Wright State University CORE Scholar

Browse all Theses and Dissertations

Theses and Dissertations

2012

# Frequency Modulated Continuous Wave Radar and Video Fusion for Simultaneous Localization and Mapping

Andrew Stephen Kondrath Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd\_all

Part of the Engineering Commons

#### **Repository Citation**

Kondrath, Andrew Stephen, "Frequency Modulated Continuous Wave Radar and Video Fusion for Simultaneous Localization and Mapping" (2012). *Browse all Theses and Dissertations*. 1099. https://corescholar.libraries.wright.edu/etd\_all/1099

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

# FREQUENCY MODULATED CONTINUOUS WAVE RADAR AND VIDEO FUSION FOR SIMULTANEOUS LOCALIZATION AND MAPPING

A dissertation submitted in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy** 

By

Andrew Kondrath

B. S., Wright State University, Dayton, Ohio, 2004

M. S., Wright State University, Dayton, Ohio, 2006

2012 Wright State University

#### WRIGHT STATE UNIVERSITY

#### SCHOOL OF GRADUATE STUDIES

August 09, 2012

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Andrew S. Kondrath ENTITLED Frequency Modulated Continuous Wave Radar and Video Fusion for Simultaneous Localization and Mapping 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 Program

Committee on Final Examination

Andrew Hsu, Ph.D. Dean of Graduate Studies

Lang Hong, Ph.D

Kefu Xue, Ph.D

Fred Garber, Ph.D

Arthur Goshtasby, Ph.D.

Richard Martin, Ph.D.

#### Abstract

Kondrath, Andrew. Ph.D., Engineering Ph.D. Program, Wright State University, 2012. Frequency Modulated Continuous Wave Radar and Video Fusion for Simultaneous Localization and Mapping.

There has been a push recently to develop technology to enable the use of UAVs in GPS-denied environments. As UAVs become smaller, there is a need to reduce the number and sizes of sensor systems on board. A video camera on a UAV can serve multiple purposes. It can return imagery for processing by human users. The highly accurate bearing information provided by video makes it a useful tool to be incorporated into a navigation and tracking system. Radars can provide information about the types of objects in a scene and can operate in adverse weather conditions. The range and velocity measurements provided by the radar make it a good tool for navigation.

FMCW radar and color video were fused to perform SLAM in an outdoor environment. A radar SLAM solution provided the basis for the fusion. Correlations between radar returns were used to estimate dead-reckoning parameters to obtain an estimate of the platform location. A new constraint was added in the radar detection process to prevent detecting poorly observable reflectors while maintaining a large number of measurements on highly observable reflectors. The radar measurements were mapped as landmarks, further improving the platform location estimates. As images were received from the video camera, changes in platform orientation were estimated, further improving the platform orientation estimates. The expected locations of radar measurements, whose uncertainty was modeled as Gaussian, were projected onto the images and used to estimate the location of the radar reflector in the image. The colors of the most likely reflector were saved and used to detect the reflector in subsequent images. The azimuth angles obtained from the image detections were used to improve the estimates of the landmarks in the SLAM map over previous estimates where only the radar was used.

# Contents

1	Intr	roduction	1
	1.1	Motivation	1
	1.2	Sensors	2
	1.3	Sensor Fusion	4
	1.4	Problem Statement	7
	1.5	Approach	8
		1.5.1 Contributions	10
	1.6	Outline	10
<b>2</b>	Bac	kground	11
	2.1	SLAM	11
	2.2	Radar SLAM	12
	2.3	Video SLAM	15
	2.4	Fusion Methods	17
3	Sys	tem	20
	3.1	Radar	20
	3.2	Video Camera	23
	3.3	Platform and Landmark Descriptions	25
		3.3.1 State Variables	25
		3.3.2 Motion Models	27
	3.4	Radar and Video Calibration	28
		3.4.1 Radar Calibration	28
4	Rac	lar SLAM	30
	4.1	Filters	30
		4.1.1 Kalman Filter	30

	4.1.2	Unscented Kalman Filter	31	
	4.1.3	Linear Regression Kalman Filter	35	
	4.1.4	Consistency Test	35	
4.2	Orient	tation Change Estimation	37	
	4.2.1	Frequency Bins of Likely Reflectors	38	
	4.2.2	Orientation Estimation	39	
4.3	Veloci	ty Estimation	39	
	4.3.1	Down-Chirp Frequency Relation to Up-Chirp Frequency	39	
	4.3.2	Range Over Time	40	
	4.3.3	Down-Chirp Relation to Up-Chirp with Range Change $\ldots$ .	41	
	4.3.4	Magnitude and Angle of Arrival Response Descriptions $\ . \ . \ .$	41	
	4.3.5	Estimating Velocity	41	
	4.3.6	Alternative Velocity Estimation	42	
4.4	State	Propagation With Radar Dead-Reckoning	43	
4.5	Landr	nark Detection	43	
	4.5.1	Quality Landmarks	43	
	4.5.2	Cell-Averaging Constant False Alarm Rate Detector	44	
	4.5.3	AoA Smoothness Constraint	46	
	4.5.4	Possible Detections from Single Chirp	47	
	4.5.5	Detections from Adjacent Chirps	47	
4.6	Radar	Measurement Model	48	
4.7	Angle	Estimation	48	
4.8	Range	e Estimation	49	
4.9	Unma	tched Detection Association with Unmatched Landmarks	49	
4.10	Filter	Update	49	
4.11	.11 New Landmark Addition			

<b>5</b>	Rad	lar Vio	leo Fusion SLAM	52
	5.1	Angle	Change Estimation From Video	52
		5.1.1	Image Summation	52
		5.1.2	Pixel-Space to Angle-Space Conversion	53
		5.1.3	Correlation Between Image Summations	53
	5.2	Angle	Change Estimation Update to Filter	54
	5.3	Reflec	tor Segmentation	55
	5.4	Weigh	t Image Generation	55
		5.4.1	Color Space Reduction	59
		5.4.2	Average Color Likelihood	60
		5.4.3	Possible Reflector Collection	60
		5.4.4	Reflector Identification	62
	5.5	Reflec	tor Tracking	63
		5.5.1	Color Filtering	64
		5.5.2	Possible Detection Collection	64
		5.5.3	Reflector Measurement From Image	66
	5.6	Video	Landmark Measurement Update to Filter	66
		5.6.1	Filter Update	67
6	Exp	erime	nt	68
	6.1	Descri	ption	68
	6.2	Odom	etry Results	70
		6.2.1	Velocity Estimation	70
		6.2.2	Angle Change Estimation	74
		6.2.3	Platform Trajectory From Dead-Reckoning	77
	6.3	Radar	SLAM	 77
	0.0	631	Detections	77
		0.0.1		11

		6.3.2	Filtering	82
	6.4	Radar	Video Angle Change Estimate Fusion SLAM	86
		6.4.1	Angle Change Estimation	86
		6.4.2	Filtering	87
	6.5	Radar	Video Landmark Fusion SLAM	90
		6.5.1	Landmark Identification and Tracking	90
		6.5.2	Filtering	91
7	Con	clusio	n	100
	7.1	Summ	ary	100
	7.2	Conclu	usions	100
	7.3	Contri	butions	102
	7.4	Future	e Work	103

# List of Figures

1.1	The complementary information of radar and camera sensors. $\ldots$ .	6
1.2	Block diagram of the FMCW radar and color video fusion SLAM system.	9
3.3	Camera and radar used in the experiment	20
3.4	Radar Geometry: The reflector and the vectors are assumed to lie in	
	the x-y plane	21
3.5	The geometry for estimating the angle of arrival. The difference can	
	be approximated as $d\sin(\alpha_s)$ because $R >> d$	24
3.6	Example of the log magnitude frequency response received over time	
	by the radar. $\ldots$	24
3.7	Example of the Angles-of-Arrival received over time by the radar	25
4.8	Illustration of the unscented transform process	32
4.9	The the normalized log magnitude response from a pulse. The up-chirp	
	is blue and the down-chirp is red. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	40
4.10	Example log magnitude response	45
4.11	CFAR result with $P_{FA} = 0.065$	45
4.12	Example of the AoA response	46
5.13	An example of radar measurements projected onto an image. The	
	measurements are green. The radar estimate of the reflector is red. $% \left( {{{\bf{r}}_{{\rm{s}}}}_{{\rm{s}}}} \right)$ .	58
5.14	The image without the radar projections. The detections are from	
	orange trihedrals.	58
5.15	An example weight image generated by combining weight images from	
	measurements	59
5.16	An example reflector likelihood image, $L$ . The original image is shown	
	in Fig. 5.14	61

5.17	A binary image corresponding to possible reflectors after all possible	
	reflectors are collected	62
5.18	Image of the detected reflector.	63
5.19	2D color histogram of the segmented reflector	64
5.20	The next image in the video sequence after the image shown in Fig. 5.14.	65
5.21	Pixels in the image whose color matches the reflector color	65
6.22	Scene layout for the experiment	68
6.23	The radar waveform.	69
6.24	A matched log magnitude responses from a pulse using the shift with	
	the maximum correlation. The up-chirp is blue and the down-chirp is	
	red	72
6.25	A matched log magnitude responses from a pulse using Gauss-Newton	
	optimization. The up-chirp is blue and the down-chirp is red	73
6.26	A matched log magnitude responses from a pulse using Gauss-Newton	
	optimization accounting for AoA . The up-chirp is blue and the down-	
	chirp is red	73
6.27	Example 1 of velocity estimates overlaid on the CW log magnitude	
	response. Method 2 estimates are blue and method 3 estimates are red.	74
6.28	Example 2 of velocity estimates overlaid on the CW log magnitude	
	response. Method 2 estimates are blue and method 3 estimates are red.	75
6.29	Example of angle change estimates. The combined estimate is in black.	
	The up-chirp and down-chirp only estimates are in blue and red, re-	
	spectively	75
6.30	The platform trajectory from dead-reckoning estimates	77
6.31	A result after applying the angle smoothness constraint. $\ldots$ . $\ldots$	78
6.32	Angle smoothness based detections after threshold.	79

6.33	Result from combining CFAR and angle constraint to the increasing	
	frequency chirp observations	79
6.34	Result from combining CFAR and angle constraint to the decreasing	
	frequency chirp observations	80
6.35	Actual detections from decreasing frequency chirp observations	80
6.36	Maps for comparison of radar detection methods. The CA-CFAR map	
	is cyan. The angle constrained CA-CFAR map is magenta	82
6.37	Maps for comparison of radar detection methods when a similar num-	
	ber of landmarks are mapped. The CA-CFAR map is cyan. The angle	
	constrained CA-CFAR map is magenta	83
6.38	Map and trajectory from radar SLAM	84
6.39	A comparison of the trajectory between SLAM from radar (black) and	
	dead-reckoning only (red)	84
6.40	Path section and landmarks used in tables	85
6.41	Image sums before correlation. The previous image sum is blue. The	
	current image sum is red	87
6.42	Image sums after correlation. The previous image sum is blue. The	
	current image sum is red	88
6.43	Image on which angle change is being estimated	89
6.44	Platform orientation estimation using angle estimates from video. The	
	orientation estimates from the radar, image, and filter are colored red,	
	blue, and black, respectively	90
6.45	Map and trajectory from radar and video angle fusion SLAM. The	
	estimated landmark locations are cyan	91
6.46	The trajectories from the radar dead-reckoning and the radar with	
	video angle fusion SLAM. The dead-reckoning and radar video fusion	
	SLAM trajectories are red and black, respectively	92

6.47	Map of the landmarks tracked using the radar and video. The land-	
	marks identified in the video are cyan	93
6.48	Radar measurements corresponding to the light pole and tree overlaid	
	on their corresponding image.	93
6.49	Orientation estimates from radar dead-reckoning (red), video (blue),	
	and video landmark tracking fused result (black)	94
6.50	The difference in orientation estimates between the radar and video	
	fusion algorithms.	95
6.51	The trajectories from the radar video fusion without landmark tracking	
	(red) and the radar video fusion with landmark tracking (black)	96

# List of Tables

6.1	Average SSDs for velocity estimation methods.	71
6.2	Path lengths from integrating velocity estimates and their errors. The	
	smallest error in each row is red.	76
6.3	Actual and Estimated Path Lengths and Distances Between Land-	
	marks from Radar SLAM	86
6.4	Actual and Estimated Angles for Each Path Turn From Video $\ .\ .\ .$	87
6.5	Actual and Estimated Path Lengths and Distances Between Land-	
	marks from Radar SLAM and Video Angle Measurement Fusion $\ . \ .$	89
6.6	Actual and Estimated Path Lengths and Distances Between Land-	
	marks from Radar SLAM and Video Fusion	96
6.7	All of the path length estimates and their errors. The smallest error	
	in each row is red.	98
6.8	All of the landmark distance estimates and their errors. The smallest	
	error in each row is red.	99

## Acknowledgements

I would like to thank my advisor, Dr. Lang Hong, for his academic guidance and support for the duration of the Ph.D. program. I am grateful for his mentoring, advice, and research support.

I also wish to thank Dr. Fred Garber, Dr. Kefu Xue, Dr. Arthur Goshtasby, Dr. Richard Martin, and Dr. Devert Wicker for serving as members of my Ph.D. defense committee. I would like to extend my gratitude to the committee members for their advice and constructive criticism. I also want to thank the Ph.D. program and the Department of Electrical Engineering at Wright State University for giving me the opportunity to obtain my Ph.D. degree at Wright State University.

I would like to thank Dr. Kyle Erickson for his mentorship and advice. I also wish to thank Todd Rovito, Clare Mikula, Dr. Juan Vasquez, Christina Schutte, Dr. Dale Nelson, and all of the others at Air Force Research Lab who supported me while working on this Ph.D.

I also would like to thank Nicholas Gale, Bernard Abayowa, Weston Earick, Thomas Pestak, Veda Galigekere, Nicholas Baine, Richard Van Hook, Karmon Vongsy, and Arunesh Roy for their companionship and support along this journey.

A special thanks goes to my wonderful wife, Nisha, for her love, encouragement, patience, and support through these years.

## 1 Introduction

#### 1.1 Motivation

The Chief Scientist of the United State Air Force released a report on Technology Horizons in 2010 outlining the future direction of the Air Force over the next 20 years [39]. In this report research focus areas were listed. A recurring theme within the list of focus areas was increased autonomy. A call for research in improved precision navigation and timing in GPS-denied environments was also included in the list.

Simultaneous Localization and Mapping (SLAM) provides a way to estimate the location of a platform as it moves through a scene while providing information about the location and descriptions of landmarks in the scene. The landmark information from a scene is retained in a map for later use, when that location is revisited. The map can take on different forms. As a probabilistic grid map, a scene is divided into sections, each of which contains the probability that a location contains a landmark. A map could also consist of a list of locations of landmarks in a scene. The list could be treated as a set or an algebraic vector, influencing the way information about each landmark is treated and understood.

SLAM has been performed by mounting a variety of sensors on moving platforms and fusing information from all of the sensors. Fusing a large number of different sensors is not always feasible. Size and power constraints can limit the number and types of sensors available. It is prudent to use sensors that have multiple uses under such conditions. For example, surveillance drones typically record some form of imagery of a scene to be relayed for humans to analyze. Performing navigation with imagery would be an ideal way to use the same information from the same sensor for multiple purposes.

#### 1.2 Sensors

The variety of sensors that have been used in SLAM and tracking include electrooptic (EO) and radar in the electromagnetic spectrum. Within the realm of EO sensors are ladar, infrared, visible light, and hyperspectral sensors. Acoustic arrays and sonar have been proven useful under certain conditions. For the operating conditions of this problem, electromagnetic sensors are preferred.

Radar is an active sensor that can be used day or night. The returns from radar provide absolute measurements. Information about the range or the range rate between the radar and an object are easily obtained from the return signal the radar receives. Depending on the sensor configuration and processing, azimuth and elevation information about reflectors in a scene may be also obtained, but with high uncertainty. A radar can interrogate the scene very often. Depending on the type of information being extracted, processing can be done very quickly. The form in which the information comes from each interrogation, however, is complex. In urban areas, it can suffer from multi-bounce effects. The signal frequency, frequency bandwidth, pulse length, pulse repetition frequency, and transmit power affect how much of a scene can be observed and to what detail it can be observed. When radar is used for SLAM, a scanning millimeter wave radar is usually used. For this type of radar, the radar is pointed at a particular direction and the return signal tells the ranges to reflectors along that particular direction. It is common for scanning radars to sample a full 360° at about 1 Hz.

EO sensors can be either active or passive. Ladar, for example, is an active sensor. Visible light and hyperspectral sensors are passive. The best illumination source, the sun, limits their use to daytime. Mid-wavelength Infrared (MWIR) and Long-wavelength Infrared (LWIR) sensors do not require sunlight, but they measure heat. Obtaining scene geometry reliably from those sensors can be difficult. Despite their limitations due to weather and lighting conditions, visible and hyperspectral EO sensors are still useful in many applications. They are lightweight and require low power. There is a tradeoff between the sampling rate and image size. Using color over panchromatic imagery offers more features at the cost of sampling rate for a given image size. The information these sensors provide is particularly useful to humans because of the quick and limited amount of processing necessary for use by humans. They provide good information about the bearing to observable features in the scene.

Ladar provides good 3D information about objects it senses. It is not very good at observing large areas very quickly. As in the radar case, it does not provide good feature information about the objects in the scene without a large amount of processing and integration of information obtained over multiple interrogations. Many implementations of SLAM use a laser scanning radar. Most implementations only sweep through a plane, however, which is limiting to ground vehicles.

Time-of-flight cameras are a type of ladar that does not require scanning. A pulse of light is transmitted to the world and the return is collected by a pixel array similar to a camera. The depth of objects in the scene can be estimated from the time the returns of each pulse take to reach the array. Measurements can be obtained from this sensor up to the order of 100 Hz. This type of sensor is limited by external lighting and weather conditions. It also is limited to measuring depths less than 100 m.

The large spectral response from hyperspectral imaging provides a lot of information to distinguish features. A large number of electromagnetic wavelengths are sampled in hyperspectral imagery, typically ranging from infrared through the visible spectrum. It is very slow when compared to the other sensors being discussed. Depending on the method for generating observations, it either cannot sense the entire scene very fast or it can only provide limited spectral information about the scene over a short time frame.

Both infrared and EO sensors give good bearing information about objects in

a scene. Windows around specific points in the images are typically used to obtain features in the image. When a target is a different temperature from its surroundings, infrared can be more reliable for making detections. There is a large body of mature work done in EO. Using a monocular camera for SLAM has been around for almost a decade. The problem with monocular SLAM is the lack of range information from video. The map and camera locations can only be known up to a scale factor.

#### **1.3** Sensor Fusion

Sensor fusion has become an ubiquitous tool for automating vehicle navigation. Radar and video have been fused in the past two decades in the automotive industry for automatic cruise control and collision avoidance. Jia et al. pointed out that only within the past decade has the industry begun to incorporate SLAM systems for navigation [45]. An excellent review of SLAM is given in [60, 59]. Multi-modal SLAM systems typically consist of more than two modes for fusion. Often, multiple sensors of the same type are used to observe more of the scene around a vehicle. Systems used in the DARPA Grand Challenge and the DARPA Urban Challenge provide an example of the power of this type of fusion for navigation [43, 42, 44, 41, 40].

There are various levels of fusion within the sensor fusion framework. In highlevel fusion, detections from each sensor are tracked independently of other sensors. Tracks from different sensors are then associated to form a fused tracking solution. In low-level or feature-level fusion, information from each sensor is associated before the information is input to the tracking filter.

A high level fusion scheme wherein range features are tracked separately from image features produces better estimates than can be made by either sensor alone. Passing the processed track information between sensors is more feasible than passing raw data when communications between sensors is limited. The major difficulty then becomes associating which track from one sensor corresponds to which track from the other sensor.

Feature level fusion schemes use raw data from each sensor to aid in signal processing. The ability to obtain detections in weak signals is increased. The data association problem can be mitigated by the detection process at this level of fusion as well.

A large portion of the SLAM body of work has been concerned with ground-based platforms. Airborne platforms introduce additional challenges that will affect sensor choice. Ground-based platforms can limit landmark and platform positions to lie in a 2D plane. Scanning sensors work well under this constraint. The platform effectively only has 3 Degrees of Freedom (DoF). There are 6 DoF for airborne platforms. This leads to an expected increase in the amount of jitter the platform undergoes. Registration between scans in this type of a 3D environment becomes difficult.

When choosing which sensors to fuse, it is desirable that the sensors provide complementary, or orthogonal, information about the platform and/or landmarks. Weaknesses of one sensor should be overcome by strengths in another sensor. For better map and location estimates, this is especially true when a minimal number of sensors is used.

Two sensors that appear to complement each other well are Monopulse FMCW radar and color video. The radar provides range information to objects in the scene. It works in a variety of weather conditions. Phase-comparison monopulse provides a measure of the direction to detections which can provide a link to the video. Color video was chosen because its more accurate bearing information complements the radar well. It is also more likely to have an alternative use on the platform. By identifying the reflectors that cause the radar detection in the video, more accurate estimates of reflector and platform locations can be made. A low-level fusion scheme can accomplish this goal, while reducing the computational cost of performing detection and tracking in each individual sensor separately. An illustration of



Figure 1.1: The complementary information of radar and camera sensors.

how measurements of the same landmark from each sensor can reduce uncertainty is shown in Fig. 1.1.

SLAM will be performed by mounting the sensors in the forward looking direction of a ground-based platform. These sensors have not been fused for SLAM in any capacity previously. By limiting the platform to a ground vehicle, and keeping the map 2D, an assessment can be made of the feasibility of using these sensors on an airborne platform.

Radar SLAM will provide the foundation for the low-level fusion solution. Video information will be added as it becomes available to improve the estimates. This method is chosen for a number of reasons:

- The radar is all-weather. In the event that video data is unavailable (cloudiness, darkness, etc.), the radar should still maintain some navigation capability.
- The radar can make many more observations per second than the video. It will be shown that those observations can provide direct information about the

motion of the platform, without relying on triangulation with landmarks. The radar can provide odometry information at a rate compatible to an Inertial Measurement Unit (IMU).

- There are expected to be much fewer feature detections in the radar to track. Much work in the SLAM domain has dealt with the problem of map management and computational costs in updating. Adding landmarks to the map helps improve estimates to a point, but also increases processing requirements quadratically. Adding a dimension and tracking landmarks in 3D exacerbates the problem.
- The amount of data to process per radar observation is much smaller than the amount of data to process per video observation. Image processing is a computational bottleneck. Focusing on finding reflectors in specific regions of the images can reduce the burden.

#### **1.4 Problem Statement**

Multiple challenges need to be overcome when fusing FMCW radar and color video.

- There needs to be a solution to the monopulse radar SLAM problem. To the author's knowledge, monopulse radar SLAM has not been performed before. In some cases, landmarks might not be detectable. During these times, some form of navigation will be necessary.
- Improvements in the detection methods for monopulse radar are necessary for use in SLAM. A unique challenge with this type of radar is the instability in angle of arrival measurements. Strong reflectors in a scene should be detectable over a consecutive set of observations. There should be some continuity in the angle-of arrival (AoA) associated with those detections. Adding a constraint on

the AoA might also allow the classic detection constraints to be relaxed, making more detections possible overall.

• A way of linking the AoA measurements from the radar with features in the video needs to be improved. A strong reflector does not have to have a specific color or take a specific shape in the image. It is expected that the reflector will be discernible from its surroundings in the image.

#### 1.5 Approach

The approach taken herein is to use the radar as a foundation for a low-level SLAM fusion algorithm. Radar observations are used to perform a dead-reckoning prediction of the platform location. This is done before landmark detections are obtained from the radar. It can be performed in the absence of detectable reflectors, though it is assumed that there is some structure in the return related to the scene. Individual landmark detections from the radar observations are then found and used to improve the platform location estimate and map landmarks in the scene. Angle-of-arrival information will be used to improve the detection process. Associations will be made as part of the detection process where possible in order to reduce processing. The electro-magnetic reflectivity requirement will keep the number of landmarks small, reducing the search space necessary both in the image domain as well as for map loop closing.

As observations arrive from the video, at a slower rate from the radar, a measurement of the platform orientation change based on the video is used to update the platform estimate. An optical flow method will estimate the orientation change between the current video frame and the previous. This method will efficiently take advantage of the bearing information that the video provides. It will directly complement the range measurements provided by the radar for landmarks.

A radar-driven segmentation for association with color video is applied. It is



Figure 1.2: Block diagram of the FMCW radar and color video fusion SLAM system.

assumed that some objects in the scene will have high electromagnetic reflectivity. Without this criteria, the radar will not provide detections and the algorithm will not have a reflector for which to search. The radar-driven segmentation allows for direct association between radar detections and objects in the video.

The measurements collected from the radar detections are projected onto the image. The likely color of the reflector that generated the radar measurements is then estimated. An estimate of the radar reflector is then segmented from the image. The colors contained in that segmented object are stored as a feature describing the landmark. Once a landmark is found in the video, a color filter is applied to subsequent video observations to track the landmark. Limiting the tracked features to reflectors can reduce the search space for detections while taking advantage of the rich features provided by video.

A block diagram of the fusion approach is provided in Fig. 1.2.

#### 1.5.1 Contributions

The methods described in this work provide two solutions for fusing information from phase-comparison FMCW radar with color video. In order to navigate in scene without the presence of strong reflectors, a method for estimating dead-reckoning parameters from phase-comparison FMCW radar return waveforms is provided. A method is provided for improving reflector detections, such that the detected reflectors are more likely to be observed frequently while in the radar field of view. This work also describes a method of identifying the source of radar detections in color video.

#### 1.6 Outline

The rest of this dissertation is outlined as follows. Chapter 2 covers background information related to previous methods of performing SLAM as well as methods for radar and video fusion. In chapter 3, the sensor system used for fusion is described. Descriptions of the map and platform are also given. The algorithms necessary to carry out the radar SLAM are described in chapter 4. The radar and video fusion algorithms are provided in chapter 5. In chapter 6, the an experiment is described and the results from the experiment are shown. The dissertation is concluded in chapter 7.

### 2 Background

#### 2.1 SLAM

As its name suggests, SLAM is used to estimate the location of a sensor while estimating the locations of features in a scene. An excellent review of the SLAM process and algorithms is provided in [60] and [59]. The scene is usually static while the sensor moves through the scene, though there has been some work in which scene objects do move. Recursive updates of the current sensor location and scene map are applied as information arrives. The nonlinear measurement process of most sensors usually calls for the use of the Extended Kalman Filter (EKF) or the Unscented Kalman Filter (UKF).

An important development for SLAM was made when it was shown that as more observations are made, correlations between landmarks are built, and the map converges [22]. This causes the map to become rigid, as the locations of landmarks are known relative to each other. Much of the work over the past decade has focused on loop-closing, processing, and map management. Loop-closing occurs when a platform leaves an area of a scene, returns later, and correctly associates new measurements with landmarks observed during the previous visit. Loop-closing is an important issue because it reduces the errors that build over time as the platform moves through the scene. Ramos et al. showed that using location and appearance to associate landmarks improves the likelihood of loop closure [9]. Map management is necessary to deal with the large number of landmarks and landmark covariances that build over time. In order to reduce processing, it can be beneficial to only operate on certain landmarks or to break the map up into sections which are treated semi-independently.

Processing is an issue in SLAM because as the size of the map grows, the amount of processing necessary increases. Some processing in the propagation stage can be reduced by taking advantage of the fact that only the platform and platform covariances are changed during this step. The update stage is still a problem however, as it grows quadratically in the case of the EKF and cubically in the case of the UKF. Square-root implementations of the UKF have been able to reduce the number of computations necessary, but it is still more costly than the EKF [32].

A further improvement in alternate direction in SLAM processing came with the introduction of FasSLAM [21]. A particle filter was used in the propagation stage of the filter to estimate the platform state. The way the particle filter is used, it represents the history of the platform. It takes advantage of the fact that the landmarks are correlated over the platform history. That means that individual intra-landmark correlations do not have to be maintained and updated, and the landmarks can be updated individually by an EKF, or some other efficient filter.

These are important issues that will likely have to be addressed in future work as the scene area grows and the map sizes are increased. A benefit, and a possible problem, of the fusion method described herein is that by limiting the landmarks to radar reflectors, the map size does not grow as quickly as if video landmarks are used. Also, by keeping the map to 2D, the processing requirements are reduced.

#### 2.2 Radar SLAM

Scanning radar has been a popular tool for SLAM. This can be done either mechanically or by beam forming. It works by transmitting an electromagnetic signal with a narrow beam width along a direction and measuring the return signal. By this method, each detection has range and an angle measurement associated with it. Detections are usually made using a Constant False Alarm Rate (CFAR) detector, or some variant thereof.

An early attempt at radar SLAM was made by Clark and Durrant-Whyte [23], using the scanning radar system described in [24]. The system was mounted on a truck and driven around a scene with reflectors placed around the path. The reflector returns were polarized in 2 directions which was used to improve their detection likelihood and prevent false alarms from other objects in the scene. Encoders mounted on the vehicle drive shaft provided dead-reckoning measurements which were used in the EKF as control inputs when propagating the filter. Although the truck location uncertainty grew without any radar measurements, the control inputs provided a way to estimate how much and in what direction the vehicle moved whether or not radar measurements were obtained.

Clark and Dissanayake removed the polarized reflectors to perform SLAM with the same system in [24]. They were able to track natural features in the scene by only adding landmarks whose radar returns had polarization in two directions.

Chandran and Newman minimized a spatio-temporal cost function to estimate the platform trajectory and landmark map [13]. A weight applied to the distance between a detection obtained from the 360° scanning radar and a map landmark was increased according to the amount of time since the landmark was last observed.

A scanning radar was used to create map images by Roureure et al. Correlations between sequential images were used to estimate the platform motion between scans [8, 15, 17]. Instead of storing landmarks, an occupancy grid representation of the map was stored. The same radar and similar approach was used by Checchin et al. [11]. Correlations between images were estimated in the Fourier domain by applying the Fourier-Mellin Transform.

Mullane et al. used a Rao-Blackwellised Probability Hypothesis Density (PHD) filter to perform radar SLAM [25]. The PHD filter was used to avoid direct association between radar measurements and landmarks. In order to do this, it treated the landmarks and measurements as sets, as opposed to algebraic vectors. The radar scanned 360° and had a range of 5 km. Even though the a full scan took 0.5 seconds, the platform moved so little relative to the size of the scene being mapped, registration was ignored.

Lundquist et al. tested 3 methods for using a radar to estimate the free space in front of a vehicle [27]. The solutions involve tracking the location of the vehicle in the scene and identifying the locations of other objects in the scene. A mechanically scanning radar was used to measure the distances to objects in front of the vehicle at specific angles. The first method used an occupancy grid representation to estimate the likelihood that regions in the scene contained a reflector. The second method modeled the shape of the road borders using a polynomial. A quadratic constraint over the polynomial was used to smooth the estimates obtained from the measurements. The third method was to track points and lines in front of the vehicle. Extended objects such as guard rails were better modeled as lines on the 2D map. In this method, focus was placed on the appropriate way to associate measurements with the points and lines in the map.

Yokoo et al. fused 2 radars mounted on a vehicle in concert with a gyro sensor to perform navigation [28]. Velocity measurements were obtained from each radar. The average of the velocities was taken as the platform velocity. The difference in the velocities provided a measure of the platform angular velocity. They identified that incorporating AoA measurements in estimating the velocities improved the results. Their method of velocity estimation used a phase derivation technique, which was possible when 2 or more reflectors were observed by the radar.

Using the phase-comparison monopulse radar avoids the registration errors with which scanning radars must contend. Each observation occurs over the entire FOV of the radar. When a reflector is detected, it can be measured in the next observation, without waiting for the radar to scan through individual angles. Associations are made when possible during the detection process, reducing the burden on costly association methods later in the filter process.

The phase-comparison radar enables an additional constraint on obtaining radar detections. In the radar SLAM work described above, variations of CFAR that operated on the magnitude or phase response of the returned radar signal were used. The AoA measurements from phase-comparison radar provide an additional constraint that can improve the quality of detections.

The method used to estimate the velocity and platform orientation change only requires the single monopulse radar. There is not a requirement for detectable reflectors to be present in the scene.

#### 2.3 Video SLAM

Video SLAM was first done by Davison [58] using an EKF. A large number of landmarks, with rich feature descriptors, can be obtained using imagery. Because video only provides bearing information, the need for an increased number of landmarks is greater. The Scale Invariant Feature Transform (SIFT) [35] and a more efficient variant, Speeded Up Robust Features (SURF) [34], have recently become common tools to find features in the images. Initially, the Shi and Tomasi feature detector [36] was used to find features and correlation methods were used to track those features.

A key requirement in video SLAM was the use of an inverse depth representation of each landmark. It increased the state size, but allowed landmarks more mobility to correct as their ranges were more accurately estimated. Feature initialization has been an issue with monocular SLAM. Bearing measurements for a single landmark typically have to be observed multiple times before they can be added to the filter. Since then, improvements have been made to reduce the state size and for quick initialization [38, 37]. Performing video SLAM requires obtaining features in an image and tracking them in subsequent frames. Whether the images are monochrome or color, video processing is computationally intensive. As stated before, monocular video can only map the scene up to a scale factor.

The EKF has been a common tool in video SLAM due to the large number of landmarks and the increased state size necessary for inverse depth representation. The nonlinearities of the motion and measurement models can allow bias to enter the map and platform location estimates. Sunderhauf et al. showed the feasibility of applying the UKF to monocular SLAM [33]. An obstacle they had to overcome was determining how to handle negative inverse depth sigma points. The Square-Root Unsceneted Kalman Fitler (SRUKF) was applied to monocular SLAM by Holmes et al. [32]. The state was aligned to reduce computational costs in addition to the reduction afforded by the SRUKF alone.

Airborne video SLAM has been attempted by fusion with IMU information. The IMU provides a measure of scale in the scene. Kim and Sukkarieh used an EKF to perform fusion [10]. An experiment was carried out where white markers were placed on the ground to be easily identified in monochrome video by thresholding intensity. IMU data was input to the filter as rotation and acceleration information. They also described an indirect fusion method where the IMU data was integrated to provide position, orientation, and velocity measurements separately from the filter. The purpose was to maintain an up-to-date estimate of the platform parameters as information arrived. The method provided by Sjanic et al. used Square-root Smoothing and Mapping (SAM) [29]. The objective of SAM is to minimize a quadratic cost function based on the error in the platform trajectory and the measurements. An EKF was used to provide initial estimates of the map and trajectory. After enough observations were made, SAM was applied, improving the estimates.

The methods previously used to perform video SLAM relied on searching the images for possible features. The approach taken here is to reduce the amount of image processing necessary. Estimating the platform orientation change through the video takes advantage of the much finer bearing resolution that the video provides over the radar. By searching for radar reflectors, the range uncertainty problem can be avoided. The video can improve the landmark estimates while knowing the range can keep the necessary state size lower. Initialization issues encountered in video can also be avoided, as the landmarks have already been initialized in the filter by the time they are found in the images.

#### 2.4 Fusion Methods

There has been a great deal of work in data fusion. For relevance purposes, schemes involving track-to-track fusion are avoided and the following section is limited to work wherein AoA and bearing information are used to fuse a range-based modality with an image.

Heisele et al. did early work in automatic cruise control [52]. Information from a millimeter-wave (MMW) radar and color video sensing the scene in front of the vehicle were fused. The images were segmented using a fast color connected component algorithm. Voxels of interest were obtained from the radar. Regions of interest were projected onto the images from 3D radar detections. Color segments with a minimum amount of overlap with a region of interest were associated with that radar detection. Different color segments were grouped if they belonged to the same radar detection.

Haselhoff et al. fused information from three radars and a monochrome camera [51]. The radars produced regions-of-interest (ROI) that were projected onto the images. The size of each ROI was set to approximately 5 m by 4 m. The AdaBoost algorithm was run on sub windows of each ROI in the image to detect rear views of cars [53]. In order to perform the detection, images of car rear views were used to train the algorithm.

Mahlisch et al. fused video with a lidar for detecting obstacles in front of a vehicle [49]. The 16-channel multi-beam lidar measured distances to objects at set angles. Ellipses were projected onto the images corresponding to the regions where objects were expected to be. The ellipse size associated with each beam was dependent on the range estimated from that beam, i.e. larger distances had larger ellipses. Objects were detected in the image using a cascaded AdaBoost detector. The object search space in the image was limited to regions near the lidar projections. Detected objects were clustered and associated with the lidar measurements. Further processing was carried out to classify detections as either clutter or cars.

Gern et al. used a radar with AoA measurements and video [50]. The objective of this work was to track cars in front of the platform. A search area was set for template matching by finding areas with high vertical and horizontal symmetry. Templates and symmetry were useful because, typically, only the rears of preceding vehicles were visible. Rears of cars tend to have a box shape. Symmetry was found by looking at gradients in the intensity of the image. Matches close to radar projections were associated with those detections.

Bombini et al. fused a scanned radar with a grayscale camera [48]. Vertical symmetry was used again to detect the rears of vehicles. A search for horizontal symmetry was also performed, but the focus was on finding the more stable, dark undersides of vehicles. This approach was also used by Alessandretti et al. in [47].

Roy et al. fused radar and video for surveillance [46]. In that work, the radar and camera system was stationary. Radar reflectors were assumed to be moving cars. The cars were segmented by using change detection on the images. Radar detections that were found to be close to image detections were associated.

The approach described herein uses the detections from a radar with 2 receivers to drive image segmentation in a color video. Estimated locations of radar detections are projected onto the image. The algorithm searches for colors that appear often near those projections, but not often in areas of the image more distant from radar detection projections.

Apart from providing range measurements between the platform and landmarks, the radar provides odometry information similar to what might be provided by an IMU. The orientation change measurements provided by the video provide a higherlevel fusion specific to SLAM that was previously provided by IMUs or wheel measurements in previous fusion algorithms.



Figure 3.3: Camera and radar used in the experiment.

## 3 System

#### 3.1 Radar

The system under consideration consists of a Phase Comparison Monopulse Frequency Modulated Continuous Wave (FMCW) radar and a color video camera. The system is shown in Fig. 3.3. The radar has 2 adjacent receivers which allows for angle-ofarrival estimation to a reflector. For a unit direction vector,  $\underline{r}$ , describing the direction between the radar and a reflector and another unit direction vector,  $\underline{d}$ , describing the direction of the displacement between the two radar receivers, the angle of arrival,  $\alpha$ is given by

$$\alpha = \frac{\pi}{2} - \arccos(\underline{r}^T \underline{d}). \tag{3.1}$$

Fig. 3.4 describes the radar geometry.

The displacement between the radar receiver is assumed to be parallel to the image plane of the camera and also parallel to the x axis of the image plane. These constraints make a majority of the effective field of view of the radar overlap with the



Figure 3.4: Radar Geometry: The reflector and the vectors are assumed to lie in the x-y plane.

field of view of the camera. For a sensor separation of less than 10 cm and a minimum distance to any reflector of 10 m, the sensors can be treated as if they are coincident, and  $\alpha$  can then be treated as an azimuth measurement in the image domain.

The radar transmits a series electromagnetic wave pulses. The pulses can consist of a variety of modulation schemes. The basic signals take the form

$$s(t) = A\cos\left(\omega(t)t\right) \tag{3.2}$$

where  $\omega(t)$  is a linear function of t of the form

$$\omega(t) = \alpha t + \beta \tag{3.3}$$

where  $\alpha$  is 0 for a constant frequency, positive for an increasing frequency chirp, and negative for a decreasing frequency chirp. The signal is typically periodic, and could consist of the same increasing frequency chirp (sawtooth), an increasing chirp followed by a decreasing chirp (triangle), etc. When the signal comes into contact with the *i*th reflector, a time delayed version of the signal is returned to the radar as

$$r_i(t) = A_i \cos\left((\alpha(t-t_i)+\beta)t + \nu_i t\right),\tag{3.4}$$

where  $A_i$  is an attenuation factor,  $\nu_i$  is a Doppler shift due to the relative velocity between the reflector and radar, and  $t_i$  is the time delay of propagation, and i is the
reflector index. When multiple reflectors are in the scene, the received signal is a combination of the returns from each of the reflectors as

$$r(t) = \sum_{i} r_i(t)$$
  
=  $\sum_{i} A_i \cos\left(\omega_i(t)(t-t_i) + \nu_i t\right).$  (3.5)

The  $\omega_i(t)$  term is a delayed version of the original chirp. It is shifted by  $t_i$ . The received signal is mixed with a copy of the transmitted signal and filtered, bringing the result from radio frequencies down to intermediate frequencies. The mixing and filtering creates a signal of the form

$$m(t) = \frac{A}{2} \sum_{i} A_i \cos\left(-\omega(t)t_i + \nu_i t\right).$$
(3.6)

Because the chirps are linear, each cosine term in Eq. 3.6 has a constant frequency. Taking the Fourier Transform (FT) of Eq. 3.6 should result in a set of shifted delta functions. Because of time-windowing in the received signal, the energy from each reflector is spread around the peak. Non-linearities in the chirp will also cause spreading.

For explanation purposes, only a decreasing frequency chirp is described further. The received signal is broken into segments corresponding to each of the waveform types of each transmitted signal pulse, i.e. only the time portion of the received signal period pertaining to the decreasing chirp is considered. A FT is performed on each segment of the mixed and filtered signal. A reflector, s, in the scene induces a response,  $A_s e^{-j\phi_s}$ , at a frequency,  $f_s$ . This frequency shift,  $f_s$ , is described by

$$f_s = \frac{2}{c} \left( \frac{\beta \rho}{\tau} + f_c v \right), \tag{3.7}$$

where  $\rho$  is the range to the reflector, c is the speed of light,  $f_c$  is the baseband of the transmitted chirp frequency,  $\beta$  is the chirp bandwidth,  $\tau$  is the chirp period, and v is the magnitude of the velocity along the line of sight direction between the radar and

the reflector. For the case where the platform and reflector are stationary, Eq. 3.7 would simplify to

$$f_s = \frac{2}{c} \frac{\beta \rho}{\tau}.$$
(3.8)

Frequency bins with large magnitudes in the FT then can be said to correspond to reflectors. By using the frequencies corresponding to bins with large magnitudes, the ranges to the reflectors can then be found according to Eq. 3.8. It may be necessary to filter the resulting FT to limit the number of false detections. A Constant False Alarm Rate (CFAR) filter can be effective for this purpose.

For a phase-comparison monopulse system, the phase angles at corresponding frequency bins from both sets of return signals are used to estimate the angle-ofarrival. The difference in range between the reflector and each receiver causes a difference in the amount of time required for the return signal to reach each receiver. This time difference causes a difference in the phases between the received signals. For a corresponding pair of detections (one from each receiver), the angle of arrival may be calculated using the difference in phase at each detection frequency as

$$\alpha_s = \arcsin\left(\frac{c(\phi_{s1} - \phi_{s2})}{2\pi f_c \parallel \underline{d} \parallel_2}\right),\tag{3.9}$$

where  $\phi_{s1}$  is the phase at  $f_s$  from receiver 1,  $\phi_{s2}$  is the phase at  $f_s$  from receiver 2, and  $\|\cdot\|_2$  represents the  $\ell_2$  norm operation.

Examples of the log magnitude and AoA responses are shown in Figs. 3.6 and 3.7, respectively. The frequencies of large magnitude responses decrease over time as the radar moves closer to the reflectors. One peak stands out as decreasing faster than the others. That peak is due to a person who was walking towards the radar.

# 3.2 Video Camera

The video camera is a color camera with approximately the same field of view (FOV) as the radar. Each frame consists of an  $m \times n \times 3$  array of 8-bit integers. The



Figure 3.5: The geometry for estimating the angle of arrival. The difference can be approximated as  $d\sin(\alpha_s)$  because R >> d.



Figure 3.6: Example of the log magnitude frequency response received over time by the radar.

value of each array element can range from 0 to 255. Each subarray along the third dimension of the array corresponds to a component of color in the RGB color space. The values in the first subarray correspond to the amount of red in the image. The color components of the second and third subarrays are green and blue, respectively.

# 3.3 Platform and Landmark Descriptions

### 3.3.1 State Variables

For 2D SLAM, there are 3 variables for representing the state. The platform location is marked in Cartesian coordinates as  $x_p$  and  $y_p$ . The orientation, or viewing direction, of the platform is denoted by  $\theta_p$ . The orientation is also assumed to be the direction of platform motion. Each landmark has an x and a y coordinate associated with it. The coordinates of the  $i^{\text{th}}$  landmark are denoted by  $x_i$  and  $y_i$ . The state vector,  $\underline{x}$ ,



Figure 3.7: Example of the Angles-of-Arrival received over time by the radar.

at instance k is then arranged as

$$\underline{x}_{k} = \begin{bmatrix} x_{pk} \\ y_{pk} \\ \theta_{pk} \\ x_{1}k \\ y_{1}k \\ \vdots \\ x_{N}k \\ y_{N}k \end{bmatrix}, \qquad (3.10)$$

where there are N landmarks. The portion of the state vector corresponding to the platform is denoted as  $\underline{x}_{pk}$  and consists of  $x_{kp}$ ,  $y_{pk}$ , and  $\theta_{pk}$ . Note that  $\underline{x}_{pk}$  refers to a vector describing the platform whereas  $x_{pk}$  is a scalar and refers to a Cartesian coordinate. The portion of the state vector belonging to the *i*<sup>th</sup> landmark is denoted by  $\underline{x}_{ik}$ . The state vector can then be written as

$$\underline{x}_{k} = \begin{bmatrix} \underline{x}_{pk} \\ \underline{x}_{1k} \\ \vdots \\ \underline{x}_{Nk} \end{bmatrix}, \qquad (3.11)$$

where

$$\underline{x}_{ik} = \begin{bmatrix} x_{ik} \\ y_{ik} \end{bmatrix}. \tag{3.12}$$

Again, the underline represents a vector, whereas the no underline refers to a scalar Cartesian coordinate of a landmark. The state covariance is given by

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{P}_{pp} & \mathbf{P}_{p1} & \mathbf{P}_{p2} & \dots & \mathbf{P}_{pN} \\ \mathbf{P}_{p1}^{T} & \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \mathbf{P}_{1N} \\ \mathbf{P}_{p2}^{T} & \mathbf{P}_{12}^{T} & \mathbf{P}_{22} & \dots & \mathbf{P}_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{pN}^{T} & \mathbf{P}_{1N}^{T} & \mathbf{P}_{2N}^{T} & \dots & \mathbf{P}_{NN} \end{bmatrix},$$
(3.13)

where  $\mathbf{P}_{pp}$  is the platform covariance,  $\mathbf{P}_{pi}$  is the covariance between the platform and the  $i^{\text{th}}$  landmark, and  $\mathbf{P}_{ij}$  is the covariance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  landmarks.

#### 3.3.2 Motion Models

The platform motion is modeled by

$$\underline{x}_{pk} = \underline{x}_{pk-1} + \begin{bmatrix} \frac{(v_{k-1}+n_{vk-1})t}{(\gamma_{k-1}+n_{\gamma k-1})} \left( \sin(\theta_{pk-1}+(\gamma_{k-1}+n_{\gamma k-1})) - \sin(\theta_{pk-1}) \right) \\ \frac{(v_{k-1}+n_{vk-1})t}{(\gamma_{k-1}+n_{\gamma k-1})} \left( \cos(\theta_{pk-1}) - \cos(\theta_{pk-1}+(\gamma_{k-1}+n_{\gamma k-1})) \right) \\ \theta_{pk-1} + \gamma_{k-1} + n_{\gamma k-1} \end{bmatrix}$$
(3.14)

where t is the time between propagation instances,  $v_{k-1}$  is the velocity of the platform at instance k - 1,  $n_{vk-1}$ ) is noise in the velocity estimate,  $\gamma_{k-1}$  is the change in orientation of the platform from instance k-1 to k, and  $n_{\gamma k-1}$  is noise in the platform orientation change estimate. The variables to be estimated for performing deadreckoning are  $v_{k-1}$  and  $\gamma_{k-1}$ .

Note that the velocity and orientation change of the platform are estimated independently. In the UKF implementation, the additive noise is obtained from uncertainties in the platform velocity and orientation change estimation process.

As stated above, it is assumed that the sensor viewing direction is the same as the platform motion direction  $\theta_{pk}$ . A problem arises for the UKF as the platform orientation approaches  $\pi$ . Some of the sigma points can be broken down as  $\pi - \epsilon$  or  $-\pi + \epsilon$ , where  $\epsilon$  is a small, positive number. For the later case, with a measured value of  $\pi$ , the innovation error would be  $2\pi + \epsilon$ . The true innovation error should really be  $\epsilon$ . Because the Kalman gain is a function of this innovation error, the estimate will be considerably altered. Therefore, when the orientation is close to  $\pi$ , a correction should be done on the sigma points to make sure they all have the same sign.

The landmark positions are stationary and do not change over time. Therefore, the landmarks do not need an uncertainty term during propagation. The location of the  $i^{\text{th}}$  landmark is then described by

$$\underline{x}_{ik} = \underline{x}_{ik-1} = \underline{x}_i \tag{3.15}$$

### **3.4** Radar and Video Calibration

In order to accurately track objects in space and associate measurements between sensors, the sensors must be calibrated. Calibration is performed by first estimating the intrinsic camera parameters. The intrinsic parameter estimation can be done with a checkerboard and an open source calibration implementation such as OpenCV. The radar is then calibrated to the camera.

#### 3.4.1 Radar Calibration

A set of trihedrals is placed in a scene and recorded using the camera and radar. The trihedrals should be placed at different, known, ranges from the system. The trihedrals and their AoAs are clearly distinguishable in the measurements. The relative range between each trihedral and the radar is constant, since all are stationary.

The frequency bin corresponding to the magnitude response of each trihedral is found by a person. The AoAs from each of those frequency bins are collected. The reflectors are also identified by a person in the images. To account for any possible wind or shaking of the system and trihedrals, multiple frames are used. The x-pixel measurements are converted to AoAs using

$$\alpha_{\text{camera}} = \arctan(\frac{x - p_x}{s_x}), \qquad (3.16)$$

where x is the pixel coordinate,  $p_x$  is the x-direction principle point, and  $s_x$  is the focal length of the camera along the x-direction.

The linear mapping between raw radar AoA measurements and camera AoA measurements is given by

$$\alpha_{\text{camera}} = a\alpha_{\text{radar}} + b. \tag{3.17}$$

In the above equation, a should be close to 1 and b should be close to 0. The parameters should be constant for all trihedrals over all time.

Once the measurements are collected, the radar measurements associated with the image measurements based on their proximity in time. The measurements are arranged as

$$\begin{bmatrix} \alpha_{\text{cameral1}} \\ \vdots \\ \alpha_{\text{cameralk}} \\ \vdots \\ \alpha_{\text{cameraLM}} \end{bmatrix} = \begin{bmatrix} \alpha_{\text{radar111}} & 1 \\ \vdots & \vdots \\ \alpha_{\text{radarkj}} & 1 \\ \vdots & \vdots \\ \alpha_{\text{radarLMN}} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \qquad (3.18)$$

where there are L reflectors, M images, and N radar measurements per image for the collection. A least-squares approach is then used to solve for a and b.

# 4 Radar SLAM

## 4.1 Filters

### 4.1.1 Kalman Filter

The Kalman filter is a commonly used tool in tracking and navigation. For a linear system with additive Gaussian noise, it provides the optimal estimate in a minimum mean squared error sense. The system undergoes changes according to the form

$$\underline{x}_k = \mathbf{F}_k \underline{x}_{k-1} + \mathbf{B}_k \underline{u}_k + \underline{n}_{wk}, \tag{4.19}$$

where  $\underline{x}_k$  is the state to be estimated,  $\mathbf{F}_k$  is the process the state undergoes from instance k - 1 to instance k,  $\underline{u}_k$  is a control input for the state process,  $\mathbf{B}_k$  is a linear process on the control input, and  $\underline{n}_{wk}$  is additive Gaussian noise. The subscript k denotes the instance and signifies that the variable may change from instance to instance. Measurements of the state are obtained as

$$\underline{z}_k = \mathbf{H}_k \underline{x}_k + \underline{n}_{vk}, \tag{4.20}$$

where  $\underline{z}_k$  is the measurement,  $\mathbf{H}_k$  is the measurement process, and  $\underline{n}_{vk}$  is the additive Gaussian noise associated with the measurement process.

The Kalman filter is generally treated as a 2-step process: a propagation step, followed by an update step. The propagation step attempts to predict the next state. The update step corrects the predicted estimate based on information obtained from a measurement of the state. For a state,  $\underline{x}_k$ , the state estimate covariance is given by  $\mathbf{P}_{k|k}$ , and the Kalman filter equations for the propagation step are

$$\underline{\hat{x}}_{k|k-1} = \mathbf{F}_k \underline{x}_{k-1|k-1}, \tag{4.21}$$

and

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k, \qquad (4.22)$$

$$\underline{x}_{k|k} = \underline{x}_{k|k-1} + \mathbf{K}(\underline{z}_k - \mathbf{H}_k \underline{x}_{k|k-1}), \qquad (4.23)$$

and

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_{k|k-1},\tag{4.24}$$

where **I** is the identity matrix with the same size as  $\mathbf{P}_{k|k-1}$ . The Kalman gain can be calculated as

notation convenience. The update step is described by

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{P}_{zz}^{-1}, \tag{4.25}$$

where

$$\mathbf{P}_{zz} = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k, \qquad (4.26)$$

and  $\mathbf{R}_k$  is the covariance of  $\underline{n}_{vk}$ .

Note that Eqs. 4.24 and 4.25 can be calculated independent of the measurement, and could therefore be done during the propagation step. They are put together with the update step because, in practice, the systems are non-linear and variations of the Kalman filter must be used instead. In the alternate filter types, such as the Extended Kalman filter (EKF) and the Unscented Kalman filter (UKF), those equations become dependent on the measurement.

### 4.1.2 Unscented Kalman Filter

The UKF is preferable in this instance because of the large amount of non-linearities in the system. The quasi-linearity requirement of the EKF makes it less likely to be adequate. In using the EKF, a solution to an approximation of the problem is found. The UKF attempts to approximate the solution [56, 55]. It is also known as the Sigma Point Kalman Filter (SPKF) because during processing, the state is composed of a set of sigma points, whose average is the state estimate of the filter.



(a) Sigma points are generated (b) Sigma points are propagated from state estimate and covari- through the non-linear process. ance.



(c) New state estimate and covariance are obtained from sigma points.

Figure 4.8: Illustration of the unscented transform process.

The Unscented Transform (UT) is used in the UKF where a process is non-linear. For the models used in this work, both the propagation and measurement processes are non-linear, so it is used in both the propagation step and update step. The UT process consists of generating a set of sigma points based on the state estimate and its covariance. The sigma points are processed through the non-linear function. A new estimate of the state and its covariance can then be made by taking the weighted average of the sigma points and their covariance.

The UKF attempts to estimate the state,  $\underline{x}_k$  of a system which undergoes a propagation of the form

$$\underline{x}_k = f(\underline{x}_{k-1}, \underline{n}_{wk}) \tag{4.27}$$

and measurement of the form

$$\underline{z}_k = h(\underline{x}_k, \underline{n}_{vk}). \tag{4.28}$$

In the above equations,  $f(\cdot)$  is a propagation function which alters the state from one iteration to the next.  $\underline{n}_{wk}$  is noise that represent the uncertainty in the state propagation. The measurement,  $\underline{z}_k$ , is obtained by passing the state through the measurement function,  $h(\cdot)$ . The uncertainty associated with the measurement process is given by  $\underline{n}_{vk}$ .

When implementing the UKF, it is common to augment the state and state covariance with the expected noise values (should be 0) and the noise covariances as

$$\underline{x}_{k}^{a} = \begin{bmatrix} \underline{x}_{k} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
(4.29)

and

$$\mathbf{P}_{k|k}^{a} = \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{k} \end{bmatrix}$$
(4.30)

For notational convenience, it is assumed from this point that references to the state and state covariance in this section refer to the augmented state and the augmented state covariance.

To implement the filter, the state covariance,  $\mathbf{P}_{k|k}$ , is decomposed to obtain a matrix,  $\mathbf{C}$ , such that  $(\kappa+N)\mathbf{P}_{k|k} = \mathbf{C}\mathbf{C}^T$ . N is the number of elements in the state and  $\kappa$  determines the weight for the sigma point corresponding to the true state estimate. Possible decomposition methods include Cholesky, LDL, and Eigendecomposition. Sigma points are generated as

$$\chi_{k-1|k-1} = \begin{bmatrix} \underline{\chi}_1 & \underline{\chi}_2 & \cdots & \underline{\chi}_{2N+1} \end{bmatrix}$$
  
= 
$$\begin{bmatrix} \underline{x} & \underline{x} + \underline{\sigma}_1 & \cdots & \underline{x} + \underline{\sigma}_N & \underline{x}_{k-1|k-1} - \underline{\sigma}_1 & \cdots & \underline{x} - \underline{\sigma}_N \end{bmatrix}$$
(4.31)

where

$$\mathbf{C} = \begin{bmatrix} \underline{\sigma}_1 & \dots & \underline{\sigma}_N \end{bmatrix}. \tag{4.32}$$

Each of the sigma points is propagated through the filter such that

$$\chi_{k|k-1} = \left[\underline{f}(\underline{\chi}_1) \quad \underline{f}(\underline{\chi}_2) \quad \dots \quad \underline{f}(\underline{\chi}_{2N+1})\right].$$
(4.33)

The propagated state estimate and its covariance can then be obtained by

$$\underline{x}_{k|k-1} = \chi_{k|k-1}\underline{w} \tag{4.34}$$

and

$$\mathbf{P}_{k|k-1} = \varepsilon_{k|k-1} \operatorname{diag}(\underline{w}) \varepsilon_{k|k-1}^{T}$$
(4.35)

where  $\underline{w} = \begin{bmatrix} \kappa/d & 1/d & \dots & 1/d \end{bmatrix}$  is a weighting vector and  $\varepsilon$  is made by subtracting  $\underline{x}_{k|k-1}$  from each column of  $\chi_{k|k-1}$ .

It should be noted that the size of the augmented state vector gets reduced after being run through  $f(\cdot)$ , since the process noise components are used and drop out of the state during propagation. The sigma points are then propagated through the measurement function as

$$\mathbf{Z}_{k} = \begin{bmatrix} \underline{h}(\underline{\chi}_{1}) & \underline{h}(\underline{\chi}_{2}) & \dots & \underline{h}(\underline{\chi}_{2N+1}) \end{bmatrix}, \qquad (4.36)$$

which can be combined to make the estimated measurement and estimated measurement covariance by

$$\underline{\zeta}_k = \mathbf{Z}_k \underline{w} \tag{4.37}$$

and

$$\mathbf{P}_{\mathbf{z}\mathbf{z}} = \vartheta_{k|k-1} \operatorname{diag}(\underline{w}) \ \vartheta_{k|k-1}^T \tag{4.38}$$

where  $\vartheta_k$  is obtained by subtracting  $\underline{\zeta}_k$  from each column of  $\mathbf{Z}_k$ .

Once the measurement,  $\underline{z}_k$ , arrives, the state and state covariance estimates may then be updated by

$$\underline{x}_{k|k} = \underline{x}_{k|k-1} + \mathbf{K}(\underline{z}_k - \underline{\zeta}_k) \tag{4.39}$$

and

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}\mathbf{P}_{xz}^T, \tag{4.40}$$

where

$$\mathbf{K} = \mathbf{P}_{xz} \mathbf{P}_{zz}^{-1} \tag{4.41}$$

and

$$\mathbf{P}_{xz} = \varepsilon \operatorname{diag}(\underline{w}) \ \vartheta^T. \tag{4.42}$$

#### 4.1.3 Linear Regression Kalman Filter

It was shown that the UKF is a special case of the Linear Regression Kalman Filter (LRKF) [54].

The linear approximations of the propagation function,  $\hat{\mathbf{F}}$ , and the control input,  $\hat{\mathbf{B}}$ , can be found by

$$\begin{bmatrix} \hat{\mathbf{F}} & \hat{\mathbf{B}} \end{bmatrix} = \mathbf{P}_{yx} \mathbf{P}_{k-1|k-1}^{-1}, \tag{4.43}$$

where

$$\mathbf{P}_{yx} = \varepsilon_{k|k-1} \operatorname{diag}(\underline{w}) \varphi_{k-1|k-1}, \qquad (4.44)$$

with  $\varphi_{k-1|k-1}$  obtained by subtracting the state from  $\chi_{k-1|k-1}$ .

Likewise, estimates of the linearized measurement, **H** can be made by

$$\hat{\mathbf{H}} = \mathbf{A}_n^{-1} \begin{bmatrix} \mathbf{A}_x & \mathbf{I} \end{bmatrix}, \tag{4.45}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_x & \mathbf{A}_n \end{bmatrix}$$
  
=  $\mathbf{P}_{xz}^T \mathbf{P}_{k|k-1}^{-1}$  (4.46)

#### 4.1.4 Consistency Test

As it was stated above, the UKF is preferable to the EKF. The EKF allows bias into the state estimates over time, while the UKF is designed to prevent it. A way of identifying when bias is entering into the state estimates is through a consistency test. A consistency test provides a way to measure the performance of the filter.

The covariance associated with each state estimate provides a measure of the uncertainty in the estimate. The Mahalanobis distance provides a normalized error metric based on the state estimate and its covariance. It is given by

$$d_m = (\underline{x} - \underline{\hat{x}})^T \mathbf{P}^{-1} (\underline{x} - \underline{\hat{x}}), \qquad (4.47)$$

where  $\underline{x}$  is the true state and  $\underline{\hat{x}}$  is the estimated state. To perform a consistency test, the experiment would be run N times and the Mahalanobis distances for each estimate would be calculated and summed for that instance as

$$\kappa_k = \sum_{n=1}^N d_{mnk},\tag{4.48}$$

where n is the test number and k is the instance variable over which the filter propagates. The sum of the Mahalanobis distances is a Chi-squared variable. The consistency test is done by comparing  $\kappa_k$  to the Chi-squared test thresholds.

When the true state values are known, the consistency test can still be performed. The Mahalanobis distance between the measurement and the predicted measurement still behaves as a Chi-squared variable when normalized by the innovation covariance. In this case, the Mahalanobis distance is given by

$$d_m = (\underline{z} - \hat{\underline{z}})^T \mathbf{S}^{-1} (\underline{z} - \hat{\underline{z}}), \qquad (4.49)$$

where  $\underline{z}$  is the measurement,  $\underline{\hat{z}}$  is the predicted measurement, and **S** is the innovation covariance. The term  $\underline{z} - \underline{\hat{z}}$  is called the innovation residual.

The design of the UKF makes it such that the consistency test using Eq. 4.49 always passes. Even though a system may be non-linear and non-Gaussian, the estimated covariance should still match the true covariance of the state estimate. This property of the estimate covariance provides a way to estimate the uncertainty of the system. The trace of the covariance matrix is a measure of the uncertainty in the system.

It has been shown that when using either the EKF or the UKF, the state estimate covariance can become over-confident [31, 30]. The effect of this will be that the trace of the covariance matrix will be smaller than it should truly be. This occurs because the observability matrix for SLAM should always be singular. As observations are made and the estimates are updated, the null-space of the observability matrix is lost. The over-confidence is a result of information being added to the filter that is not really there. It was shown by Huang et al. that for the type of model being used in this work, the extra information comes in the form of a translation and rotation of the entire map. This is important to note because of the effect it has on the filter. For the implementation provided, inconsistency is unavoidable. By showing that the maps generated do not exhibit obvious signs of the effects of the inconsistency, it can be assumed that the inconsistency is minor. This allows that observations made on the state estimate covariance are still valid.

# 4.2 Orientation Change Estimation

The change in estimation can be estimated by comparing the angles-of-arrival (AoA) between successive up-chirps and down-chirps. This is done to avoid possible alignment errors between up and down-chirps. Even after an alignment is found during the velocity estimation process, the differences in frequency mapping due to the velocity and AoA make the estimation of the combined change in AoA difficult. Up-chirp and down-chirp angle biases could also cause problems in estimating the change in AoA (though this has not been definitively shown to be a problem). Because the change in range between chirps of the same type is small, responses from reflectors tend to get mapped to the same frequencies between adjacent pulses. As an example, consider a platform moving at 10 m/s and a pulse repetition frequency of 100 Hz. Between adjacent up-chirps, the platform will only move 0.1 m. If the radar has a range resolution as low as 0.2 m, the reflector is still likely to be mapped to the same bin in the frequency response. In the experiment, the platform is moving at no more than 2.5 m/s, and the range resolution is not better than 0.8 m.

The change in orientation is estimated by finding the likely locations of reflectors in adjacent chirps. The AoAs at corresponding frequency bins which are likely to contain reflectors are subtracted from each other. The average difference is taken as the change in orientation.

#### 4.2.1 Frequency Bins of Likely Reflectors

A frequency bin is said to be likely to contain a reflector response if it meets 2 criteria: 1) the magnitude response at that bin has to be detected by a constant false alarm rate (CFAR) detector, and 2) the AoAs in that bin and the surrounding bins have to be similar. Because of the 2nd condition, the probability of false alarm for the CFAR detector can be kept higher, at around 10%.

The similarity, s(i), of the AoA in a frequency bin to the AoAs of neighboring frequency bins is given by

$$s(i) = \sqrt{\sum_{j=i-3, j \neq i}^{j=i+3} w(j) \left(A(i) - A(j)\right)^2},$$
(4.50)

where A(j) is the AoA at bin, j, and w is a smoothing function.

A threshold,  $\tau$ , is placed on the similarity such that bins less than the threshold are considered to contain the AoAs of reflectors. The bins are further refined by applying a logical AND to corresponding bins from the current measurement and the previous measurement as

$$a_k(i) = (s_k(i) < \tau) \land (s_{k-1}(i) < \tau)$$
(4.51)

where  $a_k(i)$  is the proposition that the AoA measurement in bin *i* at instance *k* contains a reflector.

A logical AND is then applied between the CFAR result,  $c_k$ , and  $a_k$  to determine which frequency bins are likely to contain reflectors with consistent angle measurements. Let  $I_k$  be the set of bins containing measurements, described by

$$I_k = \{i : (c_k(i) \land a_k(i)) = \text{TRUE}\}$$

$$(4.52)$$

### 4.2.2 Orientation Estimation

The orientation change is estimated by take the average difference in AoA between measurements as

$$\delta \gamma_{pk-1} = \frac{1}{|I_k|} \sum_{i \in I_k} A_k(i) - A_{k-1}(i)$$
(4.53)

where  $|\cdot|$  represents cardinality.

## 4.3 Velocity Estimation

#### 4.3.1 Down-Chirp Frequency Relation to Up-Chirp Frequency

For the following derivation, platform rotation is ignored. The terms down-chirp and decreasing frequency chirp are synonymous, as are the terms up-chirp and increasing frequency chirp. The frequency shift induced by a reflector at a range, r, angle,  $\alpha$ , moving at a velocity, v, with respect to the radar is

$$f_{up} = \frac{2}{c} \left( \frac{B}{T} r - f_c v \cos(\alpha) \right), \tag{4.54}$$

for an up-chirp and

$$f_{down} = \frac{2}{c} \left( \frac{B}{T} r + f_c v \cos(\alpha) \right), \tag{4.55}$$

for a down-chirp. In the above equations, c is the speed of light, B is the chirp bandwidth, T is the time of the chirp, and  $f_c$  is the base-band frequency. It is assumed that the platform motion is aligned with the angle-of-arrival axis, i.e. it is looking straight ahead and moving straight ahead. An example of the log magnitude responses of an up-chirp and a down-chirp are shown in Fig. 4.9. Notice that the magnitude responses, the main peaks in particular, are nearly identical, except for a shift in frequency.



Figure 4.9: The the normalized log magnitude response from a pulse. The up-chirp is blue and the down-chirp is red.

### 4.3.2 Range Over Time

The range to the object at time, t, is given by

$$r(t) = ((x - x_p(t))^2 + (y - y_p(t))^2)^{\frac{1}{2}}$$
  
=  $(x(t)^2 + y(t)^2)^{\frac{1}{2}},$  (4.56)

where (x, y) is the location of a reflector, and  $(x_p(t), y_p(t))$  is the sensor position at time t. As the platform moves according to the motion described above, the range to an object changes as

$$r(t + \tau) = ((x(t) - \tau v)^{2} + y(t)^{2})^{\frac{1}{2}}$$
  
=  $[(r(t)\cos(\alpha) - \tau v)^{2} + (r(t)\sin(\alpha))^{2}]^{\frac{1}{2}}$  (4.57)  
=  $(r(t)^{2} - 2\tau vr(t)\cos(\alpha) + \tau^{2}v^{2})^{\frac{1}{2}}.$ 

A first order approximation of this square root is

$$r(t+\tau) \approx r(t) - \tau v \cos(\alpha) + \frac{\tau^2 v^2}{r(t)}$$

$$\approx r(t) - \tau v \cos(\alpha).$$
(4.58)

The last term was dropped because  $\tau$  is expected to be on the order of  $1e^{-2}$  while r(t) is expected to be > 1, making its contribution to  $r(t + \tau)$  insignificant.

#### 4.3.3 Down-Chirp Relation to Up-Chirp with Range Change

When accounting for the change in range, the down-chirp then becomes

$$f_{down} = \frac{2}{c} \left( \frac{B}{T} (r - \tau v \cos(\alpha)) + f_c v \cos(\alpha) \right)$$
$$= f_{up} - \frac{2}{c} \frac{B}{T} \tau v \cos(\alpha) + \frac{4}{c} f_c v \cos(\alpha)$$
$$= f_{up} - \left( \frac{2}{c} \left( \frac{B}{T} \tau - 2f_c \right) \cos(\alpha) v \right).$$
(4.59)

#### 4.3.4 Magnitude and Angle of Arrival Response Descriptions

The up-chirp and down-chirp provide the magnitude responses,  $M_{up}(f)$  and  $M_{down}(f)$ , respectively, where f is a frequency. Using the relationship between the up-chirp and down-chirp and noting that the shift,  $\tau$ , is equal to the chirp duration, T, the magnitude responses can be related as

$$M_{up}(f) = M_{down}(f + \frac{2}{c}(B - 2f_c)\cos(\alpha)v).$$
(4.60)

Using the first order Taylor Series approximation, this can also be described as  $M_{down}(f + \frac{2}{c}(B - 2f_c)\cos(\alpha)v) \approx M_{down}(f) + \frac{\partial M}{\partial v}v$   $= M_{down}(f) + \frac{2}{c}(B - 2f_c)\cos(\alpha)\frac{\partial M}{\partial f}v.$ (4.61)

Likewise, there are angle-of-arrival responses,  $A_{up}$  and  $A_{down}$ , with the same relationship

$$A_{up}(f) = A_{down}(f + \frac{2}{c}(B - 2f_c)\cos(\alpha)v).$$
(4.62)

These angles of arrival are estimates of  $\alpha$  corresponding to each frequency.

### 4.3.5 Estimating Velocity

The goal is to estimate the v which minimizes the cost function

$$J = \sum_{i} \left[ M_{down}(f_i + \frac{2}{c}(B - 2f_c)\cos(\alpha)v) - M_{up}(f_i) \right]^2.$$
(4.63)

This function represents the sum-of-squared errors between the magnitudes of the signals. When the correct velocity is applied, in combination with the AoA effects, the signals should be identical, in the absence of noise.

The cost function is minimized using the Gauss-Newton method. By taking the derivative of the cost function with respect to v, one obtains

$$\frac{\partial J}{\partial v} = 2\sum_{i} \frac{2}{c} (B - 2f_c) \cos(\alpha) \frac{\partial M}{\partial f_i} \Big[ M_{down} (f_i + \frac{2}{c} (B - 2f_c \cos(\alpha))v) - M_{up}(f_i) \Big].$$
(4.64)

Substituting the approximation from Eq. 4.61, one obtains

$$\frac{\partial J}{\partial v} = 2\sum_{i} \frac{2}{c} (B - 2f_c) \cos(\alpha) \frac{\partial M}{\partial f_i} \Big[ M_{down}(f_i) + \frac{2}{c} (B - 2f_c) \cos(\alpha) v \Big] - M_{up}(f_i) \Big].$$
(4.65)

By setting Eq. 4.65 to 0 and rearranging, the incremental change in v,  $\delta_v$ , can be estimated from

$$\sum_{i} \left(\frac{2}{c}(B-2f_c)\cos(\alpha)\frac{\partial M}{\partial f_i}\right)^2 \delta_v = \sum_{i} \frac{2}{c}(B-2f_c)\cos(\alpha)\frac{\partial M}{\partial f_i} \Big[M_{up}(f_i) - M_{down}(f_i,v^*)\Big]$$
(4.66)

at each iteration, where  $M_{down}(f, v^*)$  is the shifted version of  $M_{down}$ , and  $v^*$  is the estimated velocity. The update from iteration k - 1 to k is then

$$v_k^* = v_{k-1}^* + \delta_v \tag{4.67}$$

#### 4.3.6 Alternative Velocity Estimation

An alternative cost function can be used to estimate the velocity, when AOA measurements are not available. It is expected that not accounting for the AoA would lead to an underestimate of the velocity. It also means that the shift between signals can be treated as constant for all points in the signal. The linear shift should enable a reduction in processing. The cost function can be approximated as

$$J = \sum_{i} \left[ M_{down} (f_i + \frac{2}{c} (B - 2f_c)v) - M_{up}(f_i) \right]^2.$$
(4.68)

The velocity estimation could then be made using the simplified search equation

$$\sum_{i} \left(\frac{2}{c}(B-2f_c)\frac{\partial M}{\partial f_i}\right)^2 v = \sum_{i} \frac{2}{c}(B-2f_c)\frac{\partial M}{\partial f_i} \Big[M_{up}(f_i) - M_{down}(f_i)\Big].$$
(4.69)

### 4.4 State Propagation With Radar Dead-Reckoning

The platform variables are the only variables that change during propagation. The platform state and state covariance are augmented with the dead-reckoning variables as

$$\underline{x}_{p}^{a} = \begin{bmatrix} \underline{x}_{p} \\ v \\ \gamma \end{bmatrix}$$
(4.70)

and

$$\mathbf{P}^{a} = \begin{bmatrix} \mathbf{P}_{pp} & \underline{0} & \underline{0} \\ \underline{0}^{T} & P_{vv} & 0 \\ \underline{0}^{T} & 0 & P_{\gamma\gamma} \end{bmatrix}, \qquad (4.71)$$

where  $\mathbf{P}_{pp}$  is the platform estimate covariance,  $P_{vv}$  is the variance of the velocity estimate, and  $P_{\gamma\gamma}$  is the orientation change estimate variance.

After the platform is propagated,  $\mathbf{P}_{k|k-1}$  can be found. The propagated platform covariance can be found according to Eq. 4.35. The covariances between the landmarks remain unchanged, with only the covariances between the platform and each of the landmarks left to be calculated. For the  $i^{\text{th}}$  landmark, the covariance between the platform and the landmark is updated as

$$\mathbf{P}_{pik|k-1} = \mathbf{P}_{pik-1|k-1}\mathbf{\hat{F}},\tag{4.72}$$

where  $\hat{\mathbf{F}}$  is estimated from Eq. 4.43.

### 4.5 Landmark Detection

#### 4.5.1 Quality Landmarks

A few criteria make a landmark useful for SLAM purposes. At the very least, it should be stationary and repeatedly observable. In order to be observed by the radar, the landmark should be metal. It is not expected that all landmarks should be identifiable in the video, but it is preferred. In order to meet this criteria, it should have a unique color from the background and have a cross-sectional area that occupies a minimum number of pixels in the images. The size requirement is determined by the camera and radar range limits. There are cases where it is preferable not to incorporate a reflector into the filter. If the reflector is only observable for a few measurements, it will not improve much on the platform estimate and have a large estimate uncertainty. At the same time, it will increase the computational load of the system. Moving objects that are not handled as such will also cause problems with the SLAM estimate.

### 4.5.2 Cell-Averaging Constant False Alarm Rate Detector

A Cell-Averaging Constant False Alarm Rate (CA-CFAR) detector is used on the magnitude response from each radar chirp as part of the detection process. For this, the radar chirp magnitude response is convolved with a kernel and thresholded. The kernel is an even symmetric function. The middle value corresponds to the frequency bin under test. Guard values in the kernel are set to 0 around the bin under test in order to prevent neighboring cells from disturbing the hypothesis test in the event that the reflector is mapped to multiple frequency bins. The kernel is determined by the expected noise in the radar channel and the desired probability for false alarm. After the convolution, returned values greater than 0 are taken as detections, and returned as  $m_{bdk}(f)$  for the down-chirp response, and  $m_{buk}(f)$  for the up-chirp response.

An example log magnitude response is given in Fig. 4.10. CA-CFAR was applied to the magnitude response with a probability of false alarm of 6.5%, 3 guard cells, and 10 averaging cells. The result is shown in Fig. 4.11. Further away from the radar, the detections are distinct. Close to the radar, there are a lot of false alarms. Specifically, the set of detections starting near 8 s and at a range around 15 m are undesirable.

An image of the corresponding AoAs over time is shown in Fig. 4.12. In regions corresponding to large log magnitude responses in Fig. 4.10, the change in AoA response appears continuous and smooth. In the noisy regions in Fig. 4.11, the AoA response changes more rapidly. Notice that the AoA response around the previously mentioned undesirable detections are rapidly changing over time. The goal is to use



Figure 4.10: Example log magnitude response.

the smoothness in the AoA to help determine which CA-CFAR detections should provide good measurements for the SLAM filter. Recall, allowing for a large fluctuation



Figure 4.11: CFAR result with  $P_{FA} = 0.065$ .



Figure 4.12: Example of the AoA response.

in angle measurement for the same landmark would degrade the platform location and map estimates.

### 4.5.3 AoA Smoothness Constraint

The smoothness constraint on the AoA is the same as the one used in Eq. 4.50. Let the smoothness of the  $k^{\text{th}}$  down-chirp AoAs be denoted by  $s_{dk}(f)$ , and for the up-chirp,  $s_{uk}(f)$ . In order to reduce false alarms, adjacent chirps are averaged such that

$$\hat{s}_{dk}(f) = \frac{s_{dk}(f) + s_{dk-1}(f)}{2}.$$
(4.73)

A threshold is then applied to  $\hat{s}_{dk}(f)$  to determine if the AoA meets the smoothness constraint as

$$\hat{s}_{bdk}(f) = (\hat{s}_{dk}(f) > \tau_A)$$
(4.74)

where  $\tau_A$  is the threshold on the AoA constraint. The same operation is done on  $\hat{s}_{uk}(f)$  to obtain  $\hat{s}_{buk}(f)$ .

#### 4.5.4 Possible Detections from Single Chirp

Once the CA-CFAR and the AoA smoothness constraint are applied, a logical AND is applied to  $m_{bdk}(f)$  and  $\hat{s}_{bdk}(f)$  for the down-chirps to obtain

$$B_{dk}(f) = m_{bdk}(f) \wedge \hat{s}_{bdk}(f). \tag{4.75}$$

The AND is similarly performed on  $m_{buk}(f)$  and  $\hat{s}_{buk}(f)$  for the up-chirps to obtain  $B_{uk}(f)$ . Groups of adjacent frequency bins where  $B_{dk}$  is TRUE are collected. It is expected that responses from reflectors will occupy a minimum number of bins, therefore, groups comprised of less than a minimum number of bins are removed. The center bin of each group is taken to be the location of the reflector. The set of possible reflector bins is described by

$$P_{dk} = \{ p_{d1k}, p_{d2k}, \dots, p_{dNk} \}, \tag{4.76}$$

where  $p_{di}$  is the center bin of the  $i^{\text{th}}$  group in  $B_{dk}$  and N is the number of groups.

#### 4.5.5 Detections from Adjacent Chirps

In order to be confirmed as a detection, the possible detection must be present for 2 consecutive chirps. Adding this constraint prevents many of the moving targets from generating false detections. This means that for a detection to be made from the  $k^{\text{th}}$  down-chirp, associations must be made between elements from the sets  $P_{dk}$  and  $P_{uk}$ . For the  $k^{\text{th}}$  up-chirp, associations should be made between elements of  $P_{uk}$  and  $P_{dk-1}$ . Associations are made by comparing the distances between the elements. Note that the frequency bins for a reflector will be shifted by the velocity of the platform. The shift due to the platform velocity was already estimated by Eq. 4.66. Twice this shift amount is subtracted from the distances between the elements. The distance between the *i*<sup>th</sup> detection in  $P_{dk}$  and the *j*<sup>th</sup> detection in  $P_{uk}$  is given by

where  $b_v$  is the induced shift in the number of bins by the platform velocity. A global nearest neighbor algorithm is used to perform the associations. Distances greater than a maximum threshold are removed as association possibilities. The associations are made such that the sum of the distances between the detections is minimal.

Every detection is associated with a landmark at some point in the filtering process. Detections which are not from a landmark currently in the filter will be retroactively associated with that landmark when it is added to the filter at a later step. Because of this, the associations between landmarks and elements of  $P_{uk}$  can be carried over to the elements of  $P_{dk}$  which are associated with those elements in  $P_{uk}$ . By carrying these associations over in the detection process, the search space for associating unmatched detections can be greatly reduced.

### 4.6 Radar Measurement Model

Once detections are found, they are converted to the form given by the radar measurement model. The measurement model for landmarks assumes that each measurement contains a range and an angle. The angle measurement,  $\alpha_{ik}$ , of the  $i^{\text{th}}$  landmark at instance k is

$$\alpha_{ik} = \arctan \frac{y_i - y_{pk}}{x_i - x_{pk}} - \theta_{pk}.$$
(4.78)

The radar measurement,  $r_{ik}$  of the same landmark is

$$r_{ik} = \sqrt{(x_i - x_{pk})^2 + (y_i - y_{pk})^2}.$$
(4.79)

The combined measurement is written as

$$\underline{z}_{ik} = \begin{bmatrix} r_{ik} \\ \alpha_{ik} \end{bmatrix} \tag{4.80}$$

### 4.7 Angle Estimation

For each detection, the AoAs in the group of frequency bins corresponding to that detection are collected. The median of the AoAs is taken as the AoA of that detection.

### 4.8 Range Estimation

The frequency bin,  $f_{ik}$  of a detection, from the set  $P_{dk}$  (or  $P_{uk}$ ), is a function of the range and the platform velocity. The range of the detection is calculated as

$$r_{ik} = \frac{c}{2} \frac{T}{B} (f_{ik} - f_{\text{Dopp}}),$$
 (4.81)

where  $f_{\text{Dopp}}$  is the Doppler shift obtained from the estimation in Eq. 4.66.

# 4.9 Unmatched Detection Association with Unmatched Landmarks

A global nearest-neighbor association is used on unmatched detections and landmarks to determine if any of the detections might belong to any landmarks. Since the deadreckoning provides a good estimate of the current radar location and orientation, only landmarks within the FOV of the radar are considered for association.

The measurements for the landmarks within the FOV of the radar are estimated from the propagated state according to Eqs. 4.78 and 4.79. The Mahalanobis distances between each of the measurements and the estimated measurements are calculated. The Mahalanobis distance,  $d_M(\underline{z}_i, \underline{x}_j)$ , between the  $i^{\text{th}}$  measurement and the  $j^{\text{th}}$  landmark is given by

$$d_M(\underline{z}_i, \underline{x}_j) = (\underline{z}_i - \underline{\hat{z}}_j)^T \mathbf{R}^{-1} (\underline{z}_i - \underline{\hat{z}}_j), \qquad (4.82)$$

where **R** is the radar measurement error covariance and  $\underline{\hat{z}}_{j}$  is the estimated measurement to be obtained from the  $j^{\text{th}}$  landmark. The associations are made similar to the method used in subsection 4.5.5.

### 4.10 Filter Update

The first step of the filter update is to extract only the state estimate and state estimate covariance matrix corresponding to the platform and the landmarks that

$$\underline{x}_d = \begin{bmatrix} \underline{x}_p \\ \underline{x}_2 \\ \underline{x}_5 \end{bmatrix}$$
(4.83)

and

$$\mathbf{P}_{d} = \begin{bmatrix} \mathbf{P}_{pp} & \mathbf{P}_{p2} & \mathbf{P}_{p5} \\ \mathbf{P}_{p2}^{T} & \mathbf{P}_{22} & \mathbf{P}_{25}^{T} \\ \mathbf{P}_{p5}^{T} & \mathbf{P}_{25}^{T} & \mathbf{P}_{55} \end{bmatrix}.$$
 (4.84)

Sigma points are generated from these and the steps in Eqs. 4.36- 4.38. The linearized measurement matrix,  $\hat{\mathbf{H}}_d$ , is then estimated for the extracted state according to Eq. 4.45. From the example above, measurement matrix would have the form

$$\hat{\mathbf{H}}_d = \begin{bmatrix} \hat{\mathbf{H}}_p & \hat{\mathbf{H}}_2 & \hat{\mathbf{H}}_5 \end{bmatrix}$$
(4.85)

where  $\hat{\mathbf{H}}_p$  corresponds to the measurement of the platform and  $\hat{\mathbf{H}}_i$  corresponds to the measurement of the  $i^{\text{th}}$  landmark. The linearized measurements matrix for the entire state,  $\hat{\mathbf{H}}$ , could then be estimated as

$$\hat{\mathbf{H}} = \begin{bmatrix} \hat{\mathbf{H}}_p & \mathbf{0} & \hat{\mathbf{H}}_2 & \mathbf{0} & \mathbf{0} & \hat{\mathbf{H}}_5 & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix},$$
(4.86)

where **0** represents a matrix of zeros corresponding to the measurement matrix of unobserved landmarks.  $\hat{\mathbf{H}}$  can then be used to update the entire state and state covariance as in the normal Kalman filter Eqs. 4.23 - 4.25.

An alternative method to update the filter, the filter could be updated one landmark at a time, with the process in Eqs. 4.25-4.83 performed each time. The benefit of this method is a reduction in the size of the Cholesky decomposition, whose computational costs are generally cubic in the size of the state. A drawback of this method is that it neglects the correlations between the landmarks in the state when generating the measurement matrix. Because the expected number of measurements obtained in each observation is less than 10, the cost of these calculations is relatively small. It was determined that the computational cost was not detrimental enough to warrant the separate processing. Another alternative method to the state update is to not update the entire state, but only the landmarks within a maximum distance of the current observations. This can work because landmarks that are not observed together will not develop high correlations between each other. This method was not chosen here again because of the relatively small number of observations. The computational cost to update the state grows quadratically with the size of the state. When there is a large number of landmarks observed over time, this can become necessary.

## 4.11 New Landmark Addition

New landmarks are appended to the filter by estimating the position of the landmarks in the map. This is done by first extracting the platform portion of the state estimate covariance and appending the measurement covariance to it to make

$$\mathbf{P}_{a} = \begin{bmatrix} \mathbf{P}_{pp} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}.$$
(4.87)

A state estimate is formed as

$$\underline{x}_a = \begin{bmatrix} \underline{x}_p \\ 0 \\ 0 \end{bmatrix} \tag{4.88}$$

Sigma points are formed for this state estimate as in Eq. 4.31. The mapping function

$$\underline{x}_{N+1} = \underline{x}_p + r \begin{bmatrix} \cos(\alpha - \theta_p) \\ \sin(\alpha - \theta_p) \end{bmatrix}$$
(4.89)

is applied to the sigma points and they are collected as in Eqs. 4.37 and 4.38 to obtain  $\underline{x}_{N+1}$  and  $\mathbf{P}_{N+1,N+1}$ .  $\underline{x}_{N+1}$  is appended to the state and  $\mathbf{P}_{N+1,N+1}$  is appended to the state estimate covariance.

# 5 Radar Video Fusion SLAM

The main benefit of adding video measurements to radar SLAM is the improvement in angle measurements. A drawback of using imagery is the added computational complexity. For this reason, 2 methods of fusion are employed. The first method is used to take advantage of the improved bearing estimates for navigation purposes while attempting to keep computational complexity low. The second method is intended to improve estimates of landmarks at an added cost of computational complexity. Even in the second method, however, the search space for features is limited to the few radar reflectors visible in the images.

## 5.1 Angle Change Estimation From Video

The first method is to use the video to estimate the change in platform orientation from one frame to the next. Because the platform is assumed to move on a plane, rotations are expected to mostly occur about the axis perpendicular to the plane. A correlation between images could be performed to estimate the change in angle from frame to frame. In addition, a reduction in processing can be made if the images are summed before the correlation is performed. This is done for 2 main reasons. Translations should only occur along 1 axis, so this reduces the number of computations in the correlation process. Acting on a sum of pixels is more robust to slight out of plane translations or rotations that may occur which are not along the axis perpendicular to the plane.

### 5.1.1 Image Summation

It is assumed the image plane is perpendicular to the x-z plane shown in Fig. 3.4. It is also assumed that  $\underline{d}$ , shown in Fig. 3.4, is parallel to the image plane. The first step

in the orientation change estimation is to sum the image along the z-axis to obtain

$$\iota_l(x) = \sum_z I_l(x, z), \tag{5.90}$$

where I(x, z) is an index to the pixel at location (x, z) in the image at instance l. When comparing 2 images using this method, there will be some errors introduced by this summation. In particular, these are due to the non-linear effects of the image process and radial distortion. These errors have not been found to noticeably interfere with the correlation process, however.

### 5.1.2 Pixel-Space to Angle-Space Conversion

While each pixel is typically treated as a sample along a Cartesian space, it can also be treated as a sample in a spherical coordinate space as well. For this method, the pixel x-coordinates are converted to angle coordinates according to

$$\Theta(i) = \arctan(\frac{x(i) - p_x}{\alpha_x}), \qquad (5.91)$$

where x(i) is the pixel-space value of x in the  $i^{\text{th}}$  column of the image,  $\Theta(i)$  is the corresponding angle value,  $p_x$  is the location of the principal point along the x-axis, and  $\alpha_x$  is the focal length.

After the angles are calculated,  $\iota_l(\Theta)$  is resampled such that the new values of  $\Theta$  are spaced equally. While the next step could be performed without resampling, equal spacing simplifies calculations and processing.

### 5.1.3 Correlation Between Image Summations

By describing the summed images as a functions of angles, a shift between the 2 summed images represents a rotation. The relation between  $i_l$  and  $i_{l-1}$  is described by

$$\iota_l(\Theta) = \iota_{l-1}(\Theta + \gamma_I). \tag{5.92}$$

This was done under the assumption that the translation that the platform undergoes between image samples is minor. A first order Taylor series approximation yields

$$\iota_l(\Theta) \approx \iota_{l-1}(\Theta) + \gamma_I \frac{\partial \iota_{l-1}}{\partial \Theta}$$
(5.93)

The correlations between angles is estimated using the Gauss-Newton method. The cost function is

$$J_{I} = \sum_{\Theta} \left( \iota_{l}(\Theta) - \iota_{l-1}(\Theta) - \gamma_{I} \frac{\partial \iota_{l-1}(\Theta)}{\partial \Theta} \right)^{2}.$$
 (5.94)

Taking the derivative with respect to  $\gamma_I$  yields

$$\frac{J_I}{\partial \gamma_I} = -2\sum_{\Theta} \left( \iota_l(\Theta) - \iota_{l-1}(\Theta) - \gamma_I \frac{\partial \iota_{l-1}(\Theta)}{\partial \Theta} \right) \frac{\partial \iota_{l-1}(\Theta)}{\partial \Theta}.$$
 (5.95)

The update to  $\gamma_I$  at the *i*<sup>th</sup> iteration is then

$$\gamma_{Ii} = \gamma_{Ii-1} + \frac{\alpha_n}{\alpha_d},\tag{5.96}$$

where

$$\alpha_n = \sum_{\Theta} \left( \iota_l(\Theta) - \iota_{l-1}(\Theta) \right) \frac{\partial \iota_{l-1}(\Theta)}{\partial \Theta}$$
(5.97)

and

$$\alpha_d = \sum_{\Theta} \left( \frac{\partial \iota_{l-1}(\Theta)}{\partial \Theta} \right)^2.$$
(5.98)

# 5.2 Angle Change Estimation Update to Filter

Once the orientation change between instances l and l - 1, the platform orientation can be updated. The measurement rates for the radar and video are different, but it is assumed that the current frame being processed was observed at the same time instance as the last radar pulse to be processed. It is also assumed that the measurements from the last radar pulse have been processed. With these assumptions, only an update needs to be performed, and not a propagation. To account for this, the change in orientation is adjusted as a change from the current orientation as

$$\theta_I = \gamma_I - \theta_{pl} + \theta_{pl-1}. \tag{5.99}$$

The measurement is then treated as a direct observation of the current platform orientation. This is a linear process, so the update process can be performed using Eqs. 4.24 - 4.26 from the standard Kalman filter, where

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}. \tag{5.100}$$

# 5.3 Reflector Segmentation

It is assumed that the color of the reflector is unique with respect to its surroundings and the reflector consists of continuous, smooth surfaces. Based on this, a color-based segmentation and identification scheme is used for finding the reflector.

If the reflector for a landmark has not been identified, and the landmark has been observed by the radar between the last frame and the current frame, a search is performed for the reflector corresponding to that landmark in the image. This requires that measurements between the last frame and the current frame be collected and stored according to which landmark they belong.

# 5.4 Weight Image Generation

For each reflector to be found, a weight image is generated. The weight image is the same size as the video image. Each pixel in the weight image represents the likelihood that the image of reflector is contained in the corresponding pixel in the video image. The weight image is generated based on the idea that each measurement creates a potential field, F, of probability in 3D space that the reflector is at some location in the 3D space.

For a single reflector, an equipotential surface on this field would be in the shape of a lens. Consider the case where the origin of the field is located at the radar center, with the radar beam centered along the x-axis, and the radar azimuth measurement is aligned with the spherical coordinate frame azimuth (centered at x = 0). The magnitude of the field at a location where  $\rho$ ,  $\alpha$ , and  $\psi$  provide the spherical range, azimuth, and elevation coordinates, respectively, can be given by

$$F(\rho, \alpha, \psi) = k \exp\left(-\frac{(\rho - \rho_m)^2}{2\sigma_\rho^2} - \frac{(\alpha - \alpha_m)^2}{2\sigma_\alpha^2} - \frac{\psi^2}{2\sigma_\psi^2}\right),$$
 (5.101)

where  $\sigma_{\rho}^2$  is the radar range measurement variance,  $\sigma_{\alpha}^2$  is the radar azimuth measurement variance, and  $\sigma_{\psi}^2$  is determined by the radar beam width. The scale factor is

$$k = (2\pi)^{-\frac{3}{2}} (\sigma_{\rho} \sigma_{\alpha} \sigma_{\psi})^{-1}.$$

Assuming the camera and radar are co-located with the same viewing axis and orientation, a weight image,  $\mathbf{W}(\underline{x})$ , is defined as the integral of  $F(\rho, \alpha, \psi)$  along the line joining the world origin with the point  $\underline{x}'$  on the plane at x = 1. The variable  $\underline{x}$  represents a pixel location. The intrinsic camera matrix transforms  $\underline{x}'$  to  $\underline{x}$ . If the camera were not located at the origin, the integral would be along the line joining the camera center to the point  $\underline{x}'$  on the plane orthogonal to the camera optical axis and located one unit distance from the camera center.

Consider the case of projecting the 3D Gaussian integration onto the unit sphere. The integral would then be

$$\mathbf{W}(\alpha,\psi) = \int_{-\infty}^{\infty} F(\rho,\alpha,\psi)\partial\rho$$
  
=  $(2\pi\sigma_{\alpha}\sigma_{\psi})^{-1} \exp\left(-\frac{(\alpha-\alpha_m)^2}{2\sigma_{\alpha}^2} - \frac{\psi^2}{2\sigma_{\psi}^2}\right),$  (5.102)

and the result would be a 2D Gaussian. If the unit sphere result were then projected onto the plane at x = 1, the projection would be approximately Gaussian. This approximation method will be inadequate when the radar and camera are not colocated. The projection would not have any symmetry about the mean.

This approximation can be made by treating the projection as a measurement function and applying Eqs. 4.36-4.38 to the radar measurements. Before Eq. 4.36 can be applied, the radar measurement must first be projected to 3D space.

For a measurement,  $\underline{m} = [r, \alpha, 0]$ , with covariance matrix,  $\mathbf{M} = \text{diag}\{[\sigma_r^2, \sigma_\alpha^2, \sigma_\psi^2]\}$ , the covariance matrix is decomposed as

$$3\mathbf{M} = \mathbf{S}^T \mathbf{S}.\tag{5.103}$$

Sigma points,  $\underline{\zeta}_i$ , are given by

$$\mathbf{Z} = \begin{bmatrix} \underline{\zeta}_0 & \underline{\zeta}_1 & \cdots & \underline{\zeta}_6 \end{bmatrix} \\ = \begin{bmatrix} \underline{m} & \underline{m} & \cdots & \underline{m} \end{bmatrix} + \begin{bmatrix} \underline{0} & \mathbf{S} & -\mathbf{S} \end{bmatrix}.$$
(5.104)

The sigma point are converted from spherical coordinates to Cartesian coordinates in the platform frame and then projected onto the image plane as

$$\nu_i \underline{x}_i = \mathbf{K} \underline{X}_i \tag{5.105}$$

where **K** describes the intrinsic camera parameters,  $\underline{X}_i$  is the Cartesian coordinate representation of  $\underline{\zeta}_i$ , and  $\nu_i$  is the last element in  $\mathbf{K}\underline{X}_i$ . Note that this assumes the radar and camera are collocated and aligned. The mean and covariance of the projection are then given by

$$\underline{x} = \frac{1}{7} \sum_{i=0}^{6} \underline{x}_i \tag{5.106}$$

and

$$\mathbf{P} = \frac{1}{7} \sum_{i=0}^{6} \underline{x}_i \underline{x}_i^T \tag{5.107}$$

respectively. In this case, a uniform weight is applied to the sigma points.

When considering multiple detections from one landmark, the projection of the  $i^{th}$  measurement yields a mean image location,  $\underline{p}_i$ , and covariance,  $\mathbf{P}_i$ . The weight image from each measurement is obtained as

$$\mathbf{W}_{i}(\underline{x}) = \frac{1}{2\pi\sqrt{|\mathbf{P}_{i}|}} \exp\left\{\frac{1}{2}(\underline{x}-\underline{p}_{i})^{T}\mathbf{P}_{i}^{-1}(\underline{x}-\underline{p}_{i})\right\}.$$
(5.108)

All of the weight images are combined as

$$\mathbf{W}(\underline{x}) = \frac{1}{w} \sum_{i} \mathbf{W}_{i}(\underline{x}), \qquad (5.109)$$

where w normalizes the maximum value in  $\mathbf{W}$  to 1.

An example of the projection of measurements and their covariances onto an image is shown in Fig. 5.13. It is easier to see the orange trihedral that caused the detections on the right in Fig. 5.14. Another orange trihedral cause the detections on the left,


Figure 5.13: An example of radar measurements projected onto an image. The measurements are green. The radar estimate of the reflector is red.



Figure 5.14: The image without the radar projections. The detections are from orange trihedrals.



Figure 5.15: An example weight image generated by combining weight images from measurements.

but it is out of frame. The weight image generated by combining the weight images of each of the measurements is shown in Fig. 5.15. Only the measurements from the reflector on the right were used to generate the image. This is because a weight image is generated for only one landmark at a time.

#### 5.4.1 Color Space Reduction

In order to simplify the descriptor of the reflector, and simplify computations, the color space of the image is reduced. An RGB image typically has three 8-bit channels, one channel each for red, green, and blue.

The colors are first converted from RGB to rgs color space, where r = R/(R + G + B), g = G/(R + G + B), and s = (R + G + B)/3. The R, G, and B represent the red, green, and blue components of the color, respectively. Only the r and g elements are considered for identifying colors as the corresponding s relates to the intensity of

an (r, g) color. The range of values r and g can take is scaled to make them 4-bit, taking on values from 0 to 15. The conversions are described by

$$r = \left\lfloor \frac{48R}{R+G+B} \right\rfloor - 1 \tag{5.110}$$

and

$$g = \left\lfloor \frac{48G}{R+G+B} \right\rfloor - 1. \tag{5.111}$$

## 5.4.2 Average Color Likelihood

The set of discrete values that (r, g) can take make up the lattice **C**. The color of a pixel,  $\underline{x}$ , in rg space is denoted by  $C(\underline{x})$ . For an observed (r, g) pair, or color,  $z^c$ , the probability of  $z^c$  being a color of the reflector is estimated as

$$Pr(z^{c}|\mathbf{O}) = \mathbf{C}(z^{c})$$
$$= \frac{1}{n_{c}} \sum_{i \in \underline{x}_{c}} \mathbf{W}(i)$$
(5.112)

where  $\underline{x}_c = {\underline{x} | C(\underline{x}) = z^c}$ ,  $n_c$  is the cardinality of  $\underline{x}_c$ , and **O** is the set of measurements of the landmark.

Once the likelihoods of each color are found, a new image, L, is formed wherein the color of each pixel is replaced by the average likelihood of that color (i.e. if  $C(\underline{x}) = z^c$ , the color,  $z^c$ , then  $L(\underline{x}) = Pr(z^c | \mathbf{O})$ ). An example is shown in Fig. 5.16. The small area of bright pixels on the right corresponds to an area of the image whose colors are likely to belong to the reflector. In this case, they correspond to the orange trihedral. There are also a few bright pixels in the middle of the image. They correspond to another orange reflector.

#### 5.4.3 Possible Reflector Collection

The image L is then thresholded to determine the pixels to which the reflector is likely mapped as

$$B(\underline{x}) = (L(\underline{x}) > \tau_L) \tag{5.113}$$



Figure 5.16: An example reflector likelihood image, L. The original image is shown in Fig. 5.14.

An image closing operation is performed on B to form  $B_c$ . The connected components in  $B_c$  are collected. Let  $C_B$  be the set of connected components in B (after the closing operation). Each group in  $C_B$  consists of the indices to the pixels belonging to that group.

Size constraints on the expected size of the reflector can be applied at this point. The maximum and minimum amount of cross-area the reflector is expected to have are given by  $a_{\text{max}}$  and  $a_{\text{min}}$ , respectively. Each pixel occupies an amount of area on the image plane denoted by  $a_p$ . An estimate of the range between the camera and the landmark is known from the radar measurements. An estimate of the cross-area that each group in  $C_B$  would occupy is given by

$$a_C(i) = r | C_B(i) | a_p,$$
 (5.114)

where r is the range to the landmark and  $|C_B(i)|$  represens the cardinality of the  $i^{\text{th}}$ 



Figure 5.17: A binary image corresponding to possible reflectors after all possible reflectors are collected.

group in  $C_B$ . Groups for which  $a_C$  does not fall between  $a_{\min}$  and  $a_{\max}$  are discarded.

In the event that more than one group remains after the size constraint is applied, the group containing the pixel with the highest weight in W is chosen. This is equivalent to saying the connected component closest to the centroid of the measurement projections is chosen.

An example is shown in Fig. 5.17. In this case, the bright pixels in the middle of Fig. 5.16 were eliminated by the size constraints.

#### 5.4.4 Reflector Identification

If a group in  $C_B$  was found that meets the constraints, the set of pixels which make that group are assumed to contain the mapping of the reflector. That group is called  $C_B^*$ . The rg colors for that reflector are then extracted. Because of the closing operation, some pixels in that group may contain colors that are not unique to the Figure 5.18: Image of the detected reflector.

0

reflector. In order to keep with the uniqueness assumption, only the pixels in  $C_B^*$  that were also TRUE in *B* are used for reflector color identification. The set of colors associated with the *i*<sup>th</sup> landmark is denoted as  $L_{ci}$ .

The image of the reflector detected in the example is shown in Fig. 5.18. Note that not all of the TRUE pixels in Fig. 5.17 contribute to this image. Only the pixels that were TRUE in that image and also TRUE after thresholding the image shown in Fig. 5.16 (before image closing) contribute to the reflector segmentation. A histogram of the colors corresponding to the detected reflector is shown in Fig. 5.19.

# 5.5 Reflector Tracking

When a new image is obtained, landmarks that are in the FOV and whose reflectors have been identified can be tracked. A color feature tracker is employed to find the AoA to the reflector in the image.

#### 5.5.1 Color Filtering

The first step is to determine which colors in the image belong to the landmark being tracked. A binary image,  $B_L$ , is created whose values are TRUE for pixels whose colors match those of the reflector. A closing operation is then performed on  $B_L$ .

The next image in the video sequence from the one shown in Fig. 5.14 is shown in Fig. 5.20. The result after filtering for the color of the reflector is shown in Fig. 5.21. After closing, the result is nearly identical to the one shown in Fig. 5.18. The collection process is nearly identical to the process used in the segmentation. The absence of a weight image, however, requires that another method is used when multiple objects are detected.

# 5.5.2 Possible Detection Collection

Connected components are collected from  $B_L$ . Size constraints on the connected components are applied as in Sec. 5.4.3. The set of connected components is denoted  $C_L$ . Again, each group in  $C_L$  is a set of pixel locations corresponding to the connected



Figure 5.19: 2D color histogram of the segmented reflector.



Figure 5.20: The next image in the video sequence after the image shown in Fig. 5.14.



Figure 5.21: Pixels in the image whose color matches the reflector color.

components that make that group. It is possible that multiple connected components can meet the size constraints. Generating a weight image can be costly and should be avoided if possible. Instead, the angle to each connected component is estimated and used to estimate which connected component should be the landmark. The angle closest to the expected angle to the landmark is taken as the measurement. It should be noted that if the difference between the measured angle in the image and the expected angle is greater than a predetermined threshold, the measurement is discarded.

#### 5.5.3 Reflector Measurement From Image

For a connected component,  $C_L(i)$ , the x-direction centroid of the pixels is calculated by

$$\bar{x}_L(i) = \frac{1}{|C_L(i)|} \sum_j C_{Lxi}(j), \qquad (5.115)$$

where  $C_{Lxi}(j)$  is the x-coordinate of the  $j^{\text{th}}$  pixel location in  $C_L(i)$ . The angle to that centroid is then calculated as

$$\alpha_L(i) = \arctan\left(\frac{\bar{x}_L(i) - p_x}{\alpha_x}\right),\tag{5.116}$$

where  $p_x$  and  $\alpha_x$  are the same as in Eq. 5.91. By obtaining image measurements in this way, and because of the radar and camera configuration, the measurement model for landmarks in the images is the same as the AoA measurement model for the radar.

# 5.6 Video Landmark Measurement Update to Filter

Once all of the measurements of landmarks in the video are taken, the filter can be updated. As with the orientation change measurement, no platform propagation is necessary. The latest image in the video sequence is assumed to be observed at the same time as the latest radar pulse observation.

#### 5.6.1 Filter Update

The update process from image measurements is the same as the update process from radar measurements. The measurement model for the images is the same as in Eq. 4.78. As in the radar example in Eqs. 4.83 and 4.84, a truncated state and state estimate covariance consisting of only the platform and the measured landmarks is used to generate sigma points. The sigma points are propagated through the measurement model and an estimate of the linearized measurement matrix,  $\hat{\mathbf{H}}_d$ , is made. The linearized measurement matrix for the entire state is estimated from  $\hat{\mathbf{H}}_d$ , and the update can be completed using Eqs. 4.23 - 4.25.

# 6 Experiment

# 6.1 Description

An experiment was performed in which a radar and camera system was mounted on a cart and pushed along an outdoor path in the shape of a triangle. A map of the scene is shown in Fig. 6.22. The path, marked in blue, was traversed in the clockwise direction. Starting at the upper right vertex and moving to the bottom vertex, the distances around the path were 18.3 m, 22.9 m, and 26.5 m. Aluminum trihedrals, approximately 6 in. per side, were placed in the scene. The locations of the trihedrals are marked with yellow dots on the map. Other reflective objects, such as light poles and trees, in the scene are marked in red. The trihedrals were used



Figure 6.22: Scene layout for the experiment.

to guarantee strong reflectors would be detected in the radar domain. Along with stationary objects, people were observed during the collection process as they walked through the area. A measurement was not taken, but it was quite windy on that day, causing many of the reflectors to sway.

Two sets of data were collected and used for the experiment. For the first set, a 12 mm lens was attached to the camera. The field of view for this system was 21.5°. The path was traversed approximately 4.5 times over the span of approximately 6.5 minutes. A 6 mm lens was attached to the camera for the second collection. The field of view for this system was 42°. The path was traversed approximately 5 times over the span of approximately 6.5 minutes for that collection. The radar had a field of view of 50°. Radar measurements beyond a 40° field of view were found, however, to be unreliable.

The radar and camera were approximately co-boresighted, with the radar placed



Figure 6.23: The radar waveform.

above the camera as they sat on the cart. Recall, a picture of the sensors was shown in Fig. 3.3. Because the displacement is so small compared to the ranges to landmarks, the displacement is ignored during calculations. In addition to the sensors used for this effort, a Continuous Wave (CW) radar was mounted on the platform. The results from that sensor only provide a visual comparison to the FMCW estimates.

One period of a radar pulse consists of an increasing frequency chirp, followed by a decreasing frequency chirp. The base frequency of the chirp was 24 GHz. The bandwidth was 180 MHz. The pulse repetition frequency was 100 Hz. The radar waveform is shown in Fig. 6.23. An image was taken once every 7 radar pulses. This made the video frame-rate approximately 14.29 Hz.

# 6.2 Odometry Results

#### 6.2.1 Velocity Estimation

Three correlation-based velocity estimation methods were tried. The first method was implemented by finding the shift between adjacent rising and falling chirps. The second method was the Gauss-Newton optimization method described in Sec. 4.3.6. The estimates by that method should be improved on the first method by allowing for non-integer frequency bin shifts. The third method incorporates the AoA measurements in the optimization process. It was described in Sec. 4.3.5.

The velocity of the cart was not recorded during the collection, making a direct comparison of the estimates to the true value impossible. Alternative metrics are possible, however. All of the methods attempt to find a shift that minimizes the difference between frequency responses of adjacent chirps. The sum of squared differences (SSD) between the shifted frequency responses provides a way to measure how well each method is matching the responses. Another way to measure the accuracy of the estimates is to estimate the distance travelled over time based on the velocity measurements. The cart moved in straight lines along the path. The distances

Method	Average SSD	ReductionOver	ReductionOver
		Method 1 (%)	Method 2 $(\%)$
1	0.6577		
2	0.5947	9.71	
3	0.5939	9.58	0.14

Table 6.1: Average SSDs for velocity estimation methods.

on each side of the path are known. Integrating the velocities along the path and multiplying the result by the time travelled can provide another measure of estimate accuracy.

The SSD results are shown in Table 6.1. Both of the optimization algorithms provide a nearly 10% reduction in average SSD. The optimization that accounted for shifts due to the AoA performed negligibly better. A major contributing factor to the errors were changes in the magnitude response shape. The magnitude response is a function of many variables. No matter how accurately the velocity and AoA are estimated, changes in other variables will cause the magnitude responses to differ. Another factor is the limited range of AoA. For an AoA of 20°, the Doppler shift is reduced by less than 7%. A noteworthy improvement might be made for larger velocities and/or larger AoAs.

Examples of matched log magnitude responses for the 3 methods are shown in Figs. 6.24-6.26. The original responses were shown in Fig. 4.9. All of the shifted responses appear much more closely aligned than the original. With the peaks aligned, the major contributors to the SSD are easily seen.

Examples of the velocity estimates overlaid onto the magnitude response of the CW radar are shown in Figs. 6.27 and 6.28. Recall that the CW radar data was collected for visual comparison. The estimates from the  $2^{nd}$  and  $3^{rd}$  methods are nearly identical. Fig. 6.28 illustrates strengths of the methods. There appears to be a sinusoid along the main detection in the CW image. That is due to the reflectors swaying with the wind. The estimates are somewhat affected by this, but because

the reflectors are not all moving at the same speed, they are not affected by one particular reflector. As will be seen later, there were many instances where only one landmark was detected for use in updating the filter. If that reflector happened to be one of the swaying reflectors, large errors would be introduced in the platform location and map estimates. A second strength of this method is illustrated in its ability to estimate the velocity in the presence of moving objects. Recall that the timing for this corresponds to the log magnitude response over time shown in Fig. 3.6. The second, more negative, response is from a person moving towards the platform. The person induces a strong response, but the estimates are not affected.

Tab. 6.2 shows the results of integrating the velocity estimates over each of the sections of the path. The correlation method consistently underestimated the velocity, causing the larger errors. The other estimation methods produced much better results. As expected, there was a tendency to underestimate the true platform velocity,



Figure 6.24: A matched log magnitude responses from a pulse using the shift with the maximum correlation. The up-chirp is blue and the down-chirp is red.



Figure 6.25: A matched log magnitude responses from a pulse using Gauss-Newton optimization. The up-chirp is blue and the down-chirp is red.



Figure 6.26: A matched log magnitude responses from a pulse using Gauss-Newton optimization accounting for AoA . The up-chirp is blue and the down-chirp is red.

leading to underestimates of the path lengths.

# 6.2.2 Angle Change Estimation

A plot of the accumulated estimated changes in angle are shown in Fig. 6.29. The plot shows the radar is estimated to be turning to the right, as it did. The angle estimates should not change as much, however, during times when the platform was moving along a straight line.



Figure 6.27: Example 1 of velocity estimates overlaid on the CW log magnitude response. Method 2 estimates are blue and method 3 estimates are red.



Figure 6.28: Example 2 of velocity estimates overlaid on the CW log magnitude response. Method 2 estimates are blue and method 3 estimates are red.



Figure 6.29: Example of angle change estimates. The combined estimate is in black. The up-chirp and down-chirp only estimates are in blue and red, respectively.

-				
	Error (m)	0.0	0.47	0.26
NITO DEDITO	Method 3 (m)	19.20	22.39	25.86
0 111 CINIT	Error (m)	0.31	0.65	0.15
	Method 2 (m)	18.89	22.21	25.45
accounting for	Error (m)	1.01	2.03	1.56
D	Method 1 (m)	18.19	20.83	24.04
	True length (m)	19.20	22.86	25.60
0	Path section		2	3

is red. Ε Ē • 4+ h 4 7 4 د 4 4+ þ Table 6.2:



Figure 6.30: The platform trajectory from dead-reckoning estimates.

## 6.2.3 Platform Trajectory From Dead-Reckoning

Fig. 6.30 shows the estimated radar trajectory using only the velocity and angle change estimates, accumulated over time. It appears that the angle change estimates would allow for a loop closing if the velocity estimates were better. That is not actually the case. There appears to be a drift in the angle estimates in favor of a right turn. During the actual turns, the change in angle was actually underestimated. There was one troublesome turn in particular. The reflectors were not spaced closely enough in angle and, for a brief time, no reflectors are visible to the radar. It was estimated that there were no changes in orientation while no reflectors were visible.

# 6.3 Radar SLAM

# 6.3.1 Detections

Fig. 6.31 shows the application of the angle smoothness constraint. This corresponds to the section of the AoA response shown in Fig. 4.12. The detections after thresholding are shown in Fig. 6.32. Most of the noisy measurements at closer ranges were removed, compared to what was shown in the CA-CFAR result in Fig. 4.11.

The resulting detection image is shown in Fig. 6.33. The resulting image appears very similar to the image in Fig. 6.32. The differences arise where the angle response is smooth even though a reflector is not present.

The corresponding decreasing frequency chirp image is shown in Fig. 6.34. Recall that in order for a detection to be made, it must be seen in adjacent increasing and decreasing chirp responses. The actual detections are plotted on the binary image in Fig. 6.35.

In order to test the quality of the measurements with the detection constraints presented, the detections from CA-CFAR and CA-CFAR with the AoA constraint were used to perform SLAM and the results were compared. Because of the landmark observability issues in the experiment, fusion with video was applied to enable loop



Figure 6.31: A result after applying the angle smoothness constraint.



Figure 6.32: Angle smoothness based detections after threshold.



Figure 6.33: Result from combining CFAR and angle constraint to the increasing frequency chirp observations.



Figure 6.34: Result from combining CFAR and angle constraint to the decreasing frequency chirp observations.



Figure 6.35: Actual detections from decreasing frequency chirp observations.

closing. The detection process is independent of the SLAM process, so the only effect the fusion had was to improve the platform orientation estimate and enable loop closing.

Both of the resulting maps are shown in Fig. 6.36. The angle constrained CA-CFAR landmarks are magenta and the CA-CFAR landmarks are cyan. The CFAR probability of false alarm was set to 0.0025 for both cases. Because of the additional constraints, the CA-CFAR alone returns detections. In this case, there were 39 landmarks from the CA-CFAR detection process versus 23 for the constrained detection process. There were a total of 29386 detections from the CA-CFAR method, resulting in an average of 753.5 measurements per landmark. There were 20881 detections and an average of 907.8 measurements per landmark when using the constrained detection method. Of the extra 16 landmarks, 4 of them were repeated landmarks. Of the 12 remaining landmarks, 7 can be identified on the map as belonging to a specific object in the scene. All of the 23 landmarks from the constraint map are identifiable in the image of the scene. This means that the additional constraints helped ensure the selected landmarks were more stable.

Another way of comparing the quality of the measurements by comparing the quality of the estimates generated by using those measurements. The covariance associated with each state estimate from the SLAM filter provides a measure of the uncertainty of each measurement. The trace of the covariance matrix provides a measure of the uncertainty of the platform and landmark estimates. In the example given above, the trace of the covariance matrix for the CA-CFAR estimate was **19.19** m<sup>2</sup>. The trace of the covariance matrix for the angle constrained estimate was **4.81**. This is not an appropriate comparison, however. The smaller state can reasonably be expected to yield a smaller estimate uncertainty.

A more appropriate comparison can be made by increasing the number of landmarks obtained using the angle constrained method. By increasing the probability



Figure 6.36: Maps for comparison of radar detection methods. The CA-CFAR map is cyan. The angle constrained CA-CFAR map is magenta.

of false alarm to 0.065, 40 landmarks are tracked. The average number of detections per landmark was 1076.8. The map showing the measurements from both detectors is shown in Fig. 6.37. In this case, the trace of the covariance matrix for the system using the angle-constrained detector was  $7.25 \text{ m}^2$ .

When attempting to map the scene using the regular CA-CFAR with a probability of false alarm of 0.065, the map size quickly grew, with many incorrect associations, and became inconsistent.

## 6.3.2 Filtering

The radar measurements were estimated to have standard deviations of 0.5 m for the range and  $4^{\circ}$  for the AoA. The standard deviations for the velocity and platform change estimates were 0.5 m/s and 0.15°, respectively.

Applying the filter improved the trajectory estimation and provided a map of

reflector locations in the scene. The trajectory and the map do have some errors, however. The lack of observable landmarks on one of the turns was still an issue. The section of data used for this example started at the beginning of the first collection and continued for 120 sec. The map for one loop is shown in Fig. 6.38. As a comparison, the estimated trajectories of the platform using dead-reckoning only and radar SLAM are shown in Fig. 6.39.

The path is followed much more closely before and after the first turn. Part of the second turn is unobservable, leading to the divergence from the path. For part of the second turn, a building is less than 5 m the radar. There is effectively no change in AoA for the building, and there is nothing else to detect in the scene. A loop closing does not occur after the third turn, and some of the previous landmarks are detected as new landmarks. By the end of the experiment, 39 landmarks are mapped.



Figure 6.37: Maps for comparison of radar detection methods when a similar number of landmarks are mapped. The CA-CFAR map is cyan. The angle constrained CA-CFAR map is magenta.



Figure 6.38: Map and trajectory from radar SLAM.



Figure 6.39: A comparison of the trajectory between SLAM from radar (black) and dead-reckoning only (red).

The first part of Tab 6.3 compares the true and estimated path lengths shown in Fig. 6.40. The second part compares the distances between mapped landmarks with measurements of those distances taken by a measuring tape. The distances measured by the tape are within a decimeter of the true values. Association between map landmarks and scene elements is not possible after the second turn. Comparison of distances between landmarks that were mapped before the second turn with any degree of confidence in the association is not possible.



Figure 6.40: Path section and landmarks used in tables.

	Estimated (m)	Actual (m)	Error (m)
Section 1	17.97	19.20	1.23
Section 2	20.55	22.86	2.31
Section 3	24.52	25.60	1.04
4-6	17.97	10.73	0.19
4-7	20.55	35.77	0.81
7-8	24.56	12.16	0.44

Table 6.3: Actual and Estimated Path Lengths and Distances Between Landmarksfrom Radar SLAM

# 6.4 Radar Video Angle Change Estimate Fusion SLAM6.4.1 Angle Change Estimation

An example of the shifts found by correlation between image sums is shown in Figs. 6.41 and 6.42. The second image used for the correlation is shown in Fig. 6.43. As the figures illustrate, the correlation algorithm aligns the image sums such that they nearly overlap. The algorithm could probably be carried out with a smaller section of the image, but the larger FOV makes the algorithm more robust to errors from moving objects in the scene.

The orientation estimates obtained over the experiment are shown in Fig. 6.44. They were made by accumulating the platform angle change estimates over time. The fused orientation estimate is closely aligned with the orientation estimate from the image alone. This is a result of the much smaller estimation error from the images. The estimated path angles and true angles are shown in Tab 6.4.

Adding the angle change estimation from the video greatly improves the platform orientation estimate over radar alone. The radar observability issues are overcome by this addition.



Figure 6.41: Image sums before correlation. The previous image sum is blue. The current image sum is red.

# 6.4.2 Filtering

By the end of the experiment, 40 landmarks are mapped. Fig. 6.45 shows the path of platform and all of the mapped landmarks. A comparison of the SLAM trajectory with the trajectory from the dead-reckoning parameter estimates is shown in Fig. 6.46. The **reduction** in the **number of landmarks** over the radar only SLAM is **due to the loop closing**. Landmarks that were observed during the first pass were successfully associated with their corresponding measurements in the second pass.

	Estimated	Actual	Error
$1^{\rm st}$ turn	$72.59^{\circ}$	$73.03^{\circ}$	$0.44^{\circ}$
$2^{\rm nd}$ turn	$47.29^{\circ}$	$46.43^{\circ}$	$0.86^{\circ}$
$3^{\rm rd}$ turn	$59.5^{\circ}$	$60.54^{\circ}$	$1.04^{\circ}$

Table 6.4: Actual and Estimated Angles for Each Path Turn From Video

This attests to the **accuracy** of the platform position and orientation estimation over the experiment. The **same association parameters** were used throughout the **entire process**.

Table 6.5 shows the ranges between mapped landmarks compared with measurements of those distances taken with a tape measure. Some of the errors are larger than occurred for the radar only case. Again, the path sections and landmark pairs referenced in the table are shown in Fig. 6.40. This is likely due to the loop closing causing adjustments to the entire set of mapped landmarks. The overall error in estimated path lengths is smaller than in the radar only case. The error in the estimate of the length of the third leg of the path is larger for the fused estimate, however.



Figure 6.42: Image sums after correlation. The previous image sum is blue. The current image sum is red.



Figure 6.43: Image on which angle change is being estimated.

	Estimated (m)	Actual (m)	Error (m)
Section 1	18.47	19.20	0.73
Section 2	21.30	22.86	1.56
Section 3	24.06	25.60	1.54
4-5	4.41	3.73	0.68
4-6	10.75	10.92	0.17
4-7	36.07	36.58	0.51
4-9	45.71	46.63	0.92
5-6	8.12	9.19	1.07
7-8	12.12	12.60	0.48
7-9	23.83	24.25	0.42

Table 6.5: Actual and Estimated Path Lengths and Distances Between Landmarksfrom Radar SLAM and Video Angle Measurement Fusion



Figure 6.44: Platform orientation estimation using angle estimates from video. The orientation estimates from the radar, image, and filter are colored red, blue, and black, respectively.

# 6.5 Radar Video Landmark Fusion SLAM

## 6.5.1 Landmark Identification and Tracking

As landmarks were obtained from the radar, identification in the images was attempted. Of the 40 landmarks, 9 were identified in the images. Two of the landmarks were detected on the first pass and tracked again after loop closing. An image of the identified landmarks is shown in Fig. 6.47. All of the landmarks identified were trihedrals, except for a misidentified landmark. The object that was misidentified was identified as a trihedral in the image, however, it was not that particular trihedral. The other 10 trihedrals were not detected because they either did not fall within the range boundaries that determined which landmarks should be searched for, or were only visible for a limited number of frames (during turns). A majority of the other landmarks were light poles. The light poles were painted brown, making them blend



Figure 6.45: Map and trajectory from radar and video angle fusion SLAM. The estimated landmark locations are cyan.

in with the environment. Another reason light poles were not identified was the quality of the radar measurements obtained from them. Fig. 6.48 shows an example of radar measurements obtained from a light pole and a tree overlaid onto the image corresponding to those measurements. Compared to the trihedral overlay shown in Fig. 5.13, the measurements are fewer and more spread out.

# 6.5.2 Filtering

By the end of the experiment, 40 landmarks are mapped. This is the same number of landmarks as the radar and video angle change estimate fusion method. The estimate of the platform orientation over time is shown in Fig. 6.49. The difference in orientation estimates shown in Fig 6.49 and Fig. 6.44 is shown in Fig. 6.50. The estimates differ by a maximum of 1.16°. Once the landmarks were identified in the video, the subsequent angle of arrival errors tended to be relatively small, reducing the effects of the measurements on the platform orientation estimate.

A plot of the trajectories from the radar and video fusion algorithms is shown in Fig. 6.51. Along the first leg of the triangle (on the bottom right), the trajectories are aligned. There is a noticeable difference in the lengths of the paths, as the landmark tracking algorithm estimates the trajectory to be longer. There appears to be more trajectory correction during turns for the landmark tracking algorithm, although both algorithms exhibit corrections. After loop closing, the trajectory for the landmark tracking algorithm appears to align better with the original path estimate. Both algorithms appear to be on track to diverge from the original path, however.

Part of the reason for the divergence is that the map changes over time, especially before the loop closing. Two factors can contribute to this. One is that until correlations are built between landmarks, they are able to move with respect to each



Figure 6.46: The trajectories from the radar dead-reckoning and the radar with video angle fusion SLAM. The dead-reckoning and radar video fusion SLAM trajectories are red and black, respectively.



Figure 6.47: Map of the landmarks tracked using the radar and video. The landmarks identified in the video are cyan.



Figure 6.48: Radar measurements corresponding to the light pole and tree overlaid on their corresponding image.


Figure 6.49: Orientation estimates from radar dead-reckoning (red), video (blue), and video landmark tracking fused result (black).

other. Every time a landmark is observed, some correlation with other landmarks is built. Stronger correlations, however, are built between two landmarks when they are observed at the same time. As new landmarks are added to the filter and old ones cease to be observed, the older landmarks tend to drift a little. This can be a drawback of having so few landmarks. Only a few are observed at any given time, preventing a strong scene "shape" from being built initially. Adding in the video measurements helps strengthen the correlations between landmarks, making a stronger scene "shape" and preventing some of the landmark drift.

The other factor is that the location of the original path was changed due to a shift between the map frame of reference and the world frame of reference. This was due to "spurious" information entering the filter. In the SLAM problem, the observability matrix should always be null. For this particular model, the rank of the null space should have been 3. The null space for this model corresponds to a translation and



Figure 6.50: The difference in orientation estimates between the radar and video fusion algorithms.

rotation of the map with respect to world coordinates. Because the trajectories were so close, and considering the possible landmark drift, it can be assumed that a small amount of "spurious" information was introduced.

Table 6.6 shows the ranges between mapped landmarks compared with measurements of those distances taken with a laser range finder. Overall, the errors are smaller than in the previous cases. The overall error in estimated path lengths is smaller than in the radar only case. The error in the estimate of the length of the third leg of the path is larger for the fused estimate, however.

A comparison of all of the path length estimates is shown in Table 6.7. The interlandmark distance estimates are compared in Table 6.8. Overall, the tables show that the first fusion method performs better than the radar SLAM alone, while the second fusion method performs better than the first fusion method. Note that only the first three sets of inter-landmark distances are compared for the radar SLAM method. The



Figure 6.51: The trajectories from the radar video fusion without landmark tracking (red) and the radar video fusion with landmark tracking (black).

	Estimated (m)	Actual (m)	Error (m)
Section 1	18.73	19.20	0.97
Section 2	21.18	22.86	1.68
Section 3	24.45	25.60	1.15
4-5	4.12	3.73	0.39
4-6	10.66	10.92	0.26
4-7	36.20	36.58	0.38
4-9	45.94	46.63	0.69
5-6	8.33	9.19	0.86
7-8	12.14	12.60	0.46
7-9	24.19	24.25	0.36

Table 6.6: Actual and Estimated Path Lengths and Distances Between Landmarksfrom Radar SLAM and Video Fusion

radar SLAM underestimated the turns, making correct associations between mapped landmarks and their true reflectors in the scene impossible. In order to make a fair comparison, reflectors that were observable after the second turn were not used in the first total error computation. The second total error sum is for comparing the two fusion methods using all of the measured landmarks.

Path section	True length (m)	Radar (m)	Error (m)	Radar + Video angle (m)	Error (m) F a	Radar + Video angle und landmarks (m)	Error (m)
1	19.20	17.97	1.23	18.47	0.73	18.73	0.97
2	22.86	20.55	2.31	21.30	1.56	21.18	1.68
3	25.60	24.56	1.04	24.06	1.54	24.45	1.15
Total path error			4.58		3.83		3.3

Landmark pairs	Measured distance (m)	Radar (m)	Error (m)	Radar + Video angle (m)	Error (m) R aı	adar + Video angle id landmarks (m)	Error (m)
4-6	10.92	10.73	0.19	10.75	0.17	10.66	0.26
4-7	36.58	35.77	0.81	36.07	0.51	36.20	0.38
7-8	12.60	12.16	0.44	12.12	0.48	12.14	0.46
4-9	46.63	Ι	Ι	45.71	0.92	45.94	0.69
2-9	23.83	Ι	Ι	24.25	0.42	24.19	0.36
4-5	3.73	I	I	4.41	0.68	4.12	0.39
5-6	9.19	I	I	8.12	1.07	8.33	0.86
Total error, rows 1-3	I	I	1.44	I	1.16	I	1.1
Total error, rows 1-7	I	I	I	I	4.25	I	3.4

TADIE 0.0: AL	OF THE PARTON	LER UISUALICE	esumates a	<u>ulu ullelt errors. 1</u>	He silialles	ELTOT III EACH TOW IS LEC	-
mark pairs	Measured distance (m)	Radar $(m)$	Error (m)	Radar + Video angle (m)	Error (m)	Radar + Video angle and landmarks (m)	Er
	10.92	10.73	0.19	10.75	0.17	10.66	
	36.58	35.77	0.81	36.07	0.51	36.20	0
	12.60	12.16	0.44	12.12	0.48	12.14	0
	46.63	I	I	45.71	0.92	45.94	0
	23.83	I	I	24.25	0.42	24.19	0
	3.73	I	I	4.41	0.68	4.12	0
	9.19	I	I	8.12	1.07	8.33	0
l error, rows 1-3			1.44	1	1.16	I	

## 7 Conclusion

### 7.1 Summary

There has been a push recently to develop technology to enable the use of UAVs in GPS-denied environments. As UAVs become smaller, there is a need to reduce the number and sizes of sensor systems on board. A video camera on a UAV can serve multiple purposes. It can return imagery for processing by human users. The highly accurate bearing information provided by video makes it a useful tool to be incorporated into a navigation and tracking system. Radars can provide information about the types of objects in a scene and can operate in adverse weather conditions. The range and velocity measurements provided by the radar make it a good tool for navigation.

In this work, FMCW radar and color video were fused to perform SLAM in an outdoor environment. A radar SLAM solution provided the basis for the fusion. Correlations between radar returns were used to estimate dead-reckoning parameters to obtain an estimate of the platform location. Radar reflectors were detected and mapped, further improving the platform location estimates. As images were received from the video camera, changes in platform orientation were estimated, further improving the platform location estimates. The expected locations of radar measurements, whose uncertainty was modeled as Gaussian, were projected onto the images and used to estimate the location of the radar reflector in the image. The colors of the most likely reflector were saved and used to detect the reflector in subsequent images. The azimuth angles obtained from the image detections were used to improve the estimates of the landmarks in the SLAM map.

#### 7.2 Conclusions

In performing SLAM with a phase-comparison monopulse FMCW radar, there are a few challenges that must be overcome. One of the challenges considered in this work was a way to estimate the motion of the platform independent of the map. The other challenge was in generating as many detections as possible on stable landmarks while limiting the number of detections on landmarks that would be more sparsely observable.

For the first challenge, platform velocity and orientation change estimates were made based on the returned signal waveforms in order to perform dead-reckoning estimation. Correlating received pulse waveforms to estimate the velocity of the platform can have a tendency to underestimate the true velocity. The cosine effect induced by the angle between the velocity direction and the direction to reflectors reduces the estimated velocity. The phase-comparison monopulse radar provided an AOA estimate for each magnitude response bin. Using this information, the velocity estimates were improved. Differencing the AOA waveforms from one pulse to the next provided an estimate of the change in orientation. The estimates provided by this method generally tracked with the radar change in orientation, but were not always sufficient.

A common method of detecting reflectors in radar signals is CFAR. This method compares a cell in the magnitude response with neighboring cells to determine if it is likely to originate from a reflector. The AOA estimates provided by phase-comparison monopulse radar also contain information about the existence of a reflector in a cell. Assuming that strong reflectors are represented in a group of neighboring cells, obtaining a similar AOA across a group of cells can signify the existence of a reflector. Further, this smoothness of AOA response should persist over time. Combining estimates from these methods improved the number and quality of results. The CFAR threshold was raised, so that detections could be made that otherwise would not. The AOA constraint ensured that the measurements were stable on the map, keeping the map consistent. Keeping the smoothness constraint also enabled associations between measurements and landmarks to be made in the detection process, with little extra computation, reducing the computations later for associating unmatched measurements with landmarks.

When fusing radar and video for SLAM, it is a challenge to combine complementary information from both sensors in an efficient way. Using radar SLAM as the foundation, the need for video processing can be reduced. It was shown that adding orientation change estimates from the video improved both the trajectory and the map over radar SLAM alone, enabling loop closure.

Segmenting radar reflectors in a scene is a difficult task. In non-roadway scenarios, there can be no assumption about the shapes of the reflectors. The method provided in this work was able to detect reflectors placed in a scene based on estimating the likely colors of the reflectors. By extracting the colors of the landmarks, they could be tracked in subsequent frames without the need to account for changes in landmark size or shape in the images as the viewing location and angle changed. The increased bearing accuracy improved both the map and the platform location estimates.

### 7.3 Contributions

The contributions provided in this work are:

- A solution was provided for fusing FMCW radar and video information to perform SLAM.
- A method of extracting dead-reckoning variable estimates from FMCW radar was provided.
- A method of improving detections in FMCW radar signals for the purpose of SLAM was provided.
- A method of identifying reflectors in an image from FMCW radar measurements was provided.

#### 7.4 Future Work

The main objective behind this work was to fuse sensors that could be put on an aerial platform for SLAM. In moving to an aerial platform, some problems need to be worked on:

• The detection and dead-reckoning algorithms should be tested on a radar that provides elevation AoA measurements in addition to azimuth AoA measurements. The smoothness constraint applied to the elevation AoA could further improve detections. It might be possible that the smoothness constraint on the AoAs is sufficient and CFAR is unnecessary. In addition, the current method for estimating azimuth AoA does not account for how different elevations can affect the receiver phase differences. A system with more receivers can better approximate the AoA in both directions.

The dead-reckoning orientation change estimation is less likely to work in a 6 degree of freedom environment when only azimuth AoA is available. It is likely that the solution using both sets of measurements will be much more complex, as the relative changes in AoA between reflectors will become non-linear.

• Adding an IMU to the sensor system would remove the need for radar odometry. It would add more weight to the platform and require more power, but it could reduce the processing necessary.

In order to make the system more robust in the real world, some improvements need to be made:

• A more robust segmentation algorithm is necessary. In order to work in the real world, the distinct color requirement will have to be relaxed. It is also important to keep the size and shape constraints relaxed. A foreground segmentation algorithm such as Kernel Density Estimation (KDE) might be useful. That

algorithm initializes with a likelihood image similar to the one used in this work. A caveat in using the algorithm in its current form is that it tends to be inclusive and can end up segmenting more than the reflector. It is sensitive to the initial likelihood and can require many iterations to reach a stable segmentation.

- Incorporating moving object tracking into the solution will be an important step. It can be expected that ideal reflectors in urban areas will be vehicles. They have large cross-sections and tend to be painted colors that make them easily distinguishable from their surroundings. It should be expected that they will also move. The radar detections can provide information about the velocity of moving objects in the scene. This sensor platform should be able to successfully track moving objects, once they are identified as such.
- Identifying situations where two reflectors are at a similar range from the platform will also need to be accomplished. There is no accounting for such a case currently. A detection in the radar signal cannot distinguish between one or multiple reflectors causing a response at a particular frequency bin.

# References

- M. Monod, P. Faure, and R. Rouveure, "Intertwined linear frequency modulated radar and simulator for outdoor robotics applications," in *IEEE International Conference on Radar Systems*, October 2009, pp. 06–12.
- [2] J. Langelaan and S. Rock, "Passive GPS-free navigation for small uavs," in *IEEE Aerospace Conference*. Ieee, March 2005, pp. 1–9.
- [3] F. Gérossier, P. Checchin, C. Blanc, R. Chapuis, and L. Trassoudaine, "Trajectory-oriented EKF-SLAM using the fourier-mellin transform applied to microwave radar images," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009.* IEEE, October 2009, pp. 4925–4930.
- [4] D. Vivet, P. Checchin, and R. Chapuis, "On the fly localization and mapping using a 360 field-of-view microwave radar sensor." St Louis, MO: Citeseer, October 2009, pp. 23–28.
- [5] F. Cheein and R. Carelli, "Analysis of different feature selection criteria based on a covariance convergence perspective for a SLAM algorithm," *Sensors (Basel, Switzerland)*, vol. 11, no. 1, p. 62, 2011.
- [6] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis, "Improved fast gauss transform and efficient kernel density estimation," in *Proceedings of IEEE International Conference on Computer Vision*. College Park, MD: Ieee, October 2003, pp. 664–671.
- [7] M. I. Skolnik, Radar handbook. New York: McGraw-Hill Professional, 1990.
- [8] R. Rouveure, P. Faure, P. Checchin, and M. Monod, "Mobile robot localization and mapping in extensive outdoor environment based on radar sensor-first results," *Proc. of PSIP 2007*, Feb 2007.

- [9] F. Ramos, J. Nieto, and H. Durrant-Whyte, "Recognising and modelling landmarks to close loops in outdoor SLAM," in *IEEE International Conference on Robotics and Automation*. IEEE, April 2007, pp. 2036–2041.
- [10] J. Kim and S. Sukkarieh, "Real-time implementation of airborne inertial-SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 62–71, January 2007.
- P. Checchin, F. Gérossier, C. Blanc, R. Chapuis, and L. Trassoudaine, "Radar scan matching SLAM using the Fourier-Mellin transform," in *Field and Service Robotics*, ser. Springer Tracts in Advanced Robotics, vol. 62, Springer. Berlin / Heidelberg: Springer, 2010, pp. 151–161.
- [12] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Transactions on Aerospace and Electronic Systems*, no. 4, pp. 608–621, July 1983.
- [13] M. Chandran and P. Newman, "Motion estimation from map quality with millimeter wave radar," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, October 2006, pp. 808–813.
- [14] E. Jose and M. Adams, "Millimetre wave radar spectra simulation and interpretation for outdoor SLAM," in *IEEE International Conference on Robotics and Automation*, vol. 2. IEEE, April 2004, pp. 1321–1326.
- [15] R. Rouveure, M. Monod, and P. Faure, "Mapping of the environment with a high resolution ground-based radar imager," in 14th IEEE Mediterranean Electrotechnical Conference. IEEE, May 2008, pp. 822–828.
- [16] K. Granstrom, J. Callmer, F. Ramos, and J. Nieto, "Learning to detect loop closure from range data," in *IEEE International Conference on Robotics and Automation*. IEEE, May 2009, pp. 15–22.

- [17] R. Rouveure, M. Monod, and P. Faure, "High resolution mapping of the environment with a ground-based radar imager," in *International Radar Conference-Surveillance for a Safer World*. IEEE, October 2009, pp. 1–6.
- [18] M. Schikora and B. Romba, "A framework for multiple radar and multiple 2D/3D camera fusion," in 4th German Workshop Sensor Data Fusion: trends, Solutions, Applications (SDF), 2009.
- [19] B. Ma, S. Lakshmanan, and A. Hero III, "Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion," *IEEE Transactions* on Intelligent Transportation Systems, vol. 1, no. 3, pp. 135–147, September 2000.
- [20] R. Schweiger, H. Neumann, and W. Ritter, "Multiple-cue data fusion with particle filters for vehicle detection in night view automotive applications," in *Proceedings. IEEE Intelligent Vehicles Symposium.* IEEE, June 2005, pp. 753–758.
- [21] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fastslam: A factored solution to the simultaneous localization and mapping problem," in *Proceedings of* the National conference on Artificial Intelligence. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2002, pp. 593–598.
- [22] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, June 2001.
- [23] S. Clark and H. Durrant-Whyte, "Autonomous land vehicle navigation using millimeter wave radar," in 1998 IEEE International Conference on Robotics and Automation, vol. 4. IEEE, May 1998, pp. 3697–3702.

- [24] S. Clark and H. Durran-Whyte, "The design of a high performance MMW radar system for autonomous land vehicle navigation," in *Proc. Int. Conf. Field and Service Robotics*. Citeseer, December 1997, pp. 292–299.
- [25] J. Mullane, S. Keller, A. Rao, M. Adams, A. Yeo, F. Hover, and N. Patrikalakis, "X-band radar based SLAM in singapore's off-shore environment," in 11th International Conference on Control Automation Robotics & Vision. IEEE, December 2010, pp. 398–403.
- [26] J. Callmer, D. Törnqvist, F. Gustafsson, H. Svensson, and P. Carlbom, "Radar SLAM using visual features," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–11, 2011.
- [27] C. Lundquist, T. Schön, and U. Orguner, "Estimation of the free space in front of a moving vehicle," in *Proceedings of the SAE World Congress*, April 2009, pp. 01–1288.
- [28] K. Yokoo, S. Beauregard, and M. Schneider, "Indoor relative localization with mobile short-range radar," in *IEEE 69th Vehicular Technology Conference*. Barcelona: IEEE, April 2009, pp. 1–5.
- [29] Z. Sjanic, M. Skoglund, T. Schön, and F. Gustafsson, "Solving the SLAM problem for unmanned aerial vehicles using smoothed estimates," in *Proceedings of the Reglermöte (Swedish Control Conference)*, Lund, Sweden, June 2010.
- [30] G. Huang, A. Mourikis, and S. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, April 2010.

- [31] —, "On the complexity and consistency of UKF-based SLAM," in *IEEE In*ternational Conference on Robotics and Automation. IEEE, May 2009, pp. 4401–4408.
- [32] S. Holmes, G. Klein, and D. Murray, "A square root unscented kalman filter for visual monoslam," in *IEEE International Conference on Robotics and Automation.* IEEE, May 2008, pp. 3710–3716.
- [33] N. Sunderhauf, S. Lange, and P. Protzel, "Using the unscented kalman filter in mono-slam with inverse depth parametrization for autonomous airship control," in *IEEE International Workshop on Safety, Security and Rescue Robotics.* IEEE, September 2007, pp. 1–6.
- [34] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer, 2006, vol. 3951, pp. 404–417.
- [35] D. Lowe, "Distinctive image features from scale-invariant keypoints," International journal of computer vision, vol. 60, no. 2, pp. 91–110, 2004.
- [36] J. Shi and C. Tomasi, "Good features to track," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, June 1994, pp. 593–600.
- [37] A. Davison, I. Reid, N. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *Pattern Analysis and Machine Intelligence*, *IEEE Transactions* on, vol. 29, no. 6, pp. 1052–1067, June 2007.

- [38] J. Civera, A. Davison, and J. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, October 2008.
- [39] (2010, May). [Online]. Available: http://www.af.mil/shared/media/document/ AFD-100727-053.pdf
- [40] M. Montemerlo, J. Becker, S. Bhat, H. Dahlkamp, D. Dolgov, S. Ettinger, D. Haehnel, T. Hilden, G. Hoffmann, B. Huhnke *et al.*, "Junior: The stanford entry in the urban challenge," *Journal of field Robotics*, vol. 25, no. 9, pp. 569–597, August 2008.
- [41] S. Kammel, J. Ziegler, B. Pitzer, M. Werling, T. Gindele, D. Jagzent, J. Schröder,
  M. Thuy, M. Goebl, F. Hundelshausen *et al.*, "Team annieway's autonomous system for the 2007 DARPA urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 615–639, Aug 2008.
- [42] A. Bacha, C. Bauman, R. Faruque, M. Fleming, C. Terwelp, C. Reinholtz, D. Hong, A. Wicks, T. Alberi, D. Anderson *et al.*, "Odin: Team victortango's entry in the DARPA urban challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 467–492, June 2008.
- [43] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann *et al.*, "Stanley: The robot that won the DARPA grand challenge," in *The 2005 DARPA Grand Challenge*, ser. Spring Tracts in Advanced Robotics. Berlin / Heidelberg: Springer, 2007, pp. 1–43.
- [44] J. Bohren, T. Foote, J. Keller, A. Kushleyev, D. Lee, A. Stewart, P. Vernaza, J. Derenick, J. Spletzer, and B. Satterfield, "Little ben: The ben franklin racing team's entry in the 2007 DARPA urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 598–614, August 2008.

- [45] Z. Jia, A. Balasuriya, and S. Challa, "Autonomous vehicles navigation with visual target tracking: Technical approaches," *Algorithms*, vol. 1, no. 2, pp. 153–182, December 2008.
- [46] A. Roy, N. Gale, and L. Hong, "Automated traffic surveillance using fusion of doppler radar and video information," *Mathematical and Computer Modelling*, vol. 54, no. 1, pp. 531–543, July 2011.
- [47] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and guard rail detection using radar and vision data fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 95–105, March 2007.
- [48] L. Bombini, P. Cerri, P. Medici, and G. Alessandretti, "Radar-vision fusion for vehicle detection," in *Proceedings of International Workshop on Intelligent Transportation*, Hamburg, Germany, March 2006, pp. 65–70.
- [49] M. Mahlisch, R. Schweiger, W. Ritter, and K. Dietmayer, "Sensorfusion using spatio-temporal aligned video and lidar for improved vehicle detection," in *IEEE Intelligent Vehicles Symposium*. IEEE, September 2006, pp. 424–429.
- [50] A. Gern, U. Franke, and P. Levi, "Robust vehicle tracking fusing radar and vision," in International Conference on Multisensor Fusion and Integration for Intelligent Systems. IEEE, 2001, pp. 323–328.
- [51] A. Haselhoff, A. Kummert, and G. Schneider, "Radar-vision fusion with an application to car-following using an improved adaboost detection algorithm," in *IEEE Intelligent Transportation Systems Conference*. IEEE, Oct 2007, pp. 854– 858.

- [52] B. Heisele, H. Neef, W. Ritter, R. Schneider, and G. Wanielik, "Object detection in traffic scenes by a colour video and radar data fusion approach," in *First Australian Data Fusion Symposium*. IEEE, Nov 1996, pp. 48–52.
- [53] P. Viola and M. Jones, "Robust real-time object detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, February 2001.
- [54] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on a new method for the nonlinear transformation of means and covariances in filters and estimators [with authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, Aug 2002.
- [55] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," Proceedings of the IEEE, vol. 92, no. 3, pp. 401–422, March 2004.
- [56] M. Grewal, A. Andrews, and E. Corporation, Kalman filtering: theory and practice using MATLAB. New York, USA: John Wiley & Sons, Inc, May 2001.
- [57] D. Hall and S. McMullen, Mathematical techniques in multisensor data fusion. Norwood, MA: Artech House Publishers, 2004.
- [58] A. Davison and D. Murray, "Simultaneous localization and map-building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 865–880, July 2002.
- [59] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108–117, September 2006.
- [60] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping: Part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, June 2006.