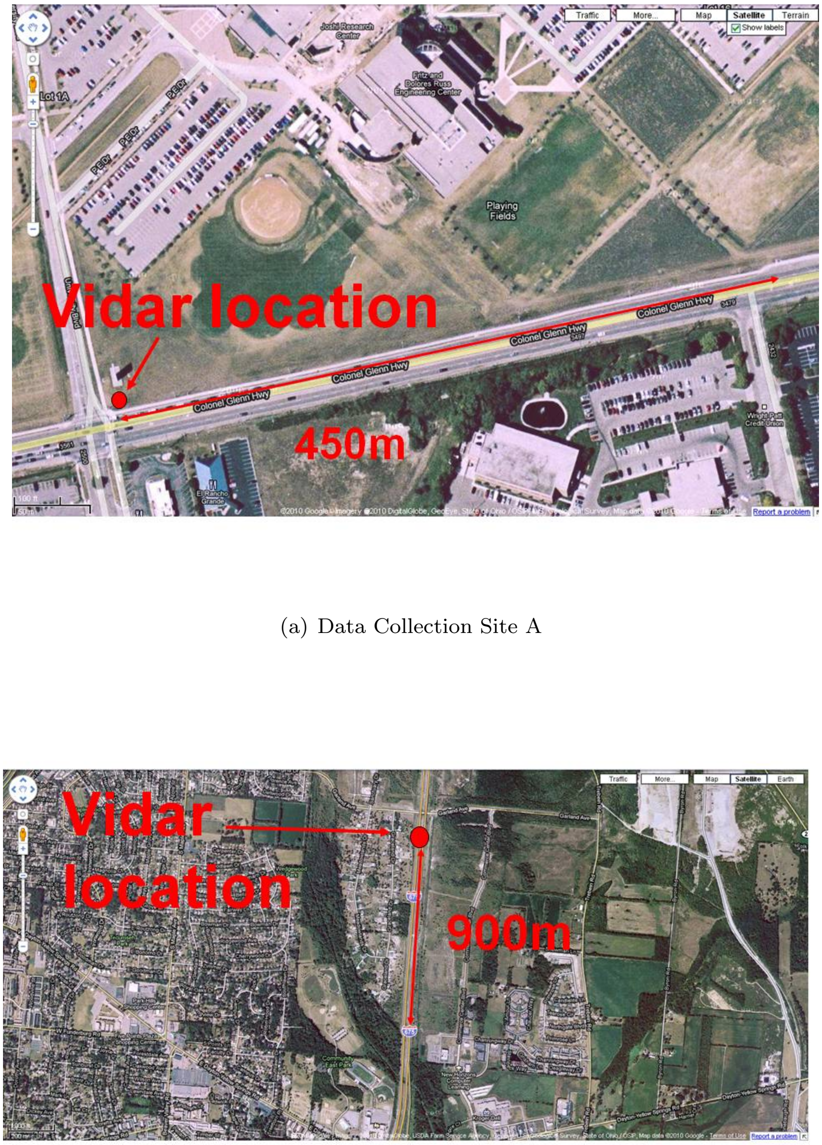
Validating the azimuth angle estimates was a more time-consuming task. After the data collection, the tracked vehicles were manually labelled. The labelling provided the image plane coordinates which were converted to azimuth angle. The ground truth azimuth angles were then compared to the azimuth angle estimated by the sensor fusion process.

Figure 5.10 shows the comparison of ground truth velocity with the estimate using Radar Only while Figure 5.12 shows the comparison of ground truth velocity with the estimate using sensor fusion. Figure 5.13 shows the azimuth angle for the same data set.

Figure 5.15(a) shows the comparison of velocity with the estimate using Radar Only and Figure 5.15(b) shows the comparison of ground truth azimuth with the estimate using Radar Only. Figure 5.17(a) shows the comparison of velocity with the estimate using Radar Only and Figure 5.17(b) shows the comparison of ground truth azimuth with the estimate using sensor fusion.

5.3 Multi-Target Tracking

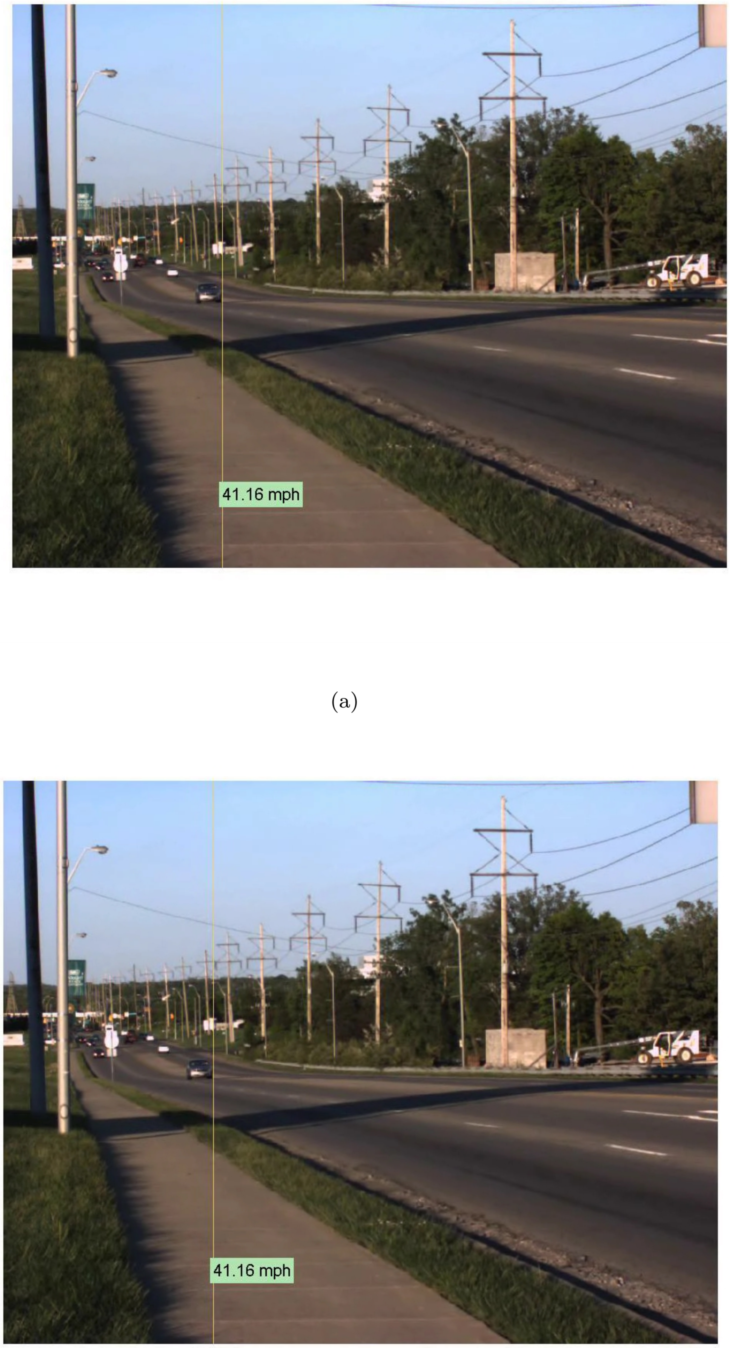
The alorithm was tested for its performance on multiple targets for data association and track accuracy. Figure 5.19(a) and 5.19(b) shows the performance of the radar only algorithm while Figure 5.21(a) and 5.21(b) shows the performance of the sensor fusion algorithm.



(a) Data Collection Site A

(b) Data Collection Site B

Figure 5.8: Data collection sites for the Vidar



(a)

(b)

Figure 5.9: Radar only snapshots(Site A

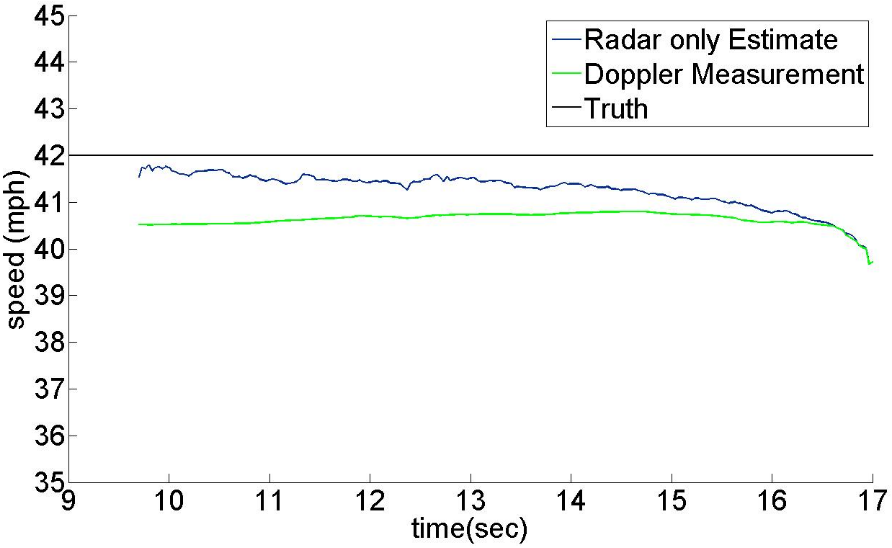
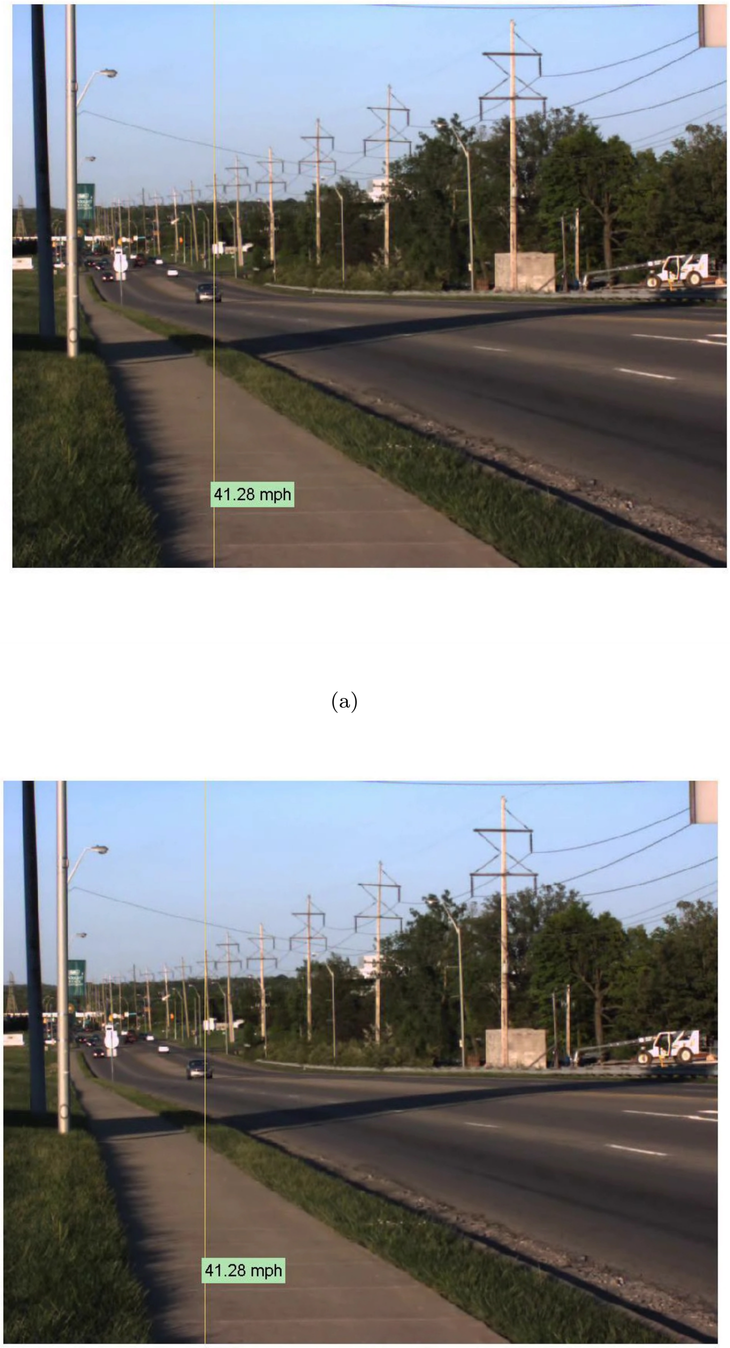


Figure 5.10: Radar only velocity ground truth (Site A)



(a)

(b)

Figure 5.11: Sensor fusion Snapshot(Site A)

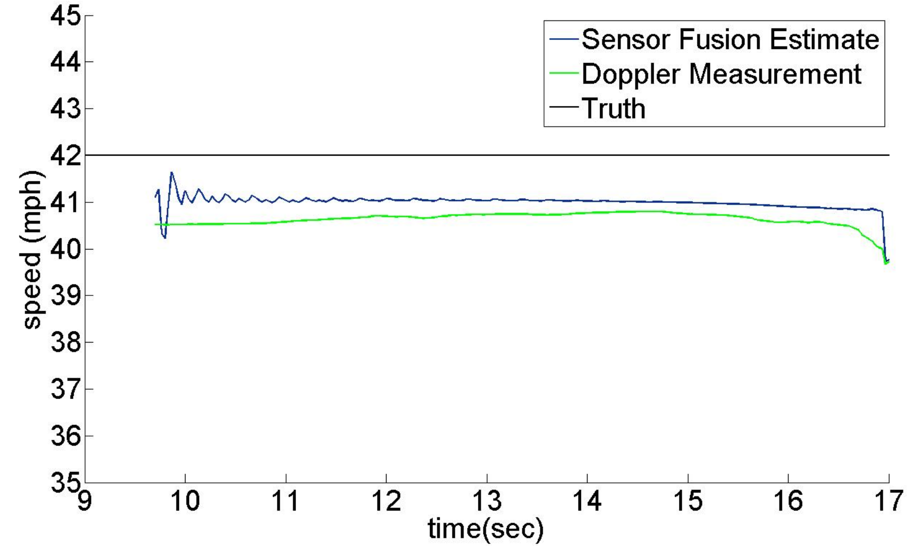


Figure 5.12: Sensor fusion velocity ground truth (Site A)

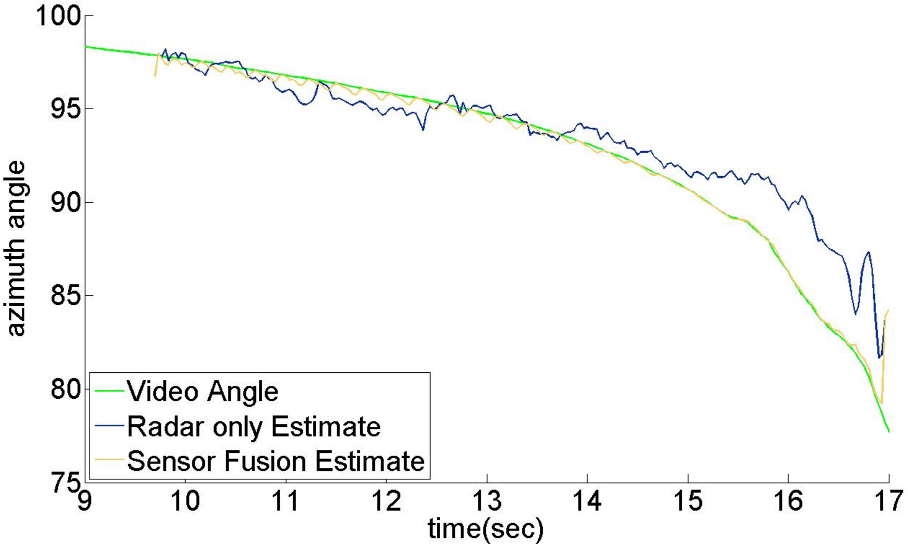
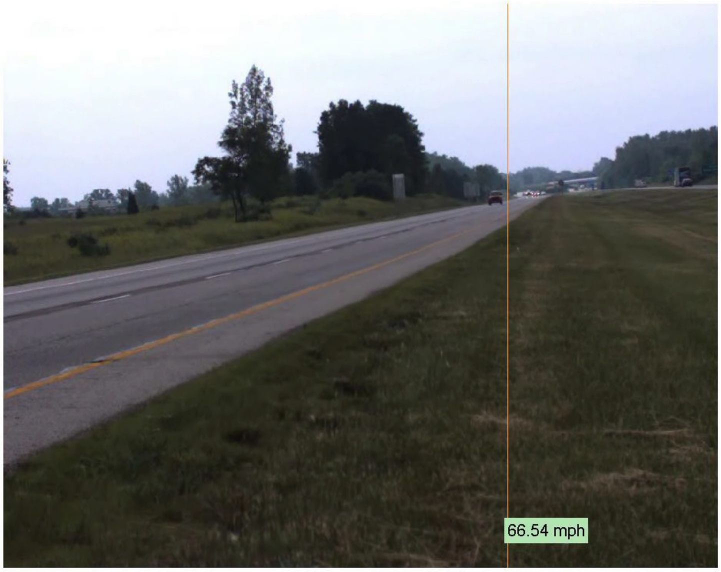
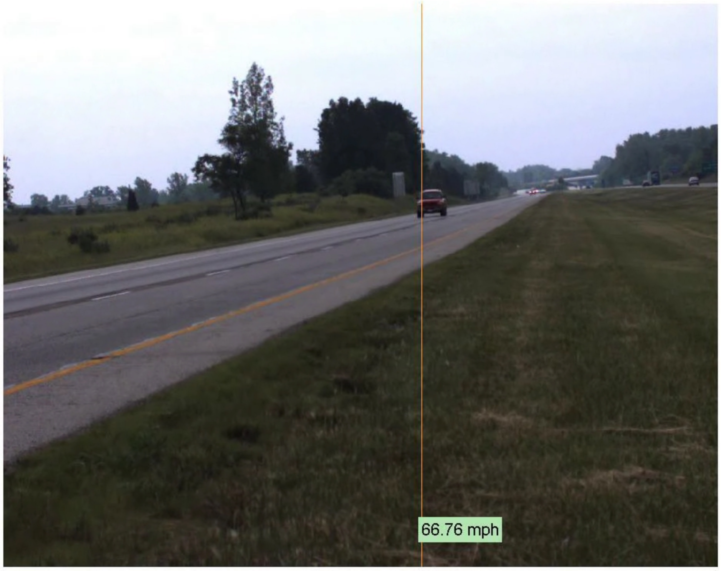


Figure 5.13: Radar only azimuth(Site A)

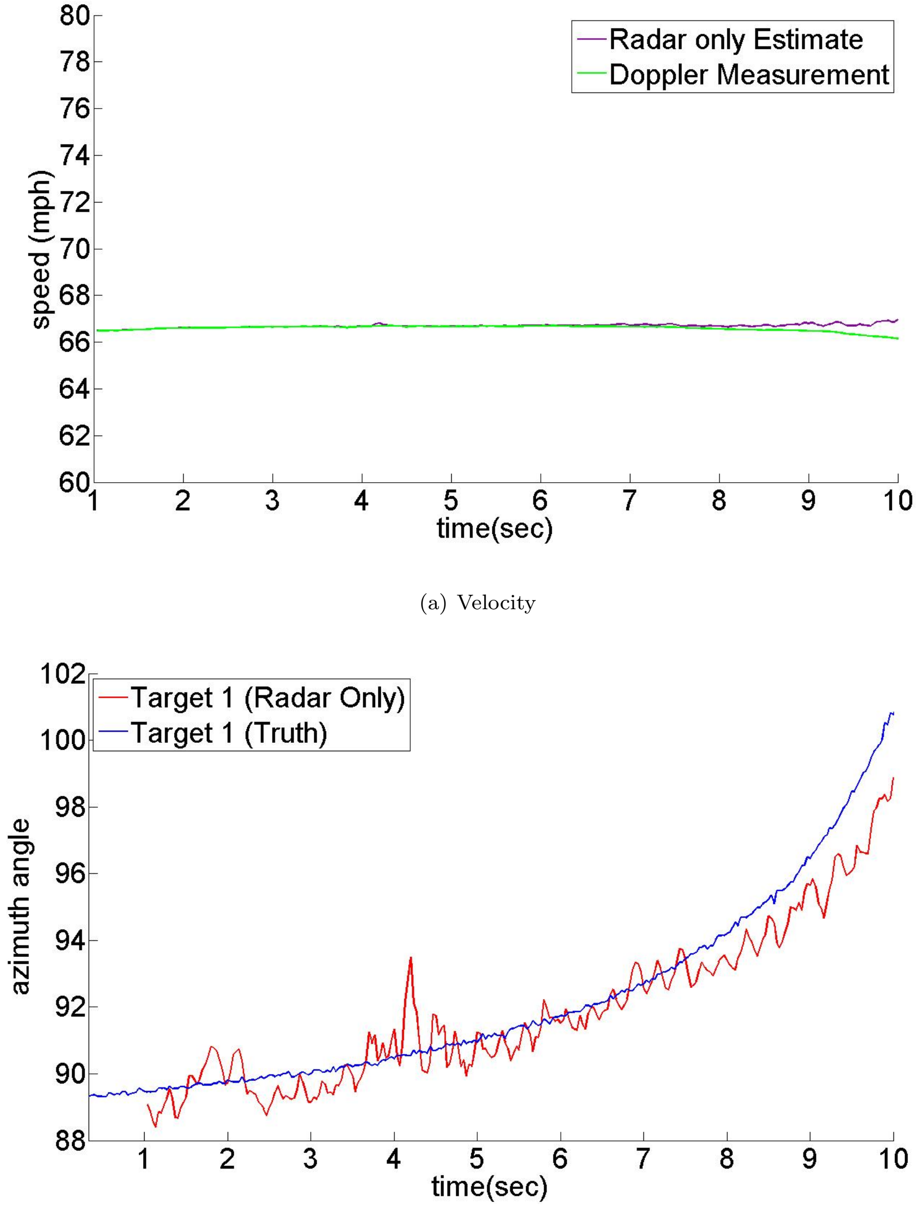


(a)



(b)

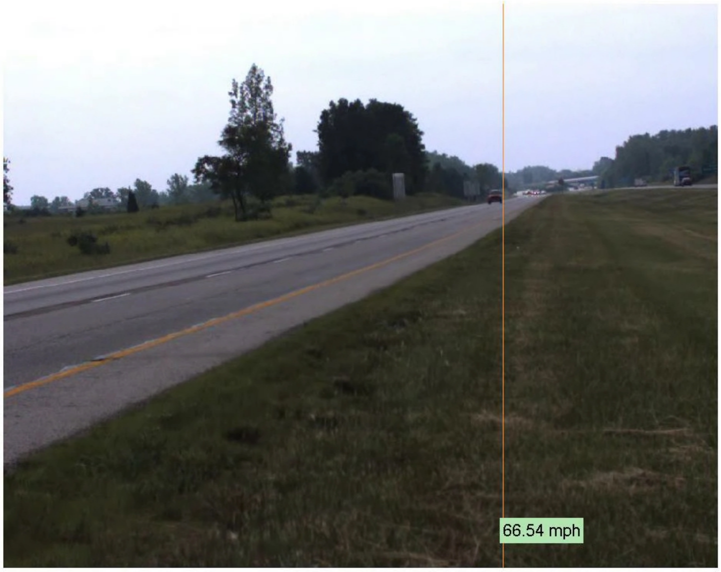
Figure 5.14: Radar only snapshots(Site B)



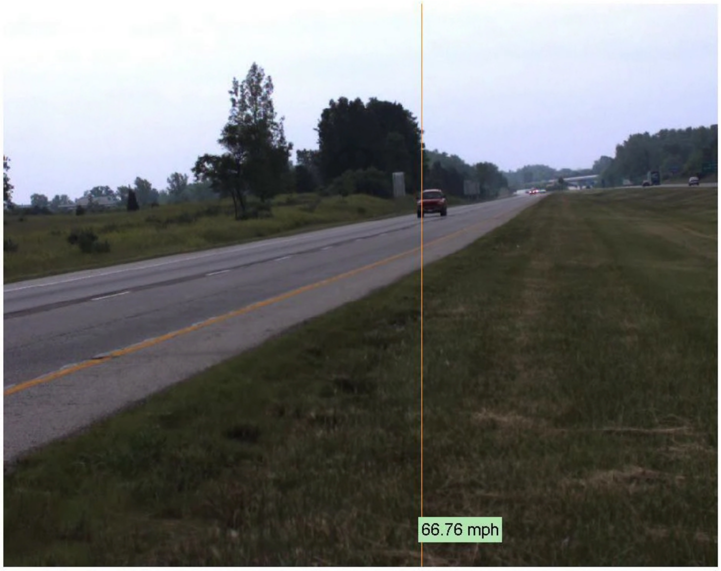
(a) Velocity

(b) Azimuth Angle

Figure 5.15: Radar only ground truth (Site B)

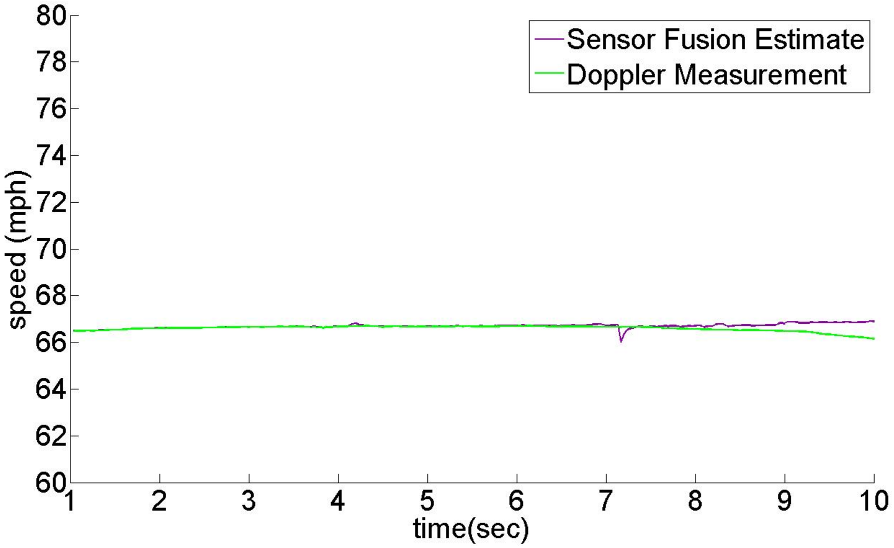


(a)

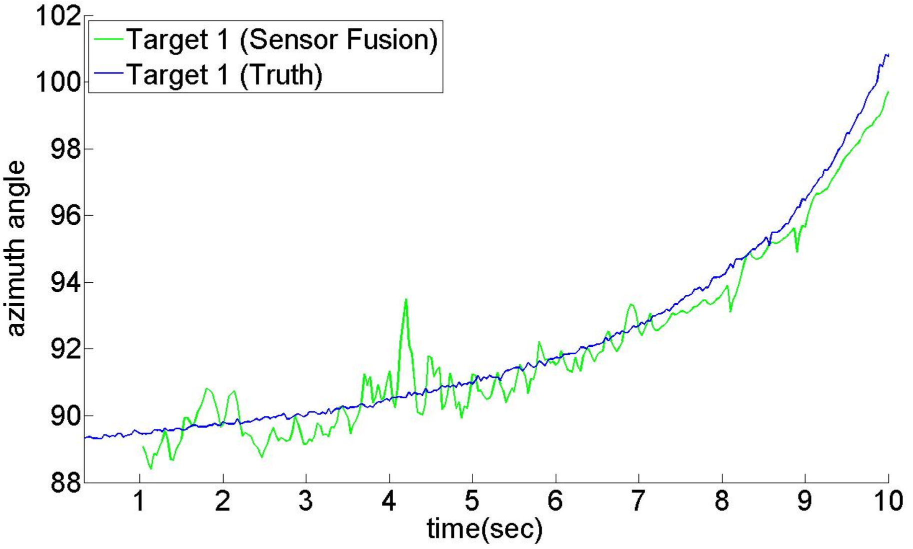


(b)

Figure 5.16: Sensor fusion snapshots(Site B)

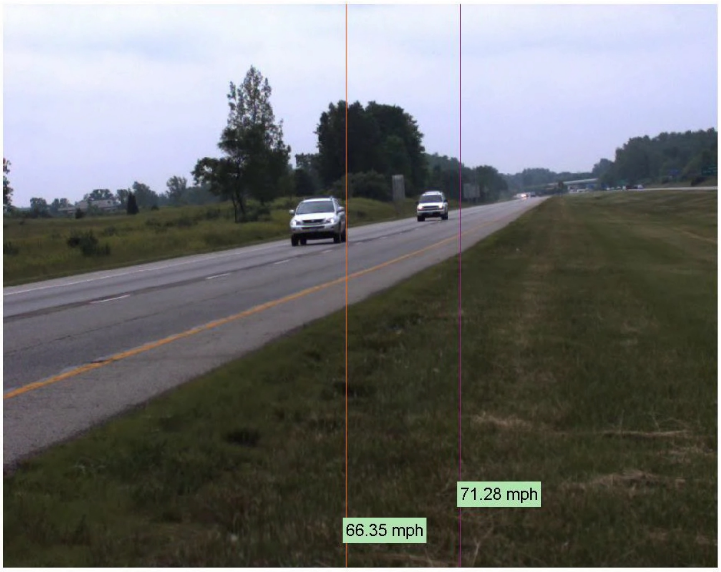


(a) Velocity

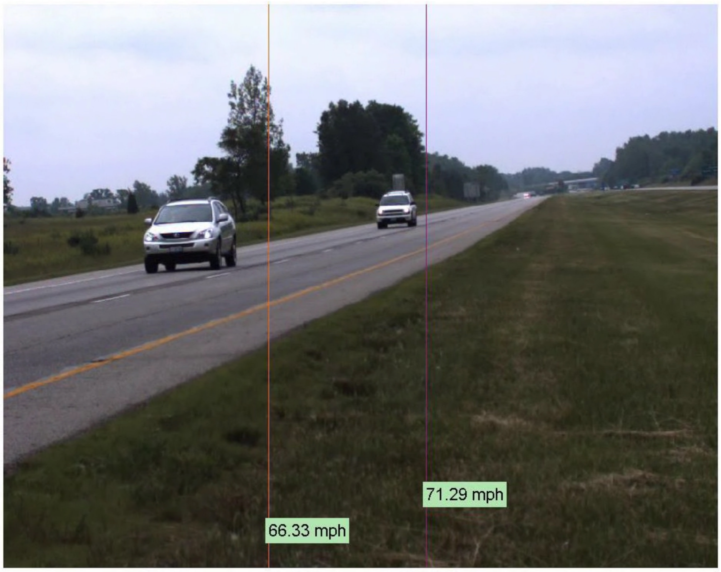


(b) Azimuth Angle

Figure 5.17: Sensor fusion ground truth (Site B)

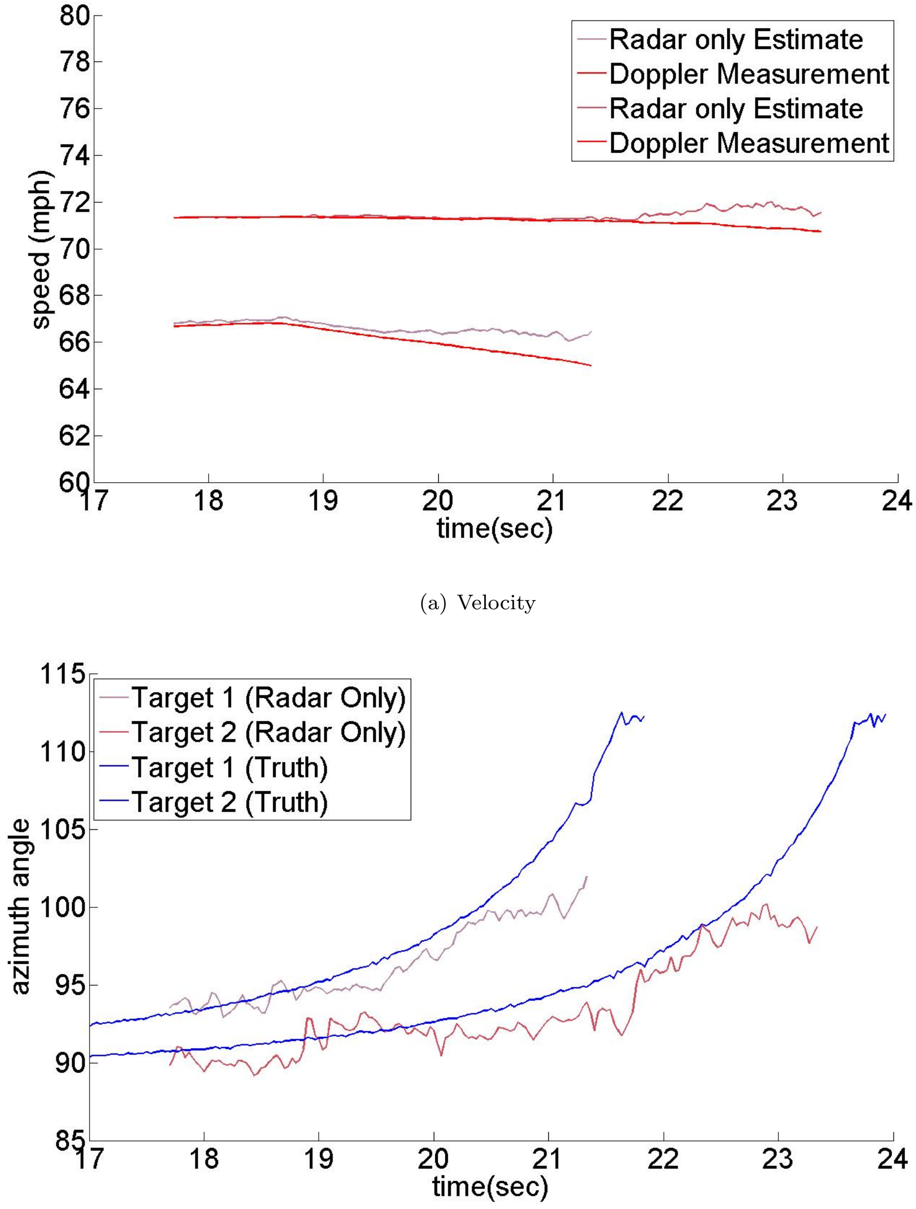


(a)



(b)

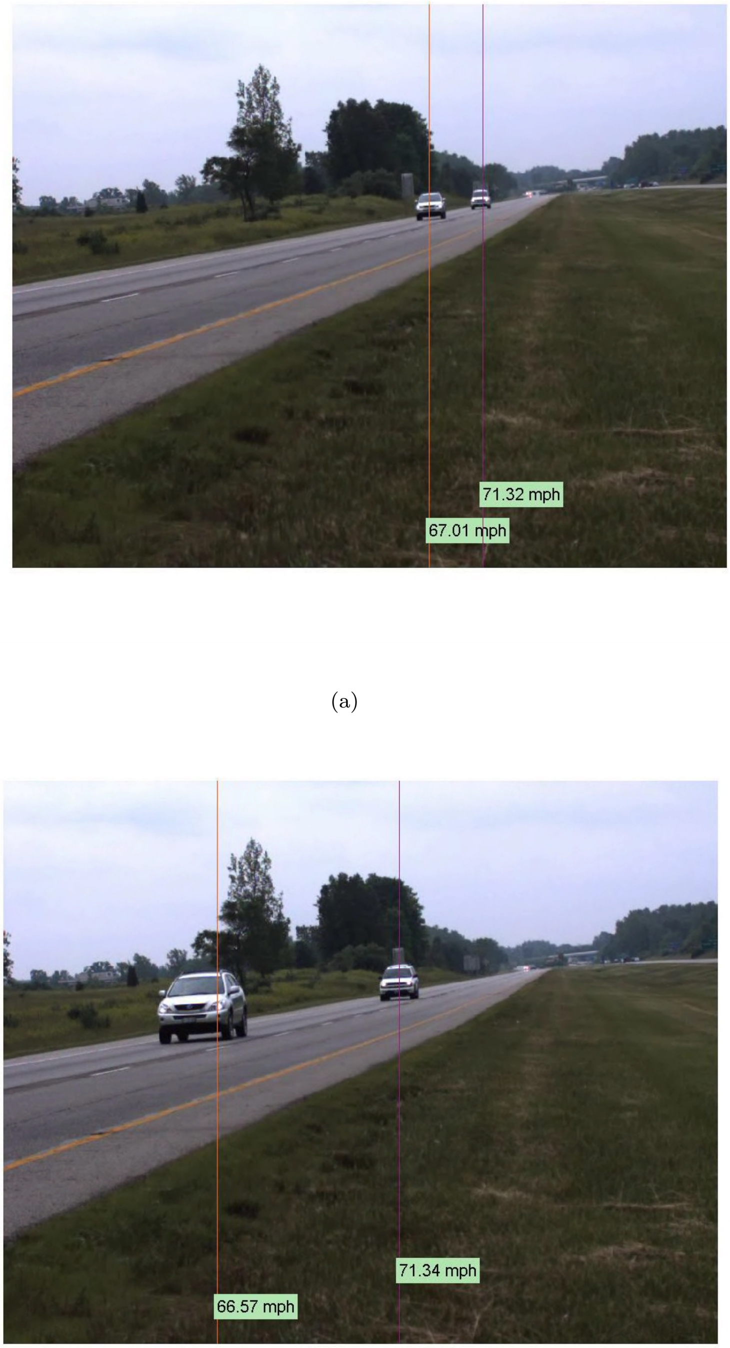
Figure 5.18: Radar only multi-target tracking Snapshots(Site B)



(a) Velocity

(b) Azimuth Angle

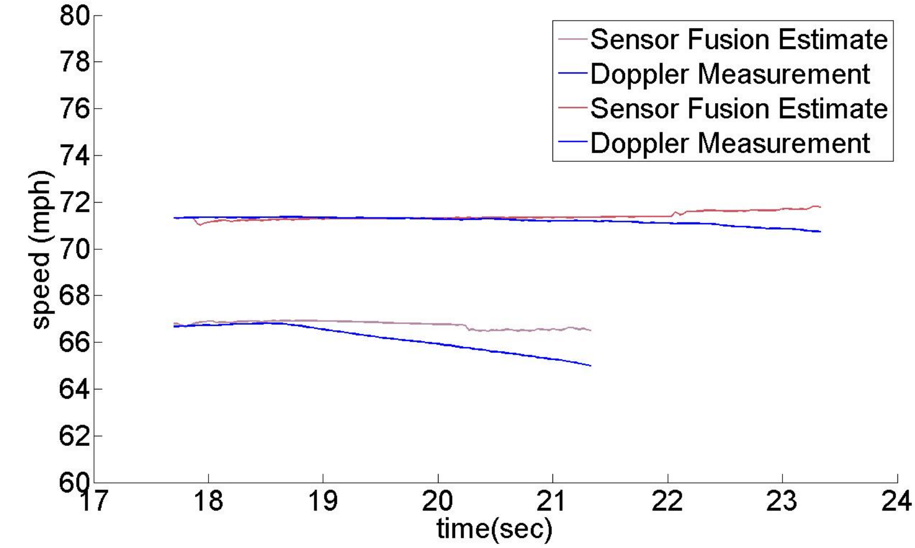
Figure 5.19: Radar only multi-target tracking (Site B)



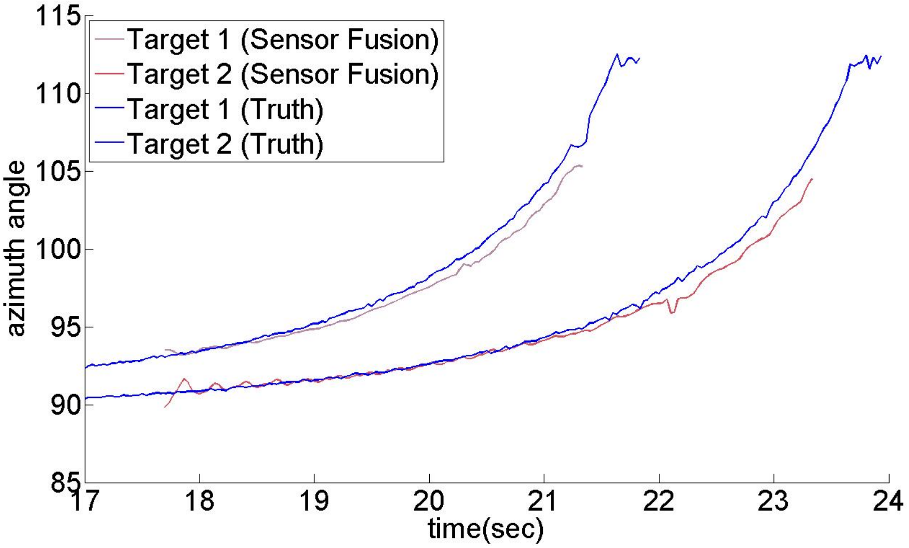
(a)

(b)

Figure 5.20: Sensor fusion multi-target tracking Snapshots(Site B)



(a) Velocity



(b) Azimuth Angle

Figure 5.21: Sensor fusion multi-target tracking (Site B)

6

Conclusion

Summary

In this work, we addressed the limitation of traditional CW Doppler radars that allows them to track only the strongest or fastest target with in the field of view. By developing sensor fusion algorithms, we were able to track multiple targets that are visible in Doppler frequency as well as by the camera.

Two different sensor suites were created and developed which were able to perform the task with improved accuracy. The systems were able to report the target velocity as well as angles which is a limitation faced by CW Doppler radars.

Sensor fusion algorithms for a Vidar traffic surveillance system were developed that match video signals to stationary Doppler radar signals and fuse the matched video and Doppler radar signals to generate moving vehicle velocity and range information.

We also applied zoom FFT techniques in conjunction with reassignment to analyze time-varying frequency signals from the CW Doppler radars. The zoom FFT allows us to process much higher resolution without the overhead of higher computational load.

Contributions

- A method of fusing video and Doppler radar information for estimating moving vehicle velocity and angle information, comprising of the steps of:
  - matching video angles to Doppler radar angles
  - fusing the matched video and radar to derive velocity and angle information of the target vehicle.
- An investigation into the use of reassignment for time-varying signals from CW Doppler radars.
- The use of the zoom FFT techniques combined with reassignment to reduce the computational load for analyzing time-varying signals from CW Doppler radars.
- A multi-sensor hardware prototype was designed with high-resolution video camera, CW Doppler radars, mechanisms for inducing motion, data processing and data recording all embedded on-board.

References

F. L. and P. David, “Improving on police radar,” *IEEE Spectrum*, vol. 29, no. 7, pp. 38–43, 1992.

“Police traffic radar handbook: A comprehensive guide to speed measuring systems,” http://www.copradar.com (accessed on July 12, 2010).

D. Middleton, D. Gopalakrishna, and M. Raman, “Advances in traffic data collection and management,” January 2003, http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13766.html (accessed on Oct 25, 2007).

J. Malik and S. Russell, “Traffic surveillance and detection technology development: New traffic sensor technology final report,” *California PATH Research Report UCB-ITS-PRR-97-6*, January 1997.

D. Dailey and L. Li, “An algorithm to estimate vehicle speed using uncalibrated cameras,” *Intelligent Transportation Systems, Proceedings of IEEE/IEEJ/JSAI International Conference on*, pp. 441 – 446, 1999.

N. K. Kanhere, S. J. Pundlik, and S. T. Birchfield, “Vehicle segmentation and tracking from a low-angle off-axis camera,” in *CVPR ’05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, Washington, DC, USA, 2005, pp. 1152–1157.

A. Gern, U. Franke, and P. Levi, "Advanced lane recognition-fusing vision and radar," *Proc. IEEE Intell. Vehicles Symp.*, pp. 45–51, 2000.

B. Steux, C. Laurgeau, L. Salesse, and D. Wautier, "Fade: A vehicle detection and tracking system featuring monocular color vision and radar data fusion," *Proc. IEEE Intell. Vehicles Symp*, pp. 632–639, 2002.

A. Gern, U. Franke, and P. Levi, "Robust vehicle tracking fusing radar and vision," *Proc. Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, pp. 323–328, 2001.

N. Kawasaki and U. Kiencke, "Standard platform for sensor fusion on advanced driver assistance system using bayesian network," *Proc.IEEE Intell. Vehicles Symp.*, pp. 250–255, 2004.

A. Solea, O. Mano, G. Stein, H. Kumon, Y. Tamatsu, and A. Shashua, "Solid or not solid:vision for radar target validation," *Proc. IEEE Intell. Vehicles Symp.*, pp. 819–824, 2004.

R. Grover, G. Brooker, and H. F. Durrant-Whyte, "A low level fusion of millimeter wave radar and night-vision imaging for enhanced characterization of a cluttered environment," *in Proc. Aust. Conf. Robot. Autom.*, pp. 73–80, 2001.

S. Wu, S. Decker, P. Chang, T. Camus, and J. Eledath, "Collision sensing by stereo vision and radar sensor fusion," *Trans. Intell. Transport. Sys.*, vol. 10, pp. 606–614, 2009.

R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME Journal of Basic Engineering*, pp. 35 – 45, 1960.

G. Welch and G. Bishop, "An introduction to the kalman filter," university of North Carolina at Chapel Hill, Chapel Hill, NC, 1995.

M. Grewal and A. Andrews, *Kalman Filtering: Theory and Practice using MATLAB*, 2nd ed. Wiley, 2001.

S. Blackman, *Multiple Target Tracking with Radar Application.* Artech House, 1986.

S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems.* Artech House, 1999.

Y. Bar-Shalom, *Multi-Target-MultiSensor Tracking: Applications and Advances Volume II.* Artech House, 1992.

L. Stone, C. Barlow, and T. Corwin, *Bayesian Multiple Target Tracking.* Artech House, 1999.

M. Liggins, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. CRC Press, 2009.

D. Hall and S. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, 2nd ed. Artech House, 2004.

M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision*, 1st ed. Springer, 2004.

D. A. Forsyth and J. Ponce, *Computer Vision-A Modern Approach.* Prentice Hall, 2002.

B. Horn, *Robot Vision*, 1st ed. MIT Press, 1986.

Z. Zhang, “A flexible new technique for camera calibration,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, 2000, pp. 1330–1334.

S. C. Cheung and C. Kamath, “Robust techniques for background subtraction in urban traffic video,” in *Proceedings of Electronic Imaging: Visual Communications and Image Processing*, 2004.

"Intel opencv computer vision library (c++)," http://www.intel.com/research/mrl/research/opencv/.

G. Bradski and A. Kaehler, *Learning OpenCV.* OŔeilly Media, September 2008.

A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," *Image and Vision Computing*, vol. 24, no. 11, pp. 1233 – 1243, 2006.

"Decatur radar," http://www.decaturradar.com (accessed on July 12, 2010).

M. Solnik, *Introduction to Radar Systems*, 3rd ed. McGraw-Hill, 2001.

B. Mahafza, *Radar System Design and Analysis using MATLAB*, 2nd ed. Chapman and Hall, 2005.

K. Kodera, R. Gendrin, and C. Villedary, "Analysis of time-varying signals with small bt values," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, pp. 64–76, Feb 1978.

F. Auger and P. Flandrin, "Improving the readability of time frequency and time scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1068–1089, 1995.

F. Auger, *Time Frequency Reassignment: From principles to algorithms.* Boca Raton FL: CRC Press, 2002.

S. Hainsworth and M. Macleod, "Time frequency reassignment: A review and analysis," 2003, technical Report.

K. Fitz and S. Fulop, "A unified theory of time-frequency reassignment," *Digital Signal Processing*, 2005.

R. G. Lyons, *Understanding Digital Signal Processing*, 2nd ed. Prentice Hall, 2004.

A. Roy, N. Gale, and L. Hong, “Fusion of doppler radar and video information for automated traffic surveillance,” in *Proceedings of 12th International Conference on Information Fusion*, July 2009, pp. 1989–1996.

J.-Y. Bouguet, “Camera calibration toolbox for matlab,” http://www.vision.caltech.edu/bouguetj/calib_doc/.

L. Hong *et al.*, “A 3d video-doppler-radar (vidar) imaging system,” US Patent 200 812 255 081, Oct, 2008.

——, “Video-doppler-radar traffic surveillance system,” US Patent 200 812 266 227, Nov, 2008.

——, “Fusion algorithm for vidar traffic surveillance system,” US Patent 200 812 333 735, Dec, 2008.