MMWAVE RADAR: ENHANCING RESOLUTION, TARGET RECOGNITION, AND FUSION WITH OTHER SENSORS

by

Renyuan Zhang

Copyright © Renyuan Zhang 2019

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Renyuan Zhang, titled mmWave radar: Enhancing resolution, target recognition, and fusion with other sensors and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Siyang Cao
Date: October 22, 2019

Gregory Ditzler
Date: October 22, 2019

Ming Li
Date: October 22, 2019

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Siyang Cao
Dissertation Committee Chair
Electrical and Computer Engineering

Date: October 22, 2019
ACKNOWLEDGMENTS

Foremost, I would like to thank my advisor, Dr. Siyang Cao, for his advising during my academic research on radar signal processing. He provides me the guidance and deep knowledge of modern millimeter wave radars and digital signal processing techniques. His willingness on solving my questions and giving me advises on approaches to problem concerned is outstanding. His knowledgeable, cooperative, compassionate, flexible, supportive, inspirational, selfless qualities are worthy of praise.

I would also like to express my gratitudes to my committee members, Dr. Ming Li and Dr. Gregory Ditzler for their valuable comments and suggestions. Especially thank to Dr. Kathleen Melde and Dr. Amit Ashok for their kindly help on antenna theory, microwave and polishing up this dissertation.

Great appreciation to my colleagues at lab, Feng Jin, Arindam Sengupta and Nathan Madsen for their kindly help in my research, experiment and presentations. Thank to Jean de Dieu Mutangana, Yiming Zhang, Kemeng Chen, Xin Xiao, Xiao Han and Rahul Salvatore Bhadani for their help during my Ph.D. studies.

I would extend my deepest gratitude to my parents, Jiguang Zhang and Jing Huang for their support. Without their endless support, I would not achieve today’s achievements and finish this Ph.D. dissertation.

Thank you all for your help.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>SYMBOLS</td>
<td>11</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>12</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>14</td>
</tr>
<tr>
<td>PREFACE</td>
<td>16</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>17</td>
</tr>
<tr>
<td>1.1 Dissertation Organization</td>
<td>21</td>
</tr>
<tr>
<td>2 EXTENDING RELIABILITY OF MMWAVE RADAR TRACKING AND DETECTION VIA FUSION WITH CAMERA</td>
<td>24</td>
</tr>
<tr>
<td>2.1 Background</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Methodology</td>
<td>28</td>
</tr>
<tr>
<td>2.2.1 Coordinates and Error Bounds</td>
<td>29</td>
</tr>
<tr>
<td>2.2.2 Camera and mmWave Radar Preprocessing</td>
<td>32</td>
</tr>
<tr>
<td>2.2.3 HEM</td>
<td>37</td>
</tr>
<tr>
<td>2.2.4 Fusion-EKF</td>
<td>39</td>
</tr>
<tr>
<td>2.2.5 Data Association and Sensor Synchronization</td>
<td>46</td>
</tr>
<tr>
<td>2.2.6 EB Evaluation</td>
<td>52</td>
</tr>
<tr>
<td>2.3 Experimental Results</td>
<td>54</td>
</tr>
<tr>
<td>3 REAL-TIME HUMAN MOTION BEHAVIOR DETECTION VIA CNN USING MMWAVE RADAR</td>
<td>66</td>
</tr>
<tr>
<td>3.1 Background</td>
<td>66</td>
</tr>
<tr>
<td>3.2 Methodology of Real-Time MDS Observation</td>
<td>68</td>
</tr>
<tr>
<td>3.2.1 Real-Time FMCW MDS Processing</td>
<td>68</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.3.2 Radon Transform and 3D Imaging Reconstruction</td>
<td>109</td>
</tr>
<tr>
<td>5.4 Compressed Sensing for 3D Imaging Radar System</td>
<td>112</td>
</tr>
<tr>
<td>5.4.1 Compressed Sensing Review</td>
<td>113</td>
</tr>
<tr>
<td>5.4.2 Compressed Sensing on Doppler-Angle Data</td>
<td>114</td>
</tr>
<tr>
<td>5.4.3 Sensing Basis $\phi$ Selection</td>
<td>116</td>
</tr>
<tr>
<td>5.5 Simulation and Experiment</td>
<td>118</td>
</tr>
<tr>
<td>5.5.1 Simulation Setup and Results</td>
<td>118</td>
</tr>
<tr>
<td>5.5.2 Experiment Setup and Results</td>
<td>121</td>
</tr>
<tr>
<td>5.6 Discussion</td>
<td>124</td>
</tr>
<tr>
<td>5.6.1 Millimeter Wave 3D Imaging Radar</td>
<td>124</td>
</tr>
<tr>
<td>5.6.2 Comparison with the Radon Method and the Compressed Sensing Method</td>
<td>125</td>
</tr>
<tr>
<td>6 STRIP MAP SYNTHETIC APERTURE RADAR FOR AUTOMOTIVE RADAR</td>
<td>127</td>
</tr>
<tr>
<td>6.1 Background</td>
<td>127</td>
</tr>
<tr>
<td>6.2 Terrain Mapping Methodology</td>
<td>128</td>
</tr>
<tr>
<td>6.3 Experimental Results</td>
<td>130</td>
</tr>
<tr>
<td>6.3.1 Setup</td>
<td>130</td>
</tr>
<tr>
<td>6.3.2 Results</td>
<td>132</td>
</tr>
<tr>
<td>6.3.3 Discussions</td>
<td>133</td>
</tr>
<tr>
<td>7 CONCLUSION</td>
<td>137</td>
</tr>
<tr>
<td>APPENDIX A: Dissertation Equations</td>
<td>140</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>143</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Average localization error of radar-camera fusion-EKF</td>
</tr>
<tr>
<td>2.2</td>
<td>Comparison of error/EBs using radar-camera fusion EKF with other radar tracking algorithms</td>
</tr>
<tr>
<td>3.1</td>
<td>CNN summarize (Total parameters: 1,141,763; Trainable parameters: 1,141,763; Non-trainable parameters: 0)</td>
</tr>
<tr>
<td>3.2</td>
<td>CNN prediction accuracies (Collected multiple test samples and averaging results among samples)</td>
</tr>
<tr>
<td>4.1</td>
<td>FMCW MIMO radar parameters</td>
</tr>
<tr>
<td>4.2</td>
<td>FMCW MIMO radar output</td>
</tr>
<tr>
<td>5.1</td>
<td>MMWCSAR simulation setup. PRI, pulse repetition interval</td>
</tr>
<tr>
<td>5.2</td>
<td>Four targets’ scheme setup. RCS, radar cross-section.</td>
</tr>
<tr>
<td>5.3</td>
<td>Three ball targets’ scheme setup.</td>
</tr>
<tr>
<td>6.1</td>
<td>300 MHz bandwidth experiment setup</td>
</tr>
<tr>
<td>6.2</td>
<td>1 GHz bandwidth experiment setup</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>FMCW MIMO radar signal processing.</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>Dissertation pipeline.</td>
<td>21</td>
</tr>
<tr>
<td>2.1</td>
<td>Coordinates definition.</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Error bounds of fused sensor system.</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Camera calibration.</td>
<td>33</td>
</tr>
<tr>
<td>2.4</td>
<td>radar-camera fusion-EKF workflow.</td>
<td>40</td>
</tr>
<tr>
<td>2.5</td>
<td>Radar-camera association timeline alignment.</td>
<td>48</td>
</tr>
<tr>
<td>2.6</td>
<td>Radar-camera association region search. Blue dash line: EKF projected trajectory, yellow region: search region for next state for radar clusters and camera bounding box BEV projection.</td>
<td>50</td>
</tr>
<tr>
<td>2.7</td>
<td>EKF output (1).</td>
<td>55</td>
</tr>
<tr>
<td>2.8</td>
<td>EKF output (2).</td>
<td>56</td>
</tr>
<tr>
<td>2.9</td>
<td>EKF output (3).</td>
<td>57</td>
</tr>
<tr>
<td>2.10</td>
<td>EKF RMSE of radar’s range.</td>
<td>59</td>
</tr>
<tr>
<td>2.11</td>
<td>EKF RMSE of radar’s angle.</td>
<td>60</td>
</tr>
<tr>
<td>2.12</td>
<td>EKF RMSE of radar’s Doppler.</td>
<td>61</td>
</tr>
<tr>
<td>2.13</td>
<td>Fusion-EKF benchmark on EB respect to ground truth.</td>
<td>63</td>
</tr>
<tr>
<td>2.14</td>
<td>Cumulative distribution of localization errors.</td>
<td>65</td>
</tr>
<tr>
<td>3.1</td>
<td>The micro-Doppler signature of a human walking.</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>The CNN framework.</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>The whole ROS implementation on mmWave radar sensors.</td>
<td>73</td>
</tr>
<tr>
<td>4.1</td>
<td>FMCW MIMO radar signal processing.</td>
<td>79</td>
</tr>
<tr>
<td>4.2</td>
<td>Sample radar point cloud detections and density functions.</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison with GMM and DBSCAN.</td>
<td>89</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>4.4 Detecting short range targets.</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>4.5 Detecting long range wall targets.</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>4.6 Detecting large car targets.</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>4.7 Detecting multiple complicated targets.</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>5.1 Uses rotation radar to resolve 3D imaging. (a) Schematic view of resolving range using range bins; (b) the swinging within the rotation plane of the radar generates velocities in different directions; (c,d) each range bin is then projected into different velocity directions while data are collected.</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>5.2 Geometry of monostatic radar remote sensing targets (the positive y axis is the boresight direction).</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>5.3 Geometry of monostatic radar velocity projections (the positive y axis is the boresight direction).</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>5.4 From projections to radar signal datacube.</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>5.5 2D point spread function (PSF) for the point scatterer located at the center (0,0) at a range of 5 m (amplitude in dB).</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>5.6 Simulation results (front view from radar panel at 5 m, the x axis and y axis are azimuth and elevation distance from the center, respectively). (a) Inverse Radon transform (IRT) method with the perfect setup; (b) compressed sensing (CS) applying Equation (5.38) to the angle profile with $R_{CS} = 1/2$; (c) CS applying Equation (5.39) to the angle profile with $R_{CS} = 1/2$; (d) CS applying Equation (5.40) to the angle profile with $R_{CS} = 1/2$; (e) CS applying Equation (5.38) to the slow-time profile with $R_{CS} = 1/2$; (f) CS applying Equation (5.39) to the slow-time profile with $R_{CS} = 1/2$; (g) CS applying Equation (5.40) to the slow-time profile with $R_{CS} = 1/2$.</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>5.7 Experiment setup. (a) Radar position with antenna facing three ball targets; (b) three ball targets’ lineup.</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>5.8 Experiment results. (a,b) IRT method; (c,d) CS applying Equation (5.38) to the slow-time profile.</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>6.1 77 GHz automotive radar terrain mapping methodology.</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>6.2 Side looking SAR. Squinted angle and RD response accumulating algorithm.</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>6.3 77 GHz automotive radar and the autonomous driving test car.</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>6.4 Experiment test ground.</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>300 MHz bandwidth SAR terrain mapping. (a)(c)(e) Using different velocity mapping from (b)(d)(f), respectively.</td>
<td>134</td>
</tr>
<tr>
<td>6.6</td>
<td>1 GHz bandwidth SAR terrain mapping. (a)(c)(e) Using different velocity mapping from (b)(d)(f), respectively.</td>
<td>135</td>
</tr>
<tr>
<td>A.1</td>
<td>Geometry of uniform linear array antenna of MIMO radar.</td>
<td>141</td>
</tr>
</tbody>
</table>
SYMBOLS

$PRI$ pulse repetition interval
$v_d$ Doppler velocity
$TX$ transmitters
$RX$ receivers
$f_c$ center frequency
$f_{\text{start}}$ chirp starting frequency
$f_{\text{stop}}$ chirp end frequency
$T_P$ chirp pulse width
$N_R$ fast time samples
$N_D$ number of chirps
$f_s$ sampling frequency
$f_c$ center frequency
ABBREVIATIONS

mmWave    millimeter wave
FMCW      frequency modulated continuous wave
LFM       linear frequency modulation
SAR       synthetic aperture radar
ADAS      advanced driver assistance system
TX        transmitter
RX        receiver
VCO       voltage controlled oscillator
LPF       low pass filter
LNA       low noise amplifier
EM        electromagnetic
CRF       camera radar fusion
KF        Kalman filter
EKF       extended Kalman filter
CNN       convolution neural network
AoA       angle of arrival
ADC       analog-to-digital converter
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB</td>
<td>error bounds</td>
</tr>
<tr>
<td>MIMO</td>
<td>multiple-input and multiple-output</td>
</tr>
<tr>
<td>ROS</td>
<td>robotic operating system</td>
</tr>
<tr>
<td>LS</td>
<td>least squares</td>
</tr>
<tr>
<td>SVM</td>
<td>support vector machine</td>
</tr>
<tr>
<td>IRT</td>
<td>inverse Radon transform</td>
</tr>
</tbody>
</table>
ABSTRACT

Over the last decade, the advanced driver assistance system (ADAS) and autonomous driving research have grown rapidly. The entire automotive industry is looking forward to autonomous vehicles and ADAS technologies. Fully autonomous driving by the automobile model year 2021/2022 with security level 4 or 5 requires the use of multiple heterogeneous sensors’ system. Automotive sensors, such as camera, millimeter (mmWave) radar and lidar, have evolved fast in signal processing for the perception of surroundings. Sensor fusion and deep learning to understand the environment implemented in automobiles are drastically changing the current sensor research. The automotive radar has been served as an essential sensor in the race to develop ADAS and autonomous vehicles. Its affordable price and reliable detection are raising attention from both industry and academia. In 2018, shipments of passenger automotive radars have grown 54% in units compared to 2017. Another trend is that with camera and radar getting fused, it can provide more reliable ADAS capabilities.

In this dissertation, a series of signal processing techniques are studied for improving the resolution and target recognition of mmWave radar. First, a sensor fusion technique for better tracking and detecting targets using mmWave radar and camera is presented. The fusion system takes consideration of error bounds (EBs) of the two different coordinate systems from the heterogeneous sensors, and further designed a new fusion extended Kalman filter (fusion-EKF) to adapt to the two sensors. The details such as synchronization between sensors, multi-target tracking, and association are also considered and illustrated. The experiment shows that the proposed fusion system can realize a range accuracy of 0.29 m with an angular accuracy of 0.013 rad
in real-time. Therefore, the proposed fusion system is effective, reliable and computationally efficient for real-time kinematic fusion applications. A clustering method, REDBSCAN, for radar point cloud data is also presented. Secondly, for enhancing target recognition, a neural network is developed for mmWave radar to classify human behavior in real-time. Thirdly, to improve the angular resolution for mmWave radar, a circular synthetic aperture radar MMWCSAR with high-resolution technique, e.g., compressed sensing is presented.
PREFACE

Modern radars, especially millimeter wave radars, working at 77 GHz, are critical sensors at automobile field. mmWave radars are robust and reliable at detecting targets and resolving their ranges, Dopplers and angles for advanced driver assistance systems (ADAS) and autonomous driving applications. In this dissertation, I am discussing several important signal processing techniques in mmWave radar field, which includes radar tracking, sensor fusion, micro-Doppler signatures, synthetic aperture radar (SAR) and machine learning applied radar researches.
1. INTRODUCTION

RADAR is the acronym for radio detection and ranging. Radar signal processing is an essential component of modern sensors technologies and radio frequency (RF) research. mmWave radar uses 77 GHz carrier frequency, which is ideal for automobiles and surveillance systems. The mmWave radar signal processing resolves problems related to imaging, target range, and Doppler information, tracking objects and providing micro-Doppler signatures of targets under detection. Comparing to other sensors, mmWave has advantages such as that it can detect targets under adverse weather conditions and finding out depth/range information reliably.

Fig. 1.1 illustrates the typical modern mmWave MIMO radar workflow. mmWave radar uses multiple transmitting antennas (TXs) and receiving antennas (RXs). The radar TXs transmit a modulated signal, typically a frequency-modulated continuous wave (FMCW) waveform. Linear frequency-modulated (LFM) waveform is a popular choice in modern radar systems because it can achieve high-range resolution by sweeping through a wide bandwidth. Continuous wave (CW) allows radar to solve Doppler information from the scatter. Compared to traditional CW radar, it solves the range frequencies using LFM waveform. Relative to a pulse signal radar, it can lower the sampling rate of an analog-to-digital converter (ADC). The radar receiver antenna receives the electromagnetic (EM) waves that are reflected from scatters of an object. After it subsequently performs stretch processing [1, 2], the RF signal is converted to an intermediate frequency (IF) signal. It converts the IF signal to a digital signal through ADC. Range-Doppler processing solves the fast-slow time signal to range and Doppler frequencies; additionally, the 2D cell averaging constant false alarm rate (CA-CFAR) is performed to detect peaks in noise. The signal processor analyzes the
received signal to identify the range and Doppler and angle information concerning objects. Digital beamforming then estimates the scatters angles from different peaks using phase angles of MIMO antennas. RXs receive the reflected EM waves from objects. The receiving circuits amplify the signal and mix all of the receiving signals with transmitting signals. The digital data is then sent to the digital signal processor (DSP) module for further processing.

During my Ph.D. studies on radar signal processing, I focused on the high-level processing chain. My Ph.D. research was in the area of Doppler processing, SAR imaging, radar target tracking, sensors integration and machine learning applied radar research in target classification. My general areas of focus have been related to ADAS
and autonomous driving. High-level radar tracking, high-level clustering, and target and behavior recognition and imaging are the core areas addressed by this dissertation.

One of my major contributions is advanced tracking algorithms for mmWave radar. Traditional tracking technique relies on Kalman filter (KF) applied on one type of sensor. However, in autonomous driving, we cannot rely on one type of sensor. For example, camera is unstable under different light conditions or adverse weather conditions. mmWave does not have sufficient angular resolution to image targets. In this case, we take advantages of camera and improve target tracking and association with a new fusion extended Kalman filter (fusion-EKF). The sensor fusion using the cameras vision is intended to improve automotive radars angular resolution, tracking and detecting reliability and enhanced classifications. My study concerning tracking using inhomogeneous sensors fusion with cameras addresses the current deficient research on clustering, multi-target tracking, alignment, association, and data fusion on mmWave radar. With better centroid estimations of the target and noise and clutter removal, better tracking can be realized. These are new innovations on automotive radar signal processing. I have also developed some state-of-the-art algorithms for automotive radar reliability. It enhances sensing ability along with detecting accuracies to prevent a sensors failure or misdetection and minimize the error bounds (EBs) from fusion sensors. Furthermore, with multi-path effects taken into account, advanced gating and association algorithms for mmWave radar have been developed. The sensor fusion project, including EKF tracking on mmWave sensors, is presented in Chapter 2.

In addition, some post-processing clustering algorithms including radar elliptical density-based spatial clustering of applications with noise (REDBSCAN) have also been developed through my study. The new adaptive and robust clustering and labelling algorithm for radar point cloud is based on the FMCW radars parameters, and it is derived from machine learning algorithm DBSCAN. Combining density function
with mmWave radar parameters, the radar detection points in polar coordinates are thereby clustered. The clustering algorithm deals with close- and long-distance targets in radar polar coordinates with different ellipses. Improvements in the elliptical clustering REDBSCAN algorithm from original DBSCAN algorithm yields better results for close- or long-range targets for mmWave radar. Therefore, it is adaptive and robust for mmWave radar output. It can handle mmWave radar noise as well. It is the clustering algorithm used in the fusion-EKF project. The REDBSCAN algorithm is elucidated in Chapter 4.

Comparing to traditional airborne SAR imaging, the mmWave synthetic aperture radar is a novel and interesting field. In the field of mmWave SAR imaging, I completed a study concerning generating 3D imaging using modern 77 GHz radar with a single transceiver. The Radon transform, high-resolution technique of compressed sensing and some advanced radar signal processing techniques are applied. A circular synthetic aperture radar is algorithmically processed and can be applied to many scenarios, such as concealed body imaging, indoor scanning, and automotive imaging. The circular synthetic aperture radar offers many new ideas which have not been explored by previous research: using mmWave radar as a primary source, Doppler processing to obtain localization of mmWave radar, and compressed sensing to improve azimuth and elevation resolution. By applying mmWave circular SAR algorithm, the outdoor parking lot SAR imaging can also be achieved, which can enable future ADAS and autonomous driving.

In the field of Doppler processing, one of the most popular research fields is micro-Doppler signatures (MDS). MDS has been introduced in many studies [3–5]. Victor Chen has introduced the micro-Doppler signature concept of mmWave radars, which is the micro motions along bulk motion of an object. And much research has not been concerned with solving ranges by applying FMCW waveforms. In my research and with my lab mates, the FMCW radars are used in lieu of traditional CW radars.
CW radar lacks the ability to solve range frequencies. Therefore, with the help of FMCW radars, the scenario of micro-Doppler signatures of human behavior monitoring can be accomplished. Additionally, by adding clustering and Kalman filters, the micro-Doppler signatures of multiple human behaviors (particularly in hospital wards), patients care can be realized. This project will be further developed in the future. In this dissertation, the above topics are detailed and analyzed in separate chapters.

1.1 Dissertation Organization

In this dissertation, my radar signal processing pipeline can be seen in Fig. 1.2. The diagram shows the 3 major fields which are radar tracking and sensor fusion, Doppler processing and imaging. The machine learning applied research includes REDBSCAN, clustering, human behavior monitoring using CNN and researches not
included in this dissertation, \textit{e.g.}, non-synchronized integration, interference classification. The remaining chapters are organized as follows:

In Chapter 2, the radar-camera sensor fusion project is explained. A state-of-the-art radar-camera sensor fusion system uses EKF as a primary filter for tracking multiple objects simultaneously. The fusion-EKF predicts and updates multiple tracks from inhomogeneous sensors. The association is performed with coordinate transformation and EBs estimations. The noises resulting from sensors are correspondingly processed by separating covariance matrix to ensure the reliability of detections. EB estimations and the corresponding alignment of time and region searches are conducted within the fusion-EKF. The fusion system uses both sensors advantages and minimizes EBs for better resolutions on different perspectives from different sensors due to fusion EBs. The RMSEs of range, angle, and doppler are shown in the experimental results section. Cross-validating with the cameras pixel data and the radars point clouds from raw data, the radar-camera fusion-EKF system provides a more accurate and more reliable detection of targets.

In Chapter 3, a real-time behavior detection system using mmWave radar is presented. Radar is used to sense the micro-Doppler information of targets. A CNN is further implemented in the detection and classification of the human motion behaviors using this information. Both the convolution layers and the architecture of CNNs are presented. Analyses of loss and accuracy of training results are also delineated. The experimental results indicate a precise determination of human motion behavior detection.

In Chapter 4, a robust and adaptive radar point cloud clustering algorithm, referred to as radar elliptical density-based spatial clustering of applications with noise (REDB-SCAN), is presented. The proposed algorithm yields better clustering results for adapting to arbitrary shape of targets as well as any number of targets relative to traditional clustering methods. The algorithm is presented and is implemented in
experiments using the state-of-the-art mmWave radar sensor with MIMO antennas. The related signal processing chain and the clustering outcomes are also discussed.

In Chapter 5, a new millimeter wave 3D imaging radar is presented. The user must simply move the radar along a circular track, and high-resolution 3D imaging can be generated. The proposed radar uses the movement of itself to synthesize a large aperture in both the azimuth and elevation directions. It can utilize inverse Radon transformation to resolve 3D imaging. To improve the sensing result, the compressed sensing approach is further investigated. The simulation and experimental results further illustrate the design. Because a single transceiver circuit is needed, a light, affordable, and high-resolution 3D mmWave imaging radar is illustrated.

In Chapter 6, a novel approach applying traditional airborne SAR techniques to 77 GHz automotive radar range-doppler response is elucidated. A 77 GHz automotive radar SAR terrain mapping experiment is conducted. The chapter provides the terrain mapping of the parking lot. Aside from current lidar mapping techniques, a single transceiver mmWave imaging system is used. With odometer, GPS and inertial measurement unit (IMU) data from autonomous driving vehicle, a range-doppler algorithm on mapping static targets is achieved. The experiment results indicate the current research progress on automotive radar terrain mapping on moving vehicle.

Finally, the conclusions and future works are discussed in Chapter 7.
2. EXTENDING RELIABILITY OF MMWAVE RADAR TRACKING AND DETECTION VIA FUSION WITH CAMERA

2.1 Background

Radar, working at W-band, is becoming an important sensor in the advanced driver assistance system (ADAS) and autonomous driving fields [6–10]. Research and industry have done many works towards safety, reliability, compact and low-cost sensor systems. Multiple sensors of ADAS system, like radar, camera and Lidar, are used for different purposes as they have different strengths. The lastest millimeter wave (mmWave) radars at W-band are surging for automobile applications, e.g., adaptive cruise control, pedestrian detection, collision avoidance, lane changing monitoring and emergency braking. mmWave radar researches are glowing according to the recent increasing demand of automotive radars in ADAS and autonomous driving industry. For instance, the automotive target shape (height and width) estimation using relaxation algorithm [11]; the super-resolution automotive radars [12]; and the automotive radar sensor fusion with other sensors to improve target tracking and classification [9]. To take advantage of other sensors for sensing accuracy and reliability, we are focusing on the mmWave radar sensor fusing with camera sensor in this chapter.

Sensors reliability, especially for avoiding false detection, miss detection, blocking, blind spot, adverse weather and failure, are essential in ADAS and autonomous driving topics. One way to reduce all uncertainties and failures is to fuse sensor outputs [13]. Sensor fusion has been developing rapidly in recent years. However, there are limited researches fusing mmWave radars with other sensors due to the fact that
radars provide limited number of detection points representing interest targets [9], which make it difficult to recognize from a snapshot of radar detection. But if this problem of mmWave radars can be solved, it can increase the reliability of detecting moving targets, avoiding blockage and tracking dramatically. In this chapter, we aim to increase mmWave radar’s informative capability about targets and its versatility by fusion with monocameras.

mmWave radars have been implemented to fuse with Lidars [14, 15] in recent years. In [14], Hajri et al. introduced radar and Lidar real-time kinematic (RTK) sensor fusion. Provided with mean square error (MSE) with radar, Lidar and fusion variance in the experiment, the result is convincing. By default, radar and Lidar, each can provide detections in the birds-eye-view (BEV) plane which is top view onto the ground [16]. But Lidar works on the infrared band, which is prone to interference, expansive and bulky in size. Not many researches are done with associating inhomogeneous sensors like cameras or infrared sensors with radars [17, 18]. Because the camera’s and radar’s detections are unrelated and difficult to align without certain assumptions. In this case, we are implementing a homography estimation method (HEM) [19] and creating a track-oriented association and fusion algorithm for calibration and tracking.

Kalman filters (KFs) are using a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone [20]. Difference between the extended Kalman filter (EKF) [21] and traditional KF is that EKF can handle linear equations as well as nonlinearities. EKF uses the Jacobian matrix to linearize nonlinear functions, e.g., use EKF as radar’s Doppler information (range rate). In this chapter, we are not considering unscented Kalman filter (UKF) because radar nonlinearity comes from Doppler. And Doppler nonlin-
earity is the first order approximation of range. Thus UKF can be unstable and its results will be biased when estimating radars Doppler velocities.

In this chapter, a HEM is applied in associating inhomogeneous sensors via their error bounds (EBs). A new fusion-EKF is designed to take both radar and camera as inputs and associate the data inside the filter to obtain an ideal target tracking outputs. Data association of the fusion-EKF are also capable to support tracking of multi-targets. In [22], Muresan et al. presents a two-step data association method for Lidar. It is based on spreading parameter weights from KF. In this chapter, the weights are further studied to support the two heterogeneous sensors. Specifically, the EBs are introduced in data association as radar provides raw Doppler velocities and camera vision generates the bounding boxes of targets. EB supports the two sensors to form a clear gating for the association of the fusion-EKF. Thus comparing to other weight-based or probability-based association methods, the proposed fusion-EKF with EBs will provide an extra dimension in association for inhomogeneous sensors. Motion models selection for multi-target tracking can also be an important part of KF tracking. In [23], the survey on different motion models is introduced and compared. In our setup, constant acceleration (CA) model is used because we assumed targets can have the linearity of the state transition equation which allows an optimal propagation of the state probability distribution of humans and cars. Yaw angle and turn rate are less essential in these models. Thus it allows tracks to have more flexibility in association from inhomogeneous sensors.

We also noted another type of efforts for sensor fusion using machine learning methods. Their sensor fusions are typically performed via concatenating tensors inside the neural network [24]. However, it requires to build a large neural network for high resolution radar and cameras, which makes it difficult to implement for real-time applications. Similarly, in [25], Zhong et al. introduced the Texas Instruments’ (TI) commercial TDA3x/TDA2x boards for camera radar fusion. It is so far the
most reliable fusion system for mmWave radars detection compound with camera’s vision from from the best of our knowledge our best of our perspective. In their paper, a complete fusion system can provide tracking and detections over time. The shortcomings of this research is that, the object detections and classifications relies on machine learning techniques with online computer vision libraries. The complete pipeline is difficult for real-time applications. In the meantime, radar noise handling is not ideal for various scenarios. In our work, we strongly overcome these shortcomings.

In this chapter, we are using monocamera and mmWave radar as RTK sensors with real-time fusion algorithms running for tracking. A reliable detection and cross-validated target tracking is realized. sensor fusion and association are done within the fusion-EKF using HEM, timeline alignment and region search. No machine learning is applied and thus the system consumes much less calculation load. Our experiment shows that the proposed system can provide a reliable tracking and detecting with low calculation costs. An embedded system like Arduino or Raspberry Pi can process the data for real time applications.

For the new fusion-EKF, we introduced a new concept, i.e. error bounds. It is defined as the sensor’s region of approximation. EBs are not from the uncertainty of sensors [26] but the sensors’ resolutions from their perspectives. Some papers [17,18,27,28] illustrated the alignment between radars and other inhomogeneous sensors. They assumed radars and other sensors are working at identical principle. But in fact, they are inhomogeneous in dimensions, processing chain and sensing wavelength. In [27], vision lateral position improvement of radar detections are introduced. Their work works for radar alignment onto vision well using cross-correlation. However, errors are not corrected during detections. In [28], Folster et al. introduced lateral velocity estimation for automotive radar sensors for ADAS. The method simply refers to the location of the target on the actual detections. The disadvantages are obvious: not many cross-correlations between sensors; approximation relies only
on locations. To overcome limitations in these cases, EBs proposed in this chapter are essential to apply to radar-vision fusion-EKF to improve region search and tracking. The HEM further enhances the transformation and track-oriented association between inhomogeneous sensors, which makes the proposed fusion system to take both sensors’ advantages, and be robust even one sensor is temporarily out of work.

In summary, our contributions are:

1. Building a new real-time sensor fusion system for mmWave radar and camera.
2. A new fusion-EKF is designed to support the two heterogeneous sensors.
3. A new concept, EBs, are defined for data association and gating inside the fusion-EKF.
4. A HEM is proposed for building the transformation matrix between mmWave radar and cameras.

The structure of this chapter is as follows. In Section 2.2, the methodology is presented including radar and camera data preprocessing, radar-camera fusion-EKF and sensor synchronization. In Section 2.3, results of radar-camera fusion-EKF and root mean square errors (RMSEs) of improved EBs are shown.

### 2.2 Methodology

In this chapter, a new fusion system for mmWave radar and camera is proposed. We will first introduce the EBs in Section 2.2.1. It is a fundamental concept in the fusion system for data association. Then, we will discuss the processing of the two sensors and a HEM method for finding out the transformation matrix between the two sensors in Sections 2.2.2 and 2.2.3. The new fusion-EKF will be discussed in Section 2.2.4. The data association and synchronization will be introduced in
Section 2.2.5. The fusion system performance will be evaluated based on RMSE as discussed in Section 2.2.6.

2.2.1 Coordinates and Error Bounds

In this sensor fusion system, two different coordinates sensors are used, i.e., monocamera and mmWave radar. In Fig. 2.1, all the coordinates in this fusion system are presented.

With a monocamera and 1D phased array mmWave radar, the fusion can improve the detection to a 3D space. For the fused sensor system, we use a left hand co-
ordination system, shown in Fig. 2.1a. The $x$-axis is the range axis which is away from the sensor center. $y$-axis is the cross-range axis, *a.k.a.*, azimuth axis. $z$-axis is the elevation axis. In Fig. 2.1b, monocamera vision plane with monocamera sensor placed at zero is presented. $(u, v)$ pixel variables at camera plane are used with $u$ representing horizontal pixel and $v$ representing vertical pixel, respectively. In Fig. 2.1c, The mmWave radar sensor is conducting measurements with $(\rho, \theta, v_d)$, which are radar measurement range, azimuth angle and Doppler velocity (*a.k.a.* range rate), respectively.

If both sensors are presented, with camera and radar a slightly vertical difference in position, shown in Fig. 2.1d, a fused system reconstructs the coordinate system with planes of vision and mmWave radar detections perpendicular to each other. As vision data is not totally limited to one plane. The blockage of some targets may exist on different range and azimuth locations. Thus the vision plane is penetrating throughout range and azimuth. In this case, the fusion can be constructed with correlations of $(u, v)$ and $(\rho, \theta, v_d)$. Additionally, Cartesian and polar conversions are throughout the fusion process. $\theta$-axis and azimuth axis are used in following statements.

Error bounds of inhomogeneous sensors have different coordinate systems as stated above. Here we need to define how the fused system will behave compared before and after the sensor fusion.

**Definition 2.2.1** Measurement point of the next inquiry lays within a certain region of approximation, either it is the measurement from vision sensor or radar sensor. The region is then the theoretical error bounds of the sensor.

Error bounds of this sensor fusion system is defined as following.

**Definition 2.2.2** The actual localization of the detecting target lays within the approximate prediction region. The approximation is estimated by two or more sensors’
system, and the region takes advantages from sensors’ error bounds. The prediction region is then the fused error bounds.

With Definitions 2.2.1 and 2.2.2, the error bounds of fused sensor system can be plotted as Fig. 2.2. For vision error, the detecting target is difficult to solve the location along range axis. For radar error, the detecting target is limited as solving along $\theta$-angle axis, which is caused by 4-by-2 transceivers configuration (15 degrees $\theta$ resolution at center). Fusion takes both advantages, which results in better range and better angle solving. However, in this application, we do not solve elevation due to no data on elevation from radar sensor. With better mmWave radars, it is able to get the elevation error bounds.
2.2.2 Camera and mmWave Radar Preprocessing

Cameras, including stereo and mono cameras, have intrinsic and extrinsic parameters which are inherent due to the physical build of the pinhole and lens [29–32]. The extrinsic parameters with rotation and translation matrix are intended to map the world coordinates of the objects into camera coordinates. On the other hand, intrinsic parameters deals with the camera coordinates into the image pixel coordinates \((u, v)\).

In this chapter, we generally follow the four-step procedure from paper [32] and calibrate on the robotic operating system (ROS). In this case, intrinsic, extrinsic and lens distortion parameters of the camera are prepared for warping and distortion image processing [33]. The camera calibration for this fusion project is shown in Fig. 2.3. Fig. 2.3a shows the sample of detected checkerboard points in image coordinates. With help from the extrinsic visualization in Fig. 2.3b, multiple samples are collected and used to estimate intrinsic parameters.

For mmWave radars, the radar sensor configuration determines its range resolution \((\Delta \rho)\), bearing resolution \((\Delta \theta)\) and Doppler resolution \((\Delta v_d)\).

Range resolution \(\Delta \rho\) can be calculated from radar sweeping chirp bandwidth:

\[
\Delta \rho = \frac{c_0}{2B},
\]

where

\(c_0\) : the electromagnet wave speed

\(B\) : the bandwidth of the radar chirp signal.

The bearing resolution physically depends on the number of virtual antennas MIMO radar emulated. The angular resolution is given as:
(a) Checkerboard calibration for intrinsic parameters

(b) Extrinsic parameters visualization

Figure 2.3.: Camera calibration.

\[ \Delta \theta = \frac{c_0}{f_c d_{\text{RX}} N_{\text{TX}} \cos \theta_i}, \]  

(2.2)
where

\[ f_c : \text{center frequency of linear frequency modulated signal} \]
\[ d : \text{receiver element spacing, typically } \lambda/2 \]
\[ N_{RX} : \text{number of receivers (RX)} \]
\[ N_{TX} : \text{number of transmitters (TX)} \]
\[ \theta_i : \text{the angle of interest}. \]

For Doppler resolution, it is inherent detections from frequency modulated continuous wave radars, and it provides the target range changing rate. Doppler resolution is defined as:

\[ \Delta v_d = \frac{c_0}{2 f_c (PRI) N_d}, \]

(2.3)

where

\[ PRI : \text{the pulse repetition interval} \]
\[ N_d : \text{number of chirps per radar frame}. \]

PRI is commonly the total chirp time of a single frequency modulated signal. Thus \( \Delta v_d \) is typically restricted by number of chirps and chirp repetition interval.

An important part for mmWave radar tracking using EKF is that mmWave radar measures targets’ scatters with compound noises. These noises come from:

1. Multipath effect [34] results detecting targets to undergo ghost or fading where a phase-shifted reflection is produced by walls or ground.

2. mmWave radar has its inherent measurement noise (\( R_{\text{radar}} \)). In this case, we are trying to find the noise model by calculating from the provided configuration
we set at radar device and the variance of the point cloud and SNR of actual measurement.

3. Measurement processing noise from stretch processing and FFTs produces noise point cloud points. It is typically dense when multiple targets are being detected. At the meantime, a higher processing rate causes the processing noise lifting.

4. Antenna gain contributes to the SNR at different angles. This may cause SNR at the center higher than SNR at the side.

For the first case, which is the multipath effect in 77 GHz automotive radar, a typical symptom is that some ghost targets are produced with random phase but similar shape and amplitude compared to the original target. A method to reduce when applied to tracking is to use multilayer perceptron [35]. In this case, the ghost is largely reduced and then the tracking becoming more reliable.

For the second case, $R_{\text{radar}}$ obeys the Gaussian distribution. $R_{\text{radar}}$ is then

$$R_{\text{radar}} \sim \mathcal{N}\left( \begin{pmatrix} \mu_ho \\ \mu_\theta \\ \mu_v \\ d \end{pmatrix}, \begin{pmatrix} \sigma^2_\rho \\ \sigma^2_\theta \\ \sigma^2_v \\ \sigma^2_d \end{pmatrix} \right). \quad (2.4)$$

The $\begin{pmatrix} \mu_ho & \mu_\theta & \mu_v \\ d \end{pmatrix}^\top$ matrix, literally is targets’ detection matrix. It specifies target’s range, bearing and Doppler velocity. On the other hand, the $\begin{pmatrix} \sigma^2_\rho & \sigma^2_\theta & \sigma^2_v \\ \sigma^2_d \end{pmatrix}^\top$ matrix, is the variances matrix of the targets, which is the uncertainty of the target’s range, bearing and Doppler velocity. So for the EB of mmWave radars, the $\begin{pmatrix} \sigma^2_\rho & \sigma^2_\theta & \sigma^2_v \\ \sigma^2_d \end{pmatrix}^\top$ matrix represents correspondingly. With all the resolution discussed in Section 2.2.1, the relations are below:
\[
\begin{pmatrix}
\Delta \rho^2 \\
\Delta \theta^2 \\
\Delta v_d^2
\end{pmatrix}
= \begin{bmatrix}
\frac{1}{4} & 0 & 0 \\
0 & \frac{1}{4} & 0 \\
0 & 0 & \frac{1}{4}
\end{bmatrix}
\begin{pmatrix}
\sigma^2_\rho \\
\sigma^2_\theta \\
\sigma^2_{v_d}
\end{pmatrix}.
\] (2.5)

Here, we have the assumptions that the discrete Fourier transform (DFT) size is the same as input data size. However, as direction of arrival (DOA) estimation can be estimated by conventional DFT/FFT method and several high resolution algorithms, e.g., MUSIC algorithm [36] and ESPRIT algorithm [37]. The \(\Delta \theta^2\) term should be modified to \(\Delta \theta_{\text{DOA}}^2\) according to number of angle bins. The 0 represents the uncorrelated from range to angle, angle to Doppler and range to Doppler. \(1/4\) term is the minimum requirement to avoid aliasing. Therefore, the EB of mmWave radar is defined.

For the third case, assume mmWave is a linear system with an output range \(\rho_o\) which is the response of an input signal \(s\), the measurement noise \(e\) can be obtained by:

\[
\rho_o = \rho + e = hs + e, \tag{2.6}
\]

where \(\rho\) is the raw measurement from sensor and \(h\) is the transform from input signal to output signal. Using least-squares estimation from orthogonal principles [38], by choosing appropriate \(h\) to minimize the Euclidean distance, the minimized projection error \(e\) is orthogonal to every column of \(s\) if and only if [39]:

\[
s^\top e = 0, \tag{2.7}
\]

Therefore, by orthogonality,

\[
h = (s^\top s)^{-1}s^\top \rho_o. \tag{2.8}
\]
In this system, the measurement processing noise is minimized with the least sum of $\|e\|^2$ to reduce the effects onto sensors’ fusion. In later EKF stage, this measurement noise is only depending on time variance of the receiving signal from equations. And it is along the inherent measurement noise $R_{\text{radar}}$ to produce the EKF estimations.

In the EKF in this application, the processing noise is then the variance from the processing time from the radar board. Here we say the processing covariance noise from radar is $Q_{\text{radar}}$, $Q_{\text{radar}}$ is related to the motion variance of acceleration of the object and the processing time variance $\Delta t$ from radar. Similarly, we have camera’s $Q_{\text{cam}}$ which is associated with camera’s processing time variance.

For the last case, the antenna gain resulted in SNR difference. It will result in center detections less noisy than the surroundings. When it comes to sensor fusion, the radial detections around corners are noisy and may have miss detections. The resulting possible failure should be compensated by other sensors like cameras. Note that antenna gain in this application only reflects signal-to-noise ratio (SNR) at different detecting angles. A TI mmWave radar is used and it is similar when applying to other mmWave radars. Measurements in anechoic chamber is conducted and the antenna gain at 0 degrees is 10 dB compared to 0 dB at ±60 degrees. The reference can be seen in TI’s document [40].

2.2.3 HEM

The transformation between inhomogeneous sensors assumes the relation exists between targets detected in different sensors’ domain. A HEM for associating radar’s detecting plane and camera’s image plane is used:
\[
\begin{pmatrix}
  u \\
  v \\
  1
\end{pmatrix}
= \begin{pmatrix}
  x \\
  y \\
  1
\end{pmatrix},
\]
(2.9)

where \(c\) is non-zero constant and \(x, y\) is the coordinates of warped image in world coordinates. And \(T\) is the transformation matrix which we are interested in warping image plane to world plane. It is a 3-by-3 matrix and can be decomposed as:

\[
T = \begin{bmatrix}
  T_1 & T_2 & T_3
\end{bmatrix}^T.
\]
(2.10)

If we collect camera vision data with a calibrated camera. And try to associate the radar’s localization of the specific point target like corner reflector. With Moore-Penrose pseudo inverse [41], the computing of the Euclidean norm solution can be obtained by linking several measured pairs. By applying least squares to the HEM, the transformation matrix \(T\) can be obtained as:

\[
\begin{cases}
  T_1 = (J^TJ)^{-1}J^TU \\
  T_2 = (J^TJ)^{-1}J^TV \\
  T_3 = (J^TJ)^{-1}J^I
\end{cases},
\]
(2.11)
where

\[
J = \begin{bmatrix}
  x_1 & y_1 & 1 \\
  x_2 & y_2 & 1 \\
  \vdots & \vdots & \vdots \\
  x_n & y_n & 1
\end{bmatrix}
\]

\[
U = \begin{bmatrix}
  cu_1 \\
  cu_2 \\
  \vdots \\
  cu_n
\end{bmatrix}
\]

\[
V = \begin{bmatrix}
  cv_1 \\
  cv_2 \\
  \vdots \\
  cv_n
\end{bmatrix}
\]

\[I: \text{identity matrix}\]

\[n: \text{number of measurements}\]

When collecting \(n\) samples of pairs from world coordinates to the camera image plane, the warped BEV can be best estimated by the transformation matrix \(T\).

After warped view from image towards BEV is applied, the BEV view of warped image can be used to associate with radar’s detection point cloud map. The warped BEV image with BBox is consistently used in fusion-EKF for further processing. With only the \((x, y)\) coordinates for objects, the radar’s detections with range, azimuth angle and Doppler velocities are then fused.

The bounding box (BBox) generation is coming from background subtraction algorithm (BSA) [42]. In future processing within the EKF, the BBoxes and radar clusters are associated according to the algorithm introduced in Section 2.2.5.

### 2.2.4 Fusion-EKF

The full workflow of this radar-camera fusion-EKF is shown in Fig. 2.4.
From Fig. 2.4, the radar, camera and fusion updates are shown in yellow, blue and green, respectively. Radar raw data is feeding with ultra short range (USR) config
on digital signal processing (DSP) chip module on TI radar. With output through USB port to host computer, the point cloud with Doppler information is extracted. Additional filtering like moving target indicator (MTI) is applied to collaborate with camera’s BSA. DBSCAN [43] or REDBSCAN [44] clustering algorithm is applied to get the clusters and potential target shape from radar scans. These are important parts to identify targets and similarly create ”bounding box” for radar targets before the radar-camera fusion-EKF. After the first stage noise deductions and the query data is sent to radar-camera fusion-EKF for further association and tracking.

Similarly, camera raw data is feeding with intrinsic and extrinsic parameters with exact locations and directions measured during the experiment. The calibration provides the BEV from warping the image onto the perspective view. The BSA is then used for image plane BBox generation. The clustering in radar processing and the BBox generation in vision processing have inherent association. But without machine learning applied or segmentation, the association should be done within the radar-camera fusion-EKF. Details will be discussed in Section 2.2.5.

The fusion-EKF uses the CA maneuvering model in our setup. The CA model is a simplified model of Singer model [45] without acceleration factor thus lower orders in time deviation to improve the processing speed. For tracking multiple targets with different velocities and positions, the CA model can handle with different preset acceleration noises.

The key difference in our setup compared to other researches is that EBs are updated by both sensors and this radar-camera fusion-EKF updates both vision BBox plane and radar localization BEV continuously. In this case, the time stamp of prediction and update is different based on each sensor’s fps. This will be discussed in Section 2.2.5. We set the current state $k$ and thus the previous state is $k – 1$. 
Prediction

The prediction part of this radar-camera fusion-EKF is no difference from typical single sensor maneuvering tracking except the CA model. If we have a state vector of a target \( \mathbf{x} \) with position \( p \) and velocity \( v \), we can define the state vector as a four-element vector with position and velocity projections onto \( x \)-axis and \( y \)-axis. That is:

\[
\mathbf{x} = (p, v)^T = (p_x, p_y, v_x, v_y)^T.
\]  \hspace{1cm} (2.12)

State prediction does not depend on whether it is from radar or camera. But it will predict either radar or camera data is coming. Here, we need a state transition matrix which predicts the EKF output with state transition matrix:

\[
\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{w},
\]  \hspace{1cm} (2.13)

where

- \( \mathbf{F}_k \): state transition matrix
- \( \mathbf{x}_{k-1} \): previous state \( k - 1 \) state vector
- \( \mathbf{B}_k \): control input matrix
- \( \mathbf{u}_k \): control input vector
- \( \mathbf{w} \): compound noise from system
  
  \((\text{motion noise and processing noise}).\)

The \( \mathbf{w} \) part, is obeying zero mean Gaussian distribution

\[
\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q}),
\]  \hspace{1cm} (2.14)
where $Q$ is the processing noise introduced before as $Q_{\text{radar}}$ or $Q_{\text{cam}}$. As time stamps are changing, the time stamp difference $\Delta t$ should be updated throughout the process.

From the normal distribution of noise, an uncertainty covariance can be obtained for linearity. The uncertainty covariance matrix, also often referred to priori error covariance, for the prediction step is:

$$
P_k = F_k P_{k-1} F_k^\top + Q_k. \quad (2.15)
$$

### Update

Similarly to the prediction state vector definition, we need an update measurement vector $z$ which is consisting of radar’s and camera’s measurement. But both are following the same procedures. For radar data,

$$
z_{\text{radar}} = (\rho, \theta, v_d)^\top, \quad (2.16)
$$

and for camera data,

$$
z_{\text{cam}} = (u, v)^\top. \quad (2.17)
$$

The update consists a residual $y$, which is the error from measurement from the prediction

$$
z_k = H_k x_k + y_k, \quad (2.18)
$$

where

$$H_k : \text{measurement function matrix}.$$

The $y$ part, is obeying

$$y \sim N(0, R). \quad (2.19)$$
where the $\mathbf{R}$ is the inherent measurement noise introduced before as $\mathbf{R}_{\text{radar}}$ or $\mathbf{R}_{\text{cam}}$. The random variable is also a normal distribution. Thus the measurement error covariance matrix is:

$$
\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\intercal + \mathbf{R}_k .
$$

(2.20)

Here we can obtain the optimal Kalman filter gain $\mathbf{G}$:

$$
\mathbf{G}_k = \mathbf{P}_k \mathbf{H}_k^\intercal \mathbf{S}_k^{-1} .
$$

(2.21)

The state vector update $\mathbf{x}_{k(\text{update})}$ is based on prediction:

$$
\mathbf{x}_{k(\text{update})} = \mathbf{x}_k + \mathbf{G}_k \mathbf{y}_k .
$$

(2.22)

$\mathbf{x}_{k(\text{update})}$ is used for next state prediction step as $\mathbf{x}_{k-1}$. Similarly, it is also applied to the uncertainty covariance matrix $\mathbf{P}_{k(\text{update})}$:

$$
\mathbf{P}_{k(\text{update})} = (\mathbf{I} - \mathbf{G}_k \mathbf{H}_k) \mathbf{P}_k .
$$

(2.23)

The uncertainty covariance, is often used to scale the response of the Kalman filter. Commonly, in terms of representing the response, the Kalman filter gain $\mathbf{G}$ is used. It represents the relative weight of the measurements compared to the current state estimate. And it is typically changing over time as noise changing throughout time. Rewriting the Equation (2.21) with Equation (2.20):

$$
\mathbf{G}_k = \frac{\mathbf{P}_k \mathbf{H}_k^\intercal}{\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\intercal + \mathbf{R}_k} .
$$

(2.24)

Thus we can conclude on the $\mathbf{G}$ as it depends on $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\intercal$ and $\mathbf{R}_k$ terms. As $\mathbf{R}_k$ is fixed in a single measurement and not changing throughout $k$ steps. When $\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^\intercal$ is much higher than $\mathbf{R}_k$ term, the $\mathbf{G}$ is smaller. The low $\mathbf{G}$ will result in the model relying more on predictions. Hence it is less responsive to measurements. And on
the other hand, when $H_k P_k H_k^T$ is much smaller than $R_k$ term, the $G$ is larger. And
the fusion-EKF puts more weights on the current measurement. Thus it is more
responsive to measurements. The $G$ will also affect the region query thus the EBs of
both sensors and it will be discussed in Section 2.2.5.

**Nonlinearity**

One of our contributions is to use nonlinearity property in EKF to deal with radar’s
Doppler data because Doppler is the range rate, which is the partial derivative of the
range with time.

Because radar has detections in polar coordinates and BEV is in Cartesian coor-
dinates, the conversion is needed for every incoming measurement. And additionally,
from Equation (2.18), the $H$ matrix should have a convertible formation to map radar
detections from the state vector $x$ to analyze $y$. If we change the Equation (2.18) for
radar observations:

$$ z_k = h(x_k) + y_k. $$

The $h(x_k)$ function represents a way to convert position and velocity of state vector
to a radar coordinates. Hence the $h(x_k)$ function can be further updated as:

$$ h(x_k) = \begin{pmatrix} \rho \\ \theta \\ v_d \\ \frac{\partial \rho}{\partial t} \end{pmatrix} = \begin{pmatrix} \rho \\ \theta \\ \frac{p_x v_x + p_y v_y}{\sqrt{p_x^2 + p_y^2}} \\ \frac{\sqrt{p_x^2 + p_y^2}}{\arctan(p_y/p_x)} \end{pmatrix}. $$

(2.26)
Head back to the $H$ matrix, the 3-by-4 matrix has higher dimensional components as $h(x_k)$ function indicates. Thus it is a nonlinear function in terms of radar-camera fusion-EKF. A way to deal with this is to use Jacobian matrix on $H$:

$$H = \left[ \frac{\partial h(x_k)}{\partial x_k} \right]_{x_{k-1}} =
\begin{bmatrix}
\frac{\partial \rho}{\partial p_x} & \frac{\partial \rho}{\partial p_y} & \frac{\partial \rho}{\partial v_x} & \frac{\partial \rho}{\partial v_y} \\
\frac{\partial \theta}{\partial p_x} & \frac{\partial \theta}{\partial p_y} & \frac{\partial \theta}{\partial v_x} & \frac{\partial \theta}{\partial v_y} \\
\frac{\partial v_d}{\partial p_x} & \frac{\partial v_d}{\partial p_y} & \frac{\partial v_d}{\partial v_x} & \frac{\partial v_d}{\partial v_y}
\end{bmatrix}.$$ (2.27)

Combining both of Equations (2.26) and (2.28), the resulting $H$ for radar is:

$$H =
\begin{bmatrix}
\frac{p_x}{\sqrt{p_x^2 + p_y^2}} & \frac{p_y}{\sqrt{p_x^2 + p_y^2}} & 0 & 0 \\
-\frac{p_y}{p_x^2 + p_y^2} & \frac{p_x}{p_x^2 + p_y^2} & 0 & 0 \\
\frac{p_y(v_y p_y - v_x p_x)}{(p_x^2 + p_y^2)^{3/2}} & \frac{p_x(v_y p_x - v_x p_y)}{(p_x^2 + p_y^2)^{3/2}} & \frac{p_x}{\sqrt{p_x^2 + p_y^2}} & \frac{p_y}{\sqrt{p_x^2 + p_y^2}}
\end{bmatrix}. \quad (2.28)

2.2.5 Data Association and Sensor Synchronization

The track-oriented data association is implemented using timeline alignment of sensors to get updates from inhomogeneous sensor’s perspective and synchronization over certain regions of EKF output to obtain the tracked target localization. This process is overlying on the radar-camera fusion-EKF in Fig. 2.4. With HEM to transform between coordinates, EBs are recalled to optimize the association of radar
clusters and camera BBoxes. The fusion-EKF track is created or deleted according to the existence from both sensor.

Timeline Alignment

The timeline alignment for radar-camera fusion system can be seen in Fig. 2.5. From different fps of both sensors, observations of different sensor type on different time stamps are collected on the processing unit. And fusion association is shown with different $\Delta t$ with different state steps $k$ in Fig. 2.5a. The green colored step on the fusion timeline actually means dual sensors are outputting their observations at the same time ($\Delta t \leq 0.005s$). In this case, predictions and updates follow the same rule as the radar-camera fusion-EKF.

If we set the model with a random acceleration vector $\mathbf{a} = (a_x, a_y)^\intercal$ with variances of $\sigma_a x^2$ and $\sigma_a y^2$ on $x$-axis and $y$-axis, respectively. In this case, the $\mathbf{w}$ noise matrix which is the motion EB for the tracked model, can be expressed as:

$$
\mathbf{w} = \begin{pmatrix}
w_{px} \\
w_{py} \\
w_{vx} \\
w_{vy}
\end{pmatrix} = \begin{pmatrix}
a_x \Delta t^2 \\
a_y \Delta t^2 \\
a_x \Delta t \\
a_y \Delta t
\end{pmatrix}.
$$

From Equation (2.14), the noise is described by a zero mean and a covariance matrix $\mathbf{Q}$. When referring to Equations (2.15) and (2.29), with variances of $\sigma_a x^2$ and $\sigma_a y^2$ applied, the $\mathbf{Q}$ matrix can be estimated by taking expectations of the $\mathbf{w}$ and $\mathbf{w}^\intercal$ matrix as:
(a) Radar-camera association timeline. A sample state steps $k - 2, k - 1, k, k + 1$ are shown in the fusion timeline.

(b) Association actions based on timeline alignment of state steps $k - 2, k - 1, k, k + 1$.

Figure 2.5.: Radar-camera association timeline alignment.

$$Q = \mathbb{E}[ww^T]$$

$$\begin{bmatrix}
\frac{\sigma_{ax}^2 \Delta t^4}{4} & 0 & \frac{\sigma_{ax}^2 \Delta t^3}{2} & 0 \\
0 & \frac{\sigma_{ay}^2 \Delta t^4}{4} & 0 & \frac{\sigma_{ay}^2 \Delta t^3}{2} \\
\frac{\sigma_{ax}^2 \Delta t^3}{2} & 0 & \sigma_{ax}^2 \Delta t^2 & 0 \\
\end{bmatrix}. \quad (2.30)$$
Hence \( Q \) is depending on \( \Delta t \) and the random acceleration variances. As \( \Delta t \) changing from time to time because of different fps of camera and radar. When with high \( \Delta t \), the uncertainty and the large EB are generated. And vice versa.

Similarly applied to measurement noise matrix \( R_{\text{radar}} \). Combining Equations (2.1) to (2.5). The result \( R_{\text{radar}} \) is:

\[
R_{\text{radar}} = \mathbb{E}[yy^\top] = \begin{bmatrix}
\sigma^2_\rho & 0 & 0 \\
0 & \sigma^2_\theta & 0 \\
0 & 0 & \sigma^2_{v_d}
\end{bmatrix}
\]

(2.31)

In Fig. 2.5b, as we have two tuned updates for each step depending on which sensor data is obtained, the prediction and update for each observation from radar or camera are shown. As stated in the previous section, the updated \( x \) and \( P \) matrices are used for the next prediction step. In other words, the measurement update of \( x \) and \( P \) is only based on the prediction within the same step. And prediction relies on the previous update step.

**Region Search within EBs**

We have the idea of the Kalman filter gain \( G \) in the Section 2.2.4. Here is the association of radar clusters and camera BBoxes based on Kalman filter gain \( G \) using HEM. In Fig. 2.6, the fusion-EKF trajectory in blue dash line is shown with steps \( k - 1, k, k + 1 \). The different predictions after update measurements are shown with positions and velocities. The yellow region query for the upcoming measurement is set in Fig. 2.6. Here is the EBs resulting from Definitions 2.2.1 and 2.2.2:
Figure 2.6.: Radar-camera association region search. Blue dash line: EKF projected trajectory, yellow region: search region for next state for radar clusters and camera bounding box BEV projection.

1. EB along $x$-axis of radar is

$$EB_{|x_{\text{radar}}} = \sigma \rho \cos \theta_i,$$  

(2.32)

where $\theta_i$ is the angle of interest from target, typically $\mu_\theta$, which is the target angular location. EB along azimuth ($y$-axis) of radar is

$$EB_{|y_{\text{radar}}} = \sigma \rho \sin \theta_i.$$  

(2.33)

The assumption is that the radar has even spacing MIMO antenna with even angular resolution. Thus we can conclude that at the center, radar has better range EB and less accurate when it is on the side. On the other hand, radar has relative bad azimuth EB and needs other sensors to compensate for the angular EB.
2. EB along range of camera is

\[
EB|_{x_{\text{cam}}} = c(T^T)^{-1}T^rU.
\] (2.34)

And EB along azimuth of camera is

\[
EB|_{y_{\text{cam}}} = c(T^T)^{-1}T^rV.
\] (2.35)

For camera pixels using HEM, warping along the vertical axis to \(x\)-axis is typically limited compared to the horizontal axis to \(y\)-axis. Thus the range resolution is restricted especially for long ranges. And azimuth resolution is consistent when looking at horizontal pixels to angle pair. So camera EB along range is worse compared to radar’s. And EB along azimuth is better with adequacy horizontal pixels.

3. Combining radar and camera EBs. There exist two standalone query regions to search for cluster or BBox from inhomogeneous sensors.

4. Fusion EBs use EKF output of new prediction to get the estimated localization of the cluster/BBox. Thus allows missing clusters/BBoxes or multiple clusters/BBoxes. This unique feature is what fusion-EKF provides.

Even with fusion EBs, there still exists multiple clusters/BBoxes or missing clusters/BBoxes. Reasons include:

1. Concealed by front object using a camera.
2. Sensor failure by radar or camera.
3. Radar’s multipath effect.
4. Lighting condition (over exposed or no lighting) by camera.
5. Others.

In Algorithm 2, it shows the region search and track updates for tracking multiple targets. Different tracks share the incoming clusters/BBoxes measurement updates from sensors. Fusion-EKF deals with the region search based on EBs estimations. The fusion-EKF provides invisible counts for inhomogeneous sensors. If either sensor fails temporary, the tracking will continue using the other sensor’s update. The tracking will still be valid if missing several clusters/BBoxes. However, track will be terminated after a certain duration of missing detections from both sensors based on cross-validation. With updates for fusion localization, the improved EBs can provide a better result for the system thus greatly decreases uncertainty errors from either sensor. In addition, new tracks are initialized based on cross-validation from different sensors with certain threshold.

Therefore, the association is done with timeline alignment together with EBs region search within tracks. With help from different advantages from inhomogeneous sensors, fusion EBs are greatly reduced: along range, the radar provides better localization; along azimuth, the camera provides better localization. The experiment result will further approve of this association and EBs’ deduction.

2.2.6 EB Evaluation

From Fig. 2.6 and Definition 2.2.2, the EB of the next query is the region of the prediction and association. And vice versa. With RMSE, the EB evaluation can be estimated as:

\[
\text{RMSE}(\hat{o}) = \sqrt{\mathbb{E}((\hat{o} - o)^2)}
\]

\[
= \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\hat{o} - o)^2}, \tag{2.36}
\]
Algorithm 1 Pseudo-code of the tracking and association for multi-targets

Require: EKF tracks: tracks at current time stamp

Initialize: \( i = 0 \)

After radar-camer fusion-EKF, get number of tracks: \( N \)

while \( i < N \) do
  if Track valid then
    Find EBs for next step from prediction
    Find radar clusters and BBoxes within EBs
    if radarclusters not equal to 1 then
      radarInvisibleCount ++
    else
      Update radar’s localization
    end if
  end if
  if BBoxes not equal to 1 then
    camInvisibleCount ++
  else
    Update camera’s localization
  end if
  end if
  if radarInvisibleCount > threshold\text{radar} then
    Cross-validation failed.
    Track invalid, valid until camera updates
  end if
  if camInvisibleCount > threshold\text{cam} then
    Cross-validation failed.
    Track invalid, valid until radar updates
  end if
  \( i ++ \)
end while

Return: EBs with localization for each track
where \( o \) is the operator of the estimation variable.

With a window function to estimate the EB, the Equation (2.36) is modified with

\[
\text{RMSE}(\hat{o}) = \sqrt{\frac{1}{w} \sum_{n=k-(w-1)/2}^{k+(w-1)/2} (\hat{o} - o)^2},
\]

(2.37)

where \( w \) is the window size of the estimation on \( k \)-step. As introduced before, the EBS of radar are the radar range, angle and Doppler variances, \( \begin{pmatrix} \sigma^2_r & \sigma^2_\theta & \sigma^2_v \end{pmatrix} \). Here the sliding window applied RMSE is used to measure the EB on where the prediction region is. In this case, the EB evaluates radar’s fusion performance with camera’s vision.

In the experiment, the EB of radar is measured and evaluated with RMSEs in range, Doppler and angle domains, respectively.

2.3 Experimental Results

The experiment is implemented by two humans’ walking inside an indoor environment. The application implemented sensor fusion in this experiment conducts monocamera sensor from ELP USB8MP02G-L75 monocamera with \( 640 \times 480 \) YUY2 video picture encoding format [46] transmission at 30 frames per second (fps). And the mmWave radar sensor is from TI AWR1642 working at 77 GHz and 30 fps, which is a 1D 4 RX by 2 TX phased array antenna. AWR1642 is only capable of measuring target in range and azimuth plane.

The vision output and radar output from EKF can be seen from Fig. Figures 2.7 to 2.9. In the vision’s output, bounding boxes with positions and velocities on \( x \)-axis and \( y \)-axis are shown. As well as the radar’s point cloud clusters. Different colors stand for different clusters with searching radius of a human. The clusters provide median filtered centroids with green marker and black marker edge. Red and blue
Figure 2.7.: EKF output (1).
Figure 2.8.: EKF output (2).
(a) Frame number 150 radar EKF output
(b) Frame number 150 camera EKF output
(c) Frame number 160 radar EKF output
(d) Frame number 160 camera EKF output
(e) Frame number 170 radar EKF output
(f) Frame number 170 camera EKF output
(g) Frame number 180 radar EKF output
(h) Frame number 180 camera EKF output

Figure 2.9.: EKF output (3).
markers with black marker edge are the EKF output of the tracking. In the radar’s output, radar point clusters’ centroid, as well as fusion-EKF prediction localization for different tracks are presented. These figures are combinations of tracks thus showing multiple targets’ multiple sensors’ tracking results.

As we stated in the Section 2.2.4, we only have a 2D radar, i.e., the elevation information is ignored, and thus only points on the ground are shown. It is also applicable when applying elevation information to have tracking more reliable.

Additionally, in case of blocking from one target to another, the output bounding box is showing “radar only” in Fig. 2.8h and Fig. 2.9b. As invisible from the camera’s BSA. The “radar only” label still keeps tracking of different tracks until a camera’s vision validates the track. Otherwise, after certain invisible counts of the track, the track is deleted because of lack of cross-validation.

The radar’s tracking, shown in Figures 2.8e, 2.8g, 2.9a, 2.9c, 2.9e and 2.9g, provides multiple targets’ tracking. If without camera’s vision, the output may be messed up until the fusion EBs’ update. The region search and association avoids the miss assignment of tracks and improves tracking results.

To see how the performance of the radar-camera fusion-EKF system, the RMSE is used. And to see the EB variance of tracks, RMSE plots represents if the fusion-EKF takes advantages from both sensors. From Equation (2.37), the range, angle and Doppler RMSE plots for different tracks 1 and 2 are shown in Fig. 2.10, 2.11 and 2.12, respectively. The memory used for RMSE is 9, 15 and 25, which are fast, median and slow windows, respectively. With fps of 30 Hz, the window is estimating about 0.3 s, 0.5 s and 0.83 s, respectively.

As from Fig. 2.10a, after initializing the track, the RMSE drops rapidly to around 0.05 m. From Equation (2.1), the range resolution of the fusion system is 0.0436 m. And that is the theoretical range resolution. After the fusion-EKF, the range resolution achieves around 0.05 m after stabilizing from Fig. 2.10a. The weakness
(a) Radar’s range RMSE of EKF estimates with memory of 9, 15 and 25 measurements for track 1

(b) Radar’s range RMSE of EKF estimates with memory of 9, 15 and 25 measurements for track 2

Figure 2.10.: EKF RMSE of radar’s range.

of radar’s detecting cross range objects is overcome from Fig. 2.10b. As for radar’s cross range objects in a typical radar scan, the cross range velocities are around zero because Doppler velocities are range rates. So for radar-camera fusion-EKF, this problem is solved by adding the camera’s vision.
As from Fig. 2.11a, after initializing the track, the RMSE drops rapidly to around 0.015 rad, which is 0.859 degrees. From Equation (2.2), the angular resolution of radar at center FOV is 0.25 rad. And that is the theoretical angular resolution at center FOV. At outside that is $1 / \cos \theta_i$. After the fusion-EKF, the angular resolution achieves around 0.015 rad after stabilizing from Fig. 2.11a. This track is along about
Figure 2.12.: EKF RMSE of radar’s Doppler.

(a) Radar’s Doppler RMSE of EKF estimates with memory of 9, 15 and 25 measurements for track 1

(b) Radar’s Doppler RMSE of EKF estimates with memory of 9, 15 and 25 measurements for track 2

the center FOV so that it verifies that the camera improves fusion’s track significantly. In Fig. 2.11b, similar result is achieved. And confirms the camera’s improvements on radar’s EB on azimuth axis.

As from Fig. 2.12a, after initializing the track, the RMSE drops rapidly to around 0.3 m/s. From Equation (2.3), the Doppler resolution of radar is 0.616 m/s. And that
is the theoretical Doppler resolution. After the fusion-EKF, the Doppler resolution achieves around 0.3 m/s after stabilizing from Fig. 2.12a. The fusion-EKF improves Doppler estimations on both range rate and cross range rate. In Fig. 2.12b, only 0.5 m/s is achieved. But for radar’s cross range rate, it is typically measured zero. Fusion-EKF greatly improves radar’s cross range rate detections. And thus allows cross range objects reliability.

From Figures 2.10 to 2.12, two motion maneuvering models are handled with different tracks with different CA models from fusion-EKF. The result shows different CA models can handle tracking for different targets with different maneuvering behaviors.

Additional experiment analysis on this EB-based sensor fusion is shown in Fig. 2.13. This experiment is using sensor fusion result of radar and camera and compares the localization result with ground truth. The benchmark of radar-camera fusion-EKF tracking result respect to ground truth can be seen in Fig. 2.13. The RMSE of fusion localization deviates from ground truth is 0.3664 m. It is a good range to separate human targets. And it is a reasonable region search EB from Definitions 2.2.1 and 2.2.2. The average localization error comes from the track association from inhomogeneous sensors. Performance of proposed fusion method can be seen in Table 2.1. With EBs reduced in range, angle and Doppler dimensions, the radar-camera fusion-EKF provides a solution to inhomogeneous sensor fusion. Additionally, cumulative distribution of localization errors of range, azimuth angle and velocity are presented in Fig. 2.14.

Some comparison of errors or EBs between the proposed radar-camera fusion-EKF and other radar tracking algorithms are shown in Table 2.2. Because mmWave radar has better range resolution than traditional radars, range EB is superior than other lower frequency radars. Meanwhile, azimuth angle EB is improved greatly due to the
Figure 2.13.: Fusion-EKF benchmark on EB respect to ground truth.

Table 2.1: Average localization error of radar-camera fusion-EKF

<table>
<thead>
<tr>
<th>Average Localization Error</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>0.2902 m</td>
</tr>
<tr>
<td>Azimuth angle</td>
<td>0.0134 rad</td>
</tr>
<tr>
<td>Velocity</td>
<td>0.7864 m/s</td>
</tr>
</tbody>
</table>

fusion with camera. Therefore, the proposed fusion-EKF provides extra information from camera to improve the detection and tracking of mmWave radar sensor
Table 2.2: Comparison of error/EBs using radar-camera fusion EKF with other radar tracking algorithms

<table>
<thead>
<tr>
<th>Method</th>
<th>Error or EBs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA model of radar-camera fusion-EKF</td>
<td>0.2902 m, 0.0134 rad (Short range configuration with camera fusion)</td>
</tr>
<tr>
<td>CCA model from [23]</td>
<td>2.49 m (Euclidean distance without fusion)</td>
</tr>
<tr>
<td>600 m airport surveillance system from [47]</td>
<td>5 m, 0.15°</td>
</tr>
<tr>
<td>New track association algorithm from [48]</td>
<td>0.5°</td>
</tr>
</tbody>
</table>
Figure 2.14.: Cumulative distribution of localization errors.

(a) Cumulative distribution of localization error of range

(b) Cumulative distribution of localization error of azimuth angle

(c) Cumulative distribution of localization error of velocity
3. REAL-TIME HUMAN MOTION BEHAVIOR DETECTION VIA CNN USING MMWAVE RADAR

3.1 Background

Human motion behavior monitoring has been grasping great interest among researchers, especially in surveillance, tracking and patient monitoring. However, current methods are limited to using cameras and infrared sensors [49]. Though cameras, including thermal cameras, are accurate and reliable for surveillance usage, the disadvantages of cameras are obvious: they can leak privacy, and rely on light condition [50]. A more comprehensive and accurate sensor is needed for behavior monitoring. In this chapter, we are using micro-Doppler signatures resulted from mmWave radar to detect the motion behavior of people.

In recent years, small and low-cost single chip consumer radar systems operating at millimeter wave frequencies open up a vast range of new applications such as automotive radar, health monitoring radar [51], micro unmanned aerial vehicle radar [52], robot guidance radar and a host of other applications [53]. Automotive radar, especially working at 77 GHz, is an attractive research area which plays an important role in the automotive industry, autonomous cars and driving assistance systems [54]. It can capture the micro-Doppler signature to recognize target via dynamic time warping [55], study human kinetics [56] and track objects [57,58]. However, the studies on human motion behavior are bounded. In [59], Seifert provided human walking style using micro-Doppler. But how to classify the behavior is not studied. In [60], Shrestha gave a good support vector machine classification on the human motion behavior for indoor monitoring. To further classify behavior into various classes, neural network
is a promising method. Its classification result can be continuously improved as more samples are fed into the training process. However, there is only limited study using neural network for radar data [59, 60]. In this chapter, we use CNN to detect the behavior of people, and build a real-time monitoring system, which can provide a reliable classification of motion behaviors.

Micro-Doppler effect is induced by micro-motion dynamics of a target or its structure such as vibration, rotation, tumbling and coning motions, which is widely existed in bulk motion of a radar target, aka Doppler effect [4]. In this chapter, the mmWave radar provides prediction output of the target behavior using micro-Doppler effect. A convolutional neural network is investigated to train the radar system to recognize the behavior of people. In this case, we treat micro-Doppler signature as a data image with intensities. The convolution layers train the collected data into the output matrix and optimize the parameters in the network neurons. The query micro-Doppler signatures of the incoming testing dataset then can be predicted using this CNN. A probability matrix is produced from query micro-Doppler signatures.

In this chapter, we are the first to build a real-time human motion behavior detection system using the commercial mmWave radar. We developed the system using robotic operating system (ROS) to control low-level device (mmWave radar), and transmit streaming data across platforms. We also investigated how CNN can be used for detecting human motion behaviors. Meanwhile, behaviors can be defined by users, and the system can be trained offline. Consequently, the proposed system can detect human motion behavior accurately in real-time. The proposed system can have a wider application, such as autonomous driving, traffic monitoring and patient cares.

The structure of this chapter is as follows. In Section 3.2, the methodology of extracting micro-Doppler information, designing of CNN and ROS framework are introduced. In Section 3.3, training and experimental results are presented.
3.2 Methodology of Real-Time MDS Observation

3.2.1 Real-Time FMCW MDS Processing

For a mmWave radar not considering to resolve angles, the received signal processing can be classified into two different scenarios: without constant false alarm rate (CFAR) detection and with CFAR detection. 

The first scenario (without CFAR case) is to use the entire range-Doppler data. The received data cube (i.e. a series of range-Doppler data) is passed through Doppler-time extraction. Accumulated real-time data is then went through a CNN prediction node. An example can be seen in Fig. 3.1, the mmWave radar obtains the micro-Doppler signature of a human walking by raw sampling on board and processing through a host computer. Arm, leg and torso can be clearly recognized in the entire Doppler data, though the clutter is not totally removed. In the proposed system, we pre-defined the number of chirps, and pulse repetition interval to guarantee the Doppler resolution is sufficient in detecting human motion behaviors.

The second scenario (with CFAR case) is to use the point cloud data, i.e. the raw range-Doppler data are processed by the integrated CFAR algorithm on the radar board, which only picks the radar detection points and produces point cloud data with Doppler information. The point cloud data has a much less data volume for data transmission, and can be used for the real-time application.

In the remaining of the chapter, we use the point could data, and transmit the data to the host computer. The host computer can process the point cloud data using grouping and clustering algorithms and form the micro-Doppler signature data. With a fixed time frame, the micro-Doppler data are passed through a trained network. In this way, the human motion behavior can be detected in real-time.

A series of sequential range-Doppler frames are used to analyze micro-Doppler signatures. The frames are buffered and stored on host computer as input to the CNN
network. The buffered data are updated by incoming data and processed through the network. It makes the real-time motion behavior detection and classification feasible.

3.2.2 CNN

CNN is a class of deep and feed-forward artificial neural networks with multi-layer perceptrons. Conventionally, a CNN is used as a common classifier for images or visually imagery because of its convolution process emulates the response of an individual neuron to visual stimuli [61, 62]. In this chapter, the CNN is used on un-visualized micro-Doppler signatures. Since the micro-Doppler signatures is time-variant, the input layer neurons for CNN are depending on time. The Doppler effects from multi-targets are clustered using density-based spatial clustering of applications with noise (DBSCAN) algorithm or other clustering algorithms before the micro-Doppler signature classification.
Fig. 3.2 shows the structure of the CNN network from the radar raw range-Doppler response. The summarization of the CNN using micro-Doppler from Fig. 3.2 is shown in Table 3.1.

In Fig. 3.2 and Table 3.1, we shows a typical frame of convolution layer, which contains:

1. Convolution layer It performs convolution calculation based on the micro-Doppler data. The convolution coefficients are trained to calculate features.

2. Activation layer (Leaky ReLU function) Leaky ReLU is the abbreviation of Leaky version of a Rectified Linear Units. It is widely used in neural network as the activation function. It overcomes the dying ReLU issue, and is a non-saturating activation function:
\[ f(x) = \begin{cases} 
\alpha x, & x < 0 \\
 x, & x \geq 0 
\end{cases} \quad (3.1) \]

where \( \alpha \) is a very small number (typically equals 0.01) to avoid negative values not mapping appropriately.

3. Max pooling layer The Spatial pooling reduces the dimensionality of each feature map but retains the most important information. The objective of subsampling is to get an input representation by reducing its dimensions, which helps in reducing overfitting.

4. Dropout layer The dropout layer is added to avoid overfitting.

5. Fully connected layer (with flatten and dense layers) Several convolution layers are then merged to a fully connected layer. The goal of the fully connected layer is to flatten the high-level features that are learned by convolutional layers. It combines all the features from the original input.

As the CNN network is setup as Fig. 3.2. We collected the training data via performing different types of human motion behaviors, such as walking, swinging hands, standing/sitting and shifting. The training data is stored, and the training processing is performed in a high performance computing (HPC) unit with on-board graphics processing units (GPU). As coefficients are generated through the training data, the neural network model is stored as an h5py file, and uploaded to the computer for implementing the detection and classification in real-time.

3.2.3 ROS Framework

The implementation of the real-time micro-Doppler behavior monitoring is done on ROS framework. ROS provides hardware and low-level device control, message
Table 3.1: CNN summarize (Total parameters: 1,141,763; Trainable parameters: 1,141,763; Non-trainable parameters: 0)

<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output Shape</th>
<th>Parameter #</th>
</tr>
</thead>
<tbody>
<tr>
<td>conv2d_1 (Conv2D)</td>
<td>(None, 128, 25, 32)</td>
<td>320</td>
</tr>
<tr>
<td>leaky_relu_1 (LeakyReLU)</td>
<td>(None, 128, 25, 32)</td>
<td>0</td>
</tr>
<tr>
<td>max_pooling2d_1 (MaxPooling)</td>
<td>(None, 64, 13, 32)</td>
<td>0</td>
</tr>
<tr>
<td>dropout_1 (Dropout)</td>
<td>(None, 64, 13, 32)</td>
<td>0</td>
</tr>
<tr>
<td>conv2d_2 (Conv2D)</td>
<td>(None, 64, 13, 64)</td>
<td>18496</td>
</tr>
<tr>
<td>leaky_relu_2 (LeakyReLU)</td>
<td>(None, 64, 13, 64)</td>
<td>0</td>
</tr>
<tr>
<td>max_pooling2d_2 (MaxPooling)</td>
<td>(None, 32, 7, 64)</td>
<td>0</td>
</tr>
<tr>
<td>dropout_2 (Dropout)</td>
<td>(None, 32, 7, 64)</td>
<td>0</td>
</tr>
<tr>
<td>conv2d_3 (Conv2D)</td>
<td>(None, 32, 7, 128)</td>
<td>73856</td>
</tr>
<tr>
<td>leaky_relu_3 (LeakyReLU)</td>
<td>(None, 32, 7, 128)</td>
<td>0</td>
</tr>
<tr>
<td>max_pooling2d_3 (MaxPooling)</td>
<td>(None, 16, 4, 128)</td>
<td>0</td>
</tr>
<tr>
<td>dropout_3 (Dropout)</td>
<td>(None, 16, 4, 128)</td>
<td>0</td>
</tr>
<tr>
<td>flatten_1 ( Flatten)</td>
<td>(None, 8192)</td>
<td>0</td>
</tr>
<tr>
<td>dense_1 (Dense)</td>
<td>(None, 128)</td>
<td>1048704</td>
</tr>
<tr>
<td>leaky_relu_4 (LeakyReLU)</td>
<td>(None, 128)</td>
<td>0</td>
</tr>
<tr>
<td>dropout_4 (Dropout)</td>
<td>(None, 128)</td>
<td>0</td>
</tr>
<tr>
<td>dense_2 (Dense)</td>
<td>(None, 3)</td>
<td>387</td>
</tr>
</tbody>
</table>

passing and source management. For the proposed system, we implemented low-level device control of the mmWave sensor of Texas Instruments (TI) AWR1642 on ROS, and transmitted data via ROS messages. In this way, the real-time monitoring can be achieved. The workflow can be seen in Fig. 3.3. The radar data is transmitted into the host computer for behavior detection. It passes through data transmission via ROS node, and generates micro-Doppler frames in micro-Doppler signature node as input for CNN network. The CNN network is implemented in human motion behavior detection node. The neural network model is stored as an h5py file and transmitted
from HPC to host computer before launching the human motion behavior detection node.

The TI mmWave radar is operating at the frequency of 77 GHz with 25 Hz framerate. For indoor application, we configure the radar parameters to a maximum range of 17.788 m. The Doppler resolution is 0.097 m/s, and the Micro-Doppler data is at a frame rate of 25 Hz.
3.3 Results

3.3.1 Training Results

The training data obtained from testing researchers which consist hundreds of human motion behavior samples such as walking, waving hands, clapping, sitting/standing, bending, falling, etc. In this study, behaviors like walking, swinging hands for help and falling are our interest. Different behaviors are recorded and stored in our database. The database is then trained with the proposed CNN on HPC. The training time on HPC costs around 37 minutes. For the first ten epochs of CNN, the training and validation accuracies and losses are shown in Fig. 4. From the Fig. 4, validation loss and validation accuracy both are in sync with training loss and training accuracy. The model is not overfitting from the reason that validation loss is decreasing and the gap between training and validation accuracies are relatively small. Unlike the accuracy, loss is not a percentage. The training loss is a summation of the errors made for each example in training or validation sets. Loss value implies how well or poorly a certain model behaves after each iteration of optimization. Ideally, one would expect the reduction of loss after each, or several, iteration(s). The accuracy of a model is usually determined after the model parameters are learned and fixed and no more learning is taking place, i.e. loss cannot further decrease. In this study, the outcome produces an accuracy above 99.5% and loss less than 0.02. Therefore, after the model is converged, the proposed system can produce a good human motion behaviors detection result.

3.3.2 Testing Results

The testing is conducted in the lab in the Department of Electrical and Computer Engineering of the University of Arizona. The human motion behavior testing is conducted by a researcher who is trying to move into the detection area and perform
Table 3.2: CNN prediction accuracies (Collected multiple test samples and averaging results among samples)

<table>
<thead>
<tr>
<th>Human Motion Behavior</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human walking and vanish from radar</td>
<td>96.32%</td>
</tr>
<tr>
<td>Human waving hands when standing or sitting</td>
<td>99.59%</td>
</tr>
<tr>
<td>Human sitting to standing and walking transition</td>
<td>64%</td>
</tr>
<tr>
<td>Human walking back and forth</td>
<td>91.18%</td>
</tr>
<tr>
<td>No micro-Doppler detections</td>
<td>97.84%</td>
</tr>
<tr>
<td>Complex detections including all behaviors</td>
<td>95.19%</td>
</tr>
</tbody>
</table>

different behaviors like walking and waving hands. The mmWave radar is set at the height of 80 centimeters and pointing to the room area. The system running ROS human motion detection node is in real-time, and the experiment shows the real-time motion behavior detection of people. The actual test is shown in Fig. 5. In Fig. 5(a), the camera vision is provided for reference. For this project, the CNN is only applied on micro-Doppler signatures. And at the testing stage, vision analysis can provide a good reference. In Fig. 5(b), the detected human with Doppler information provided. Bulk motion of body is detected as well as the micro-motions from arms and legs. Different part of body are clearly detected with appropriate configuration on chirps and CFAR.

The testing results of CNN prediction on human motion behavior monitoring using mmWave radar can be seen in Table 3.2. Each accuracy is calculated from the prediction from node and the actual behavior labelled from host. The testing results in Table 3.2 shows a very precise prediction over different continuous behaviors. The low accuracy may be due to the insufficient data in training for the transition behaviors, such as standing and walking. These behaviors happen in a short amount of time, and we are training using only 1896 samples, instead of 9973 samples for other behaviors. For other behaviors, the accuracy is relatively high. As micro-Doppler signatures are
far more than human motion behaviors like walking and waving hands, the proposed system can be extended to include other behaviors.
4. ROBUST AND ADAPTIVE RADAR ELLIPTICAL DENSITY-BASED SPATIAL CLUSTERING AND LABELLING FOR MMWAVE RADAR POINT CLOUD DATA

4.1 Background

Modern commercial mmWave radars are compact in size, fast in processing speed but have limitations of transmitting their data through serial port or controller area network interface. Thus point cloud is widely used in modern frequency-modulated continuous-wave (FMCW) MIMO radars. Clustering and labelling problems for detecting mmWave targets have raised concerns for researchers, especially in the field of automobile, patient monitoring and intelligent traffic systems.

Radar point cloud output is typically sparse and not easy to visualize for detecting targets. A joint clustering and tracking algorithm is needed for detecting targets. Conventional density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm has advantages like it can resist noise points (outliers) effectively. And it can handle clusters of different shapes and sizes within a fixed initial core distance. However, the variant densities could result in difficulties to cluster. For instance, a variant sparse object or a variant dense object can cause false clustering. In addition, original DBSCAN is sensitive to parameters. In this case, the detecting and clustering scenarios are affecting the result of DBSCAN when applying to mmWave radar point cloud data. Because radar detects targets in polar (or sphere) coordinates, the detection of targets is dense in close range field and sparse in long range field.
Target detections are variant across radar point cloud map. Thus original DBSCAN clustering algorithm is limited.

Many clustering algorithms have been introduced in recent years. In [63], Mai et al. introduced an anisotropic density-based clustering algorithm. And they compared the algorithm with original DBSCAN and ordering points to identify the clustering structure (OPTICS) algorithm. With this method, the efficiency and quality improved significantly. In [64], Smith et al. provided the elliptical clustering algorithm. With weighted density function, classification and post-processing included, the unconstrained elliptical method had improvements and comparable performance over standard DBSCAN method. Some other $K$-means algorithms are introduced in [65,66]. Some modified DBSCAN algorithms for radar point cloud data are introduced in [67,68]. Many researches are currently using or used clustering algorithm for sensors’ fusion [69,70] and SAR [71].

In this chapter, we are proposing a new adaptive and robust clustering and labelling algorithm for radar point cloud based on FMCW radar’s parameters. The algorithm is derive from DBSCAN, but with elliptical shape from radar parameters. We name this algorithm as REDBSCAN. The elliptical shape is formed from mmWave radars range resolution and angle resolution. As range resolution is even across the detecting map but angle resolution in distance is dense at close range and sparse at long range, the elliptical shape is changing according to the position. Additionally, region query on radar points is more flexible. In this case, some large targets with shape beyond the detecting region is able to be formed and labelled correctly. It also improves edge detections of adjacent targets for mmWave radars. Small adjacent targets can be clearly distinguished and found throughout our algorithms.

The structure of this chapter is as follows. In Section 4.2, the brief radar signal processing, density function of point cloud, GMM, DBSCAN and REDBSCAN are
introduced. In Section 4.3, results of some radar detected point cloud clustering are presented.

4.2 Methodology

4.2.1 FMCW MIMO Radar Signal Processing

Modern compact patch antennas are widely used in MIMO radars. In case of FMCW MIMO radar which typically working at 77 GHz for automobiles, the signal processing chain is shown in Fig. 4.1.

In Fig. 4.1, radar signal from different transmitter is:

$$S_t(t) = \sqrt{E} \exp[j2\pi f(t)t]$$  \hspace{1cm} (4.1)
where \( E \) is the electric field amplitude and \( f(t) \) is the frequency modulated signal:

\[
f(t) = f_c + k_r t \tag{4.2}
\]

where \( f_c \) is the center frequency and \( k_r \) is the frequency slope of FMCW radar. Receiver antenna signals are processed through range compression, Doppler compression using stretch processing with mixture of transmitting signal:

\[
S_r(t - \Delta t) = \sqrt{E} \sigma \exp[j2\pi(f_c + f_d)(t - \Delta t) + \pi k_r(t - \Delta t)^2] \tag{4.3}
\]

where \( f_d \) is the Doppler frequency of the scatter and \( \Delta t \) is the time delay which can be used to obtain target’s Doppler and range information. FMCW radar parameters are added to index and for clustering. Optional moving target indication (MTI) filter is added to remove the static clutters. Different scatters are produced via 2D cell averaging CFAR (CA-CFAR) detections on range-Doppler responses.

MIMO radar array can produce a virtual antenna array. The peak values on different receiver antennas have phase differences caused by time delay \( \tau \) from scatter angle \( \theta \). Assume first virtual antenna is the origin and the \( m \)th virtual antenna has the phase difference of:

\[
u_m(t - \tau) = S_r(t - \tau) \exp(-j2\pi f_c m \tau) \exp(j2\pi ft) \tag{4.4}\]

Thus the array factor (AF) can be expressed as:

\[
AF = \sum_{m=1}^{M} \exp(-j2\pi m \frac{m d}{\lambda} \sin \theta) \tag{4.5}\]

The digital beamforming is done after 2D CA-CFAR on range-Doppler response.
With range, Doppler, angle and amplitude solved, the detection table can be produced for clustering. In this chapter, we are comparing REDBSCAN clustering algorithm with GMM and DBSCAN clustering for mmWave radar applications. In this case, a better clustering point cloud algorithm can lead better target detection and point cloud visualization.

### 4.2.2 Density Function

The REDBSCAN algorithm is based on representing different radar detecting points (from reflective scatters) as dense region in detecting space, namely, the range-azimuth map. The REDBSCAN is judging different clusters by their densities, radar parameters and approximate shapes.

Here we adopt a hyperbolic tangent kernel function and represent the overall density of the dataset. The density estimation function, denoted by $\rho_w(x)$, is defined as

$$\rho_w(x) = \frac{1}{\sum_{n=1}^{N} w_n} \sum_{n=1}^{N} w_n (1 - \tanh \|x_n - x\|^2)$$

(4.6)

where $x$ is a $D$-dimensional vector and $w$ is a $N$-dimensional weight vector. Each $w_n$ is associated with observation $x_n$. The function $\rho_w(x)$ takes the high concentration of observations of the dataset.

For a typical radar point cloud after FMCW MIMO processing chain, the weighted density functions of the point cloud points can be seen in Fig. 4.2. In Fig. 4.2a, the sample radar point cloud map is shown. In Fig. 4.2b and Fig. 4.2c, cross range densities and along range densities are presented, respectively. Clustering algorithms are representing the presence of subpopulations within an overall population by trying to solve the clusters from radar point cloud densities accordingly.
4.2.3 GMM

Gaussian mixture model (GMM) is widely used in clustering density points. For GMM in 2D space, we have to first define $K$-clusters. These $K$-clusters component weights $\phi$ yield:

$$\sum_{i=1}^{K} \phi_i = 1 \quad \quad (4.7)$$
From Equation 4.7, the component weights are constrained so that the total probability of a distribution equals 1. The probabilities of all the clusters defer Gaussian distribution:

\[ p(x) = \sum_{i=1}^{K} \phi_i \mathcal{N}(x|\mu_i, \sigma_i) \]  

(4.8)

The \( \mathcal{N}(x|\mu_i, \sigma_i) \) is the 2D Gaussian distribution which is defined as:

\[ \mathcal{N}(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left( -\frac{(x - \mu_i)^2}{2\sigma_i^2} \right) \]  

(4.9)

As we see from Equations (4.7) to (4.9), all the parameters of a Gaussian distribution yield to \( K \)-clusters. Thus the pre-defined \( K \) is critical in calculating the Gaussian distribution. However, in a radar scan, we do not know exactly how many clusters the detecting map have. Thus GMM in radar scan is limited.

In addition, normal radar points have noisy scatters from targets. But GMM always have clusters cover overall detecting map. In this case, GMM cannot handle outliers from radar points. This is also validated in Section 5.3.

### 4.2.4 DBSCAN

The conventional DBSCAN clustering is defined as follows:

1. A query point \( p \) is presented, and a point \( q \) is waiting for clustering.
2. Euclidean distance of \( p \) and \( q \) is calculated and compared to \( \epsilon \).
3. A minimum points in cluster is verified.

Thus the \( \epsilon \)-neighbors is defined as:

\[ \mathcal{N}_{\epsilon}(p) : \{ q | d(p, q) \leq \epsilon \} \]  

(4.10)

where \( d(p, q) \) is the Euclidean distance between \( p \) and \( q \).
With given $\epsilon$ and minimum points, the point cloud can be categorized into three exclusive groups:

1. Core points: the interior of a cluster with more than minimum points within $\epsilon$.

2. Border points: the neighborhood of one or more core points.

3. Outliers: noisy points that are not within any clusters.

The advantages of DBSCAN are obvious: first, it can fit radar point cloud data and it can deal with noise; second, it can handle different clusters with different shapes and sizes. However, DBSCAN is still not ideal for radar point cloud clustering as it cannot handle varying densities. And it is sensitive to parameters which is based on detecting scenarios. For instance, $\epsilon$ needs to be defined to 2 meters which is similar to a car size when it applies to traffic monitoring. But $\epsilon$ needs to be set to 0.5 meters which is approximate a human body size when it applies to indoor patient monitoring. Thus DBSCAN also has limitations apply to radar clustering problems.

### 4.2.5 REDBSCAN

Compared to circle shaped clustering algorithms, elliptical shaped clustering has many advantages. It can handle densities better. As well as it can fit radar detection polar coordinates better. In previous subsection 4.2.2, we introduced REDBSCAN and its density function. In detail, it has the following steps:

1. Detecting a new cluster and give it a label.

2. Expand the existing cluster based on its minimum points.

3. Iteration over detecting and expanding until enough points are updated.
According to these steps, we compute the gradient of Equation (4.6) with respect to \( x \). The locations of the maxima obey an auto-coherent equation iteratively. Given an estimation \( x^c \) of the location of the maxima of \( \rho_w \) computed at the current iteration \( c \) of the maximization process, the updated estimation at the next iteration \( c+1 \) is given by

\[
x^{c+1} = \frac{\sum_{n=1}^{N_w} w_n C_n x^c x^n}{\sum_{n=1}^{N_w} w_n C_n x^n}
\]

where

\[
C_n = 1 - (\tanh \|x_n - x\|^2)^2
\]

Since density function occurs close to the data points, the initial estimation \( x^0 \) can be chosen randomly. The initial ellipsoid is defined according to the radar parameters range resolution, \( dr \), and angle resolution, \( da \). Range resolution is calculated as

\[
dr = \frac{c_0}{2B}
\]

where \( c_0 \) is electromagnetic wave speed and \( B \) is the bandwidth from FMCW radar. Angle resolution is defined as

\[
da = \frac{\theta_{FOV}}{N_A}
\]

where \( \theta_{FOV} \) is the field of view (FOV) of the antenna (assumed a 2D antenna without elevation information) and \( N_A \) is the number of virtual antennas emulated by MIMO.

Finally, the cluster \( \mathcal{C} \) can be expressly by the set of data points inside the generated ellipsoid as

\[
\mathcal{C} = \{ x_n \in \mathcal{D} | (x_n - X^*)^T P^T A^{-2} P (x_n - x^*) \leq 1 \}
\]

where \( P = (u_1, u_2, ..., u_D) \) is the orthogonal basis induced by orientation of the ellipsoid and \( A \) is the diagonal matrix \( A_{ij} = a_i \delta_{ij} \) with \( \delta_{ij} \) the Kronecker delta.

In this case, we build a clustering algorithm using Equations (4.6) and (4.11) to (4.15). The main algorithm of REDBSCAN is shown in Algorithm 2. The in-
cluded functions of query region and expanding clusters are shown in Algorithm 3 and Algorithm 4.

**Algorithm 2** Pseudo-code of the REDBSCAN

```plaintext
function REDBSCAN = Xn(x, y), p, dr, da
c = 0, idx = 0, n = shape(X)
Dnn = ||X||_pairwise
Nn = bool matrix(n, 1)
while i < n do
    if X_i(x, y) not visited then
        Find neighbors using RegionQuery(i)
        if neighbors < p then
            N_i = True
        else
            c = c + 1
            ExpandCluster(i, neighbors, c)
        end if
    end if
end while
```

**Algorithm 3** Pseudo-code of the neighbors = RegionQuery(i)

```plaintext
function neighbors = RegionQuery(i)
Set up ellipse using radar parameters dr, da
Count numbers using cnt
e = \sqrt{1 - b^2/a^2}
while j < n do
    Calculations on Equation (4.15)
    Find \phi using Equation (4.15)
    \epsilon = \frac{a(1-e^2)}{a(1-e^2)}(\phi-\psi)
    if D(i, j) \leq \epsilon then
        neighbor(j) = True
        cnt = cnt + 1
    end if
end while
```

Following the Algorithms 2 to 4, the radar point cloud can be ideally clustered for pedestrian, cars and other objects using radar parameters from FMCW settings. It is adaptive with variant radar parameters. In Section 5.3, we will compare the REDB-
Algorithm 4 Pseudo-code of the ExpandCluster(i, Neighbors, c)

function ExpandCluster(i, Neighbors, c)

\[ idx(i) = c, k = 1 \]

while True do

\[ j = \text{neighbors}(k) \]

if not visited(j) then

\[ \text{visited}(j) = 1 \]

if RegionQuery(j) \( \geq \) minimum_points then

\[ \text{neighbors}.append(\text{RegionQuery}(j)) \]

end if

end if

if \( idx(j) == 0 \) then

\[ idx(j) = c \]

end if

\[ k = k + 1 \]

if \( k > \text{length(neighbors)} \) then

break

end if

end while
Table 4.1: FMCW MIMO radar parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_s$</td>
<td>Sampling frequency</td>
<td>5.209 MHz</td>
</tr>
<tr>
<td>$f_c$</td>
<td>Center frequency</td>
<td>79.21 GHz</td>
</tr>
<tr>
<td>$B$</td>
<td>Bandwidth</td>
<td>3440.2 MHz</td>
</tr>
<tr>
<td>$PRI$</td>
<td>PRI</td>
<td>96.14 µs</td>
</tr>
<tr>
<td>$T_{Fr}$</td>
<td>Frame time</td>
<td>33.333 ms</td>
</tr>
<tr>
<td>$N_R$</td>
<td>Number of range bins</td>
<td>256</td>
</tr>
<tr>
<td>$N_D$</td>
<td>Number of chirps</td>
<td>16</td>
</tr>
<tr>
<td>$N_A$</td>
<td>Number of virtual antennas</td>
<td>8</td>
</tr>
<tr>
<td>$\theta_{FOV}$</td>
<td>FOV angle</td>
<td>±60°</td>
</tr>
</tbody>
</table>

Table 4.2: FMCW MIMO radar output

<table>
<thead>
<tr>
<th>Output</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum unambiguous range</td>
<td>11.154 m</td>
</tr>
<tr>
<td>Range resolution</td>
<td>0.044 m</td>
</tr>
<tr>
<td>Maximum velocity</td>
<td>±4.921 m/s</td>
</tr>
<tr>
<td>Velocity resolution</td>
<td>0.615 m/s</td>
</tr>
<tr>
<td>Angle resolution</td>
<td>15°</td>
</tr>
</tbody>
</table>

SCAN with GMM and DBSCAN and show why REDBSCAN is a better algorithm in radar point cloud clustering.

4.3 Results

The following results of REDBSCAN is compared with GMM and conventional DBSCAN algorithms. The experimental radar setup is shown in Table 4.1. The resulting radar output of range, Doppler and angle resolutions are shown in Table 4.2.
4.3.1 Comparison with GMM Clustering

The collected radar point cloud, GMM clusters, DBSCAN clusters and REDBSCAN clusters are shown in Fig. 4.3. From Fig. 4.3a, the collected radar data is processed through FMCW MIMO processing and output as a point cloud matrix. Conversion from radar polar coordinates to Cartesian coordinates is applied. Only x-y data are shown. Doppler and amplitude is not included in the clustering comparison.

As we can see from Fig. 4.3b, 4.3c and 4.3d, the GMM, DBSCAN and REDBSCAN provide different clusters for radar point cloud data. We preset the number
of clusters \( K = 14 \). GMM provides no outliers and the clusters are different in sizes. Some clusters span around 5 meters. Some clusters are not clear compared to DBSCAN and REDBSCAN. Running again with same data on GMM, the result will change because of the optimization and iteration. As we discussed in Section 4.2.3, GMM clusters need input of \( K \)-clusters or estimate \( K \) before clustering. When the detecting scenario is unknown, GMM may not work very well on radar point cloud clusters. In addition, the GMM cannot handle noise (outliers) which make it very unsuitable when applied to radar clustering problems.

4.3.2 Comparison with Conventional DBSCAN Clustering

Because REDBSCAN is originally developed from DBSCAN and has some advantages when applying to small or large clusters, we used some corner reflectors (CRs) as small near targets and wall, car and wardrobe for large targets. We are comparing some detailed radar cluster data as following.

In Fig. 4.4, the collected radar point cloud, DBSCAN clusters and REDBSCAN clusters are shown. We put two separate CRs close enough in 1 meters’ region and obtains different clusters in DBSCAN and REDBSCAN clusters. For REDBSCAN, when applying radar’s range resolution and angle resolution, the close field targets have small elliptical region compared to fixed sized DBSCAN. It has better boundary distinction over DBSCAN.

In Fig. 4.5, wall target clusters are shown. As radar has limitations of angle resolution, walls are separated with two clusters in this detection. DBSCAN provides its conventional way to cluster as two targets. But REDBSCAN recognizes the line-shaped target with radar’s virtual antenna parameters and get a cluster with a long wall-like shape. It works well with even more detection points on wall and in long range region.
In Fig. 4.6, a car target at range around 6 meters is clustered by both algorithms. As a single car have rectangle reflective surfaces around the body, the conventional DBSCAN can only show cluster of the car body as a point with preset $\epsilon$. However, REDBSCAN provides a superior cluster result that shows left and right side of the car body with its elliptical regions. In this case, the large object is clustered effectively.

In Fig. 4.7, multiple complicated target detection scenario is shown. With DBSCAN and REDBSCAN providing different clusters for radar point cloud, recognitions for small and large targets are greatly different. DBSCAN cannot handle different detecting targets with the same parameters as before. But REDBSCAN can cluster the point cloud without difficulties. In this case, REDBSCAN can work in variant detecting scenarios. It just needs to be fed with radar parameters. However, conventional DBSCAN should have known the approximate cluster size $\epsilon$ before clustering. That means the DBSCAN algorithm needs to know the detecting scenario before it starts. Therefore, the REDBSCAN algorithm is robust and adaptive.
Figure 4.5.: Detecting long range wall targets.

Figure 4.6.: Detecting large car targets.

Figure 4.7.: Detecting multiple complicated targets.
5. 3D IMAGING MILLIMETER WAVE CIRCULAR SYNTHETIC APERTURE RADAR

5.1 Background

3D imaging radar becomes feasible in the field of remote sensing due to the advancement in solid state microwave circuits and the digital signal processor [72]. However, large antenna synthesisization is usually performed by airborne radar to interrogate the terrain [73] and is not suitable for civilian usage. Considering a wide usage of millimeter wave radar for automobile, security and surveillance, we are proposing a new imaging technique [74, 75] combining synthetic aperture radar (SAR) and millimeter wave radar. The proposed millimeter wave circular synthetic aperture radar (MMWCSAR) conducts a circular trajectory to synthesize a large aperture for achieving 3D imaging. MMWCSAR has four key aspects: high resolution, working on diversity conditions, portable and low-cost.

For high resolution, millimeter wave radar can support hundreds of megahertz for range detection. For example, automotive radar is working on millimeter wave. It can sweep hundreds of megahertz to distinguish cars and even pedestrians [76]. In addition, automotive radar has been applied on smart vehicles for 2D terrain mapping [77], collision detection, parking assistance and the blind spot indicator [78]. However, the current automotive radar uses traditional strip-map SAR imaging to generate a 2D terrain mapping. The proposed MMWCSAR aims at generating 3D imaging. It can be further applied to automotive radar for better understanding of the environment.
For working on diversity conditions, we compare millimeter wave radar with optical technologies. Despite that optical imaging techniques have better resolutions due to the high center frequency, radar imaging has numerous advantages over traditional optical imaging techniques like camera and LiDAR. It is underlined in [79] that radar has superior working capacity in any weather condition, including rain, snow and fog. The complex roadway environment for automobiles requires uninterrupted remote sensors performing consistently in inclement weather. The millimeter wave radar is capable of acquiring and tracking all obstacles in its field of view (FOV) under all weather conditions [79]. MMWCSAR is an innovative imaging device capable of working in diverse conditions compared to traditional optical sensors.

For portable and low-cost features, we compare with traditional millimeter wave imaging techniques, such as security surveillance system [80, 81], concealed weapon detection [82–84] and millimeter wave cameras [77, 85]. Most of these traditional applications, including the millimeter wave scanner, which deploys many transceivers, are bulky and high-cost system designs. On the contrary, MMWCSAR adopts a monostatic radar, which can minimize its size and cost. It ensures that MMWCSAR can provide three-dimensional microwave imaging for security, indoor surveillance and automotive target detection in a wearable and inexpensive manner.

MMWCSAR requires the motion of radar in a circular track other than a straight track as traditional strip-map SAR techniques. Our study found that the movement between radar and targets produces the projections of objects to different movement directions. One can therefore obtain a range projection angle datacube from the circular movements. From the datacube, by applying inverse Radon transform (IRT) [86, 87], the datacube can be converted to the FOV figure. The principles are similar to computed tomography [86, 88–91]. Jia et al. [92] presented a 2D imaging algorithm for circular SAR. The motion of the SAR is along a circular trajectory. The sub-spectrums of different angles are matched filtered and summed in the Fourier do-
main to obtain a 2D Fourier spectrum of the imaging. By space-invariant matched filtering and 2D inverse Fourier transform (2D-IFFT), the trajectory deviation is eliminated, and the final aperture view is presented. Additionally, Bao et al. [93] proposed a multi-circular SAR approach. Their multi-circular SAR samples on different elevation levels. Different circular rotation angles and down-looking angles are recorded according to the range data. Consequently, their signal processing procedure results in reconstruction of a 3D image from multi-circular SAR. In our approach, we can use hand swinging to obtain a full-circular projection. The circular trajectory plane is parallel to the range bin plane. The range calibration is also implemented before the IRT in order to reduce target mismatch on the volume FOV image.

Traditional spotlight SAR works on S-band or X-band. It has a low carrier frequency and uses the range information to perform inverse radon transform. In the chapter, we are investigating the W-band 3D imaging technique. We found that as the carrier frequency increases to the W-band, the synthetic aperture becomes much smaller and allows the radar to use Doppler angle data to perform inverse Radon transform. The advantages of the proposed system are multifold: First, the high carrier frequency makes the radar very small. Second, the lesser number of transceiver elements can lower the cost of the radar. Third, further study found that a high resolution result can be obtained with less samples if compressed sensing (CS) is applied.

MMWCSAR tested with hand swinging is not accurate for high resolution imaging. In order to improve the sensing results, sensing time and improve the rotating accuracy, additional signal processing procedure of CS can be applied in our MMWCSAR system. CS has been introduced in [94,95]. Many large data size applications behaving as sparse targets have been using CS for data reduction and restoration, e.g., single-pixel imaging [96] and magnetic resonance imaging (MRI) [97,98]. It has advantages for situations when sampling is expensive, slow or difficult [99]. The
CS method applied in radar signal processing can sample fewer signals for the sensors, while keeping the same quality of the generated imaging. With reducing the size of samples for the portable radar device like MMWCSAR, a fast data acquisition can be achieved by the CS method. Meanwhile, the sensing results can be improved.

Radar signal processing with CS is addressed in [73, 92, 93, 100–103]. In [100], Ender gave a full analysis on applying the CS to radar pulse compression. Sevimli [101] introduced range-Doppler compressed sensing and optimization comparison of different reconstruction algorithms. In addition to the CS used in range-Doppler response, our method of applying CS to radar signal processing is innovative. It allows not only the pulse-Doppler compressed sensing, but also the slow-time-angle compressed sensing. Recent work on CS applied to SAR systems is drawing researchers’ attention. For instance, Bao et al. [93] produced the 3D multi-circular SAR image using a “2D + 1D” mode, i.e., 2D focused SAR images are followed by 1D profile estimation of the elevation direction. With CS applied to the 2D ground plane image and 1D profile of the elevation dimension, the 3D figure can be reproduced. CS theory is also applicable in our MMWCSAR system. First, CS is applied to 2D slow-time-angle data to form a 2D FOV image on a single range bin. The volume FOV figure can be reconstructed and analyzed by applying the 1D range profile to 2D FOV images of different range bins. Second, to achieve the CS in radar signal processing, the sparsity and incoherence properties are discussed. In addition, 2D transformation from the slow-time-angle to the azimuth-elevation representation matrix is discussed. Besides, we also introduce how to choose the sensing matrix, so that the CS algorithm can be realized. Finally, to further improve the performance of the MMWCSAR system, we focus on decreasing the data acquisition time, improving imaging results and reducing errors caused by humans with the CS algorithm applied in the experiment.

The structure of this chapter is as follows. In Section 5.2, the MMWCSAR system configuration, parameters, data acquisition, resolution and its constraints of choosing
MMWCSAR system parameters are introduced. In Section 5.3, range calibration and radar imaging processing by using IRT to reconstruct the volume FOV image are presented. In Section 5.4, we propose the CS for MMWCSAR algorithm. The corresponding simulation, as well as experimental results are shown in Section 5.5. In Section 5.6, we discussed the results from the simulation and experiment.

5.2 MMWCSAR System

The relative movement between a radar and targets can be used to detect and locate targets. Examples include SAR [73,104], inverse SAR (ISAR) [105–108], moving target indicator [78], etc. To introduce the relative movement between the proposed radar and targets, the placement of the MMWCSAR system is presented in Section 5.2.1. The parameters for the monostatic radar are introduced in Section 5.2.2. In Section 5.2.3, data acquisition is shown. The resolution of the MMWCSAR system is presented in Section 5.2.4. The constraints of MMWCSAR parameters are studied in Section 5.2.5.

5.2.1 MMWCSAR Configuration

The proposed 3D imaging MMWCSAR system uses a single transceiver element to acquire data from targets by emulating multiple transceivers through movement of the single transceiver. To simplify the movement of MMWCSAR, we assume that MMWCSAR moves along a circular track. The plane of the circular track is perpendicular to the range bin axis (in Figure 5.1a). The radar moving along the circular track keeps a constant speed, but the direction changes over time (in Figure 5.1b).
Figure 5.1.: Uses rotation radar to resolve 3D imaging. (a) Schematic view of resolving range using range bins; (b) the swinging within the rotation plane of the radar generates velocities in different directions; (c,d) each range bin is then projected into different velocity directions while data are collected.

As the radar is moving inside the plane perpendicular to the range bin axis, targets detected remain stationary within its range bin while sampling. The movement of radar produces relative speeds of detecting targets. For each separate range bin, detected targets have different relative velocities depending on their azimuth and elevation locations. As the radar moves to different directions, targets within a single range bin project to different directions. If the target is at the center of its range bin, the relative velocity is zero no matter the radar’s movement direction. For targets away from the center, the relative velocities vary according to the radar’s movement directions. Figure 5.1c,d shows that targets within a range bin are projected onto different radar moving directions. Consequently, relative movement produces targets’ Doppler data, which can be used to distinguish targets within the same range bin.

As we can see from Figure 5.2, the geometry of MMWCSAR is presented. $H$ represents a monostatic radar. For different rotation positions, the range-azimuth-elevation of the radar can be presented in Cartesian coordinates as $(H_x, 0, H_z)$: $H_x = r \cos \theta$ and $H_z = r \sin \theta$, where $r$ denotes the rotation radius of the MMWCSAR system, and $\theta$ is the rotation angle at which the MMWCSAR takes a frame of 2D
range-Doppler data. \( \theta \) is related to the angular velocity \( \omega \) and acquisition time stamp \( t \), i.e.,

\[
\theta = \frac{2\pi t}{\omega}.
\] (5.1)

\( A \), \( B \) are point targets ahead of the radar with different positions, and their Cartesian representation are \((A_x, A_y, A_z)\) and \((B_x, B_y, B_z)\). As the spherical coordinates are used for signal processing, \( A \) and \( B \) have spherical coordinates’ profiles of \((R_1, \alpha_1, \beta_1)\) and \((R_2, \alpha_2, \beta_2)\). The radar-target vector components of \( \overrightarrow{HA} \) and \( \overrightarrow{HB} \) relative to radar in Cartesian coordinates are:

\[
\overrightarrow{HA} = (R_1 \sin \beta_1 \cos \alpha_1 - r \cos \theta, R_1 \sin \beta_1 \sin \alpha_1, R_1 \cos \beta_1 - r \sin \theta) \quad (5.2)
\]

and:

\[
\overrightarrow{HB} = (R_2 \sin \beta_2 \cos \alpha_2 - r \cos \theta, R_2 \sin \beta_2 \sin \alpha_2, R_2 \cos \beta_2 - r \sin \theta), \quad (5.3)
\]

respectively. Equations (5.2) and (5.3) provide a method to describe targets in terms of range, azimuth angle and elevation angle instead of range, azimuth location and elevation location. Targets’ 3D location profiles are independent of the radar movement as long as applying the range calibration of the displacement of the origin (radar rotation center) to the radar. The projections are projected onto each range bin and are associated with rotation angle \( \theta \). Therefore, the MMWCSAR addresses a unique approach of the radar remote sensing problem.

The moving direction can be recorded by angle \( \theta \) over time. The geometry of obtaining the projection angle \( \gamma \) can be seen from Figure 5.3. Doppler velocities of targets are the projections onto the radar rotation plane. \( \gamma \) can be represented using the displacement vector \( \overrightarrow{HA} \) and the velocity vector \( \overrightarrow{HV} \):
\[ \gamma = \arccos \frac{\overrightarrow{HA} \cdot \overrightarrow{HV}}{|\overrightarrow{HA}| |\overrightarrow{HV}|}. \quad (5.4) \]

From geometry, we know that:

\[
\overrightarrow{HA} \cdot \overrightarrow{HV} = (\overrightarrow{OA} - \overrightarrow{OH}) \cdot \overrightarrow{HV} \\
= \overrightarrow{OA} \cdot \overrightarrow{HV} - \overrightarrow{OH} \cdot \overrightarrow{HV} \\
= \overrightarrow{OA} \cdot \overrightarrow{HV}.
\]

\[ |\overrightarrow{HA}| \text{ is the actual range of radar and target, } R_{\text{revised}}. \] The range difference between the assumed range \( R_{\text{assumed}} = |\overrightarrow{OA}| \) and \( R_{\text{revised}} \) will be discussed in range calibration Section 5.3.1. Thus, the Equation (5.4) can be simplified as:

\[ \gamma = \arccos \frac{\overrightarrow{OA} \cdot \overrightarrow{HV}}{R_{\text{revised}} |v_{\text{radar}}|}. \quad (5.6) \]

\( |v_{\text{radar}}| \) is the magnitude of the velocity of the radar:

\[ |v_{\text{radar}}| = \frac{2\pi r}{\omega}. \quad (5.7) \]

All vectors are derived from coordinates calculations.
5.2.2 MMWCSAR Parameters

As introduced above, to build a light and low-cost imaging radar, we use a monostatic radar with a single transceiver element.
In our setups, the monostatic radar is transmitting linear frequency modulated (LFM) pulse waveforms and operating in range-Doppler mode. The intermediate frequency (IF) of the radar is defined as center frequency, $f_c$. The bandwidth (BW) of the radar determines the range resolution. In our millimeter wave design, we use a wide bandwidth chirp. Sampling frequency, $f_s$, and pulse chirp duration, $T_P$, define the number of range bins, $N_R$. Pulse repetition interval (PRI) generally determines the blind speed and hence the unambiguous Doppler frequency, $f_d[1]$; thus limiting the swinging angular velocity, $\omega$, in our MMWCSAR system. The number of Doppler bins, $N_D$, defines the velocity resolution and curbs the scanning frames, $N_{Ch}$. The number of frames collected is fundamental to the final FOV image.

5.2.3 Data Acquisition

The radar transmission signal $s(t)$ is the LFM signal. The observation of radar signal for a single scatterer after de-ramping is:

$$r(t) = k \exp\left[j2\pi\left(\frac{2R(BW)}{c}T_P + \frac{2vf_c}{c}\right)t\right] + n(t).$$  

(5.8)

where $k$ is the reflective amplitude related to the target’s radar cross-section (RCS), $n(t)$ is the white noise, $R$ is the range of the target, $v$ is the velocity of the target and $c$ is the speed of electromagnetic wave. Two terms of the frequency component in the exponential function are fast-time and slow-time samples. These samples are the frequency difference of range and Doppler, respectively.

The fast-time sampling frequency of the radar determines the number of range bins, and the pulse repetition frequency (i.e., slow-time sampling frequency) determines the Doppler bins. Radar received signals are forming a time series 1D signal after the analog-to-digital converter (ADC). In Figure 5.4, the fast-slow-time samples accumulated at each angle $\theta_1$, $\theta_2$, ..., $\theta_n$ are reshaped by the number of range
bins, $N_R$, and the number of Doppler bins, $N_D$. Hence, the sampling sequence of fast-time, slow-time and frames (angles) data are organized as a $N_R \times N_D \times N_{Ch}$ complex time domain data matrix. In our MMWCSAR system, the acquisition data format is the fast-time-slow-time-angle datacube, as the fast-time and slow-time data are associated with range and Doppler (projections), respectively. This datacube can be processed into range-Doppler angle data after pulse-Doppler processing. The IRT is related to the Doppler angle planar data for each range bin. Because the relative velocities of targets caused by circular movement project different velocities based on the azimuth/elevation location onto different angle profiles, the IRT method can be applied to reconstruct the image in each range bin of our imaging geometry, which is similar to computed tomography [86]. Some IRT-applied radar techniques can be found in [105]. Consequently, the Doppler-angle data matrix can be extracted for imaging restoration of each range bin. 3D imaging can be obtained from recovering 2D images for each range bin using the “2D + 1D” model [93].

Figure 5.4.: From projections to radar signal datacube.

For LFM waveform, its pulse-Doppler processing is coupled together [109]. However, in our approach, we implement compressed sensing (in Section 5.4) to improve
the final image quality of the proposed radar. The pulse compression is done separately from the Doppler compression.

After forming a 3D datacube, the uncompressed received echo signals are pulse compressed by discrete Fourier transform (DFT) of the transmitted signal along the range axis. In Section 5.3, we separate the range profile and process the signal on 2D profile of the datacube to obtain the FOV figure on each range bin.

From the acquisition stage, we are using the fast-time, slow-time and angles in forming a 3D datacube. The pulse compression converts the datacube into the range-slow-time-angle datacube; and applying IRT to obtain range-azimuth-elevation datacube. The last datacube is the volume FOV figure of the actual image.

5.2.4 MMWCSAR Resolution

The MMWCSAR system has 3D imaging capacity. Therefore, the resolution is an essential topic for high resolution imaging.

For the range resolution, the range profile is independent of the Doppler and angle profiles. Thus the range resolution is:

$$\Delta R = \frac{c}{2(BW)}.$$  \hfill (5.9)

For azimuth and elevation resolution, they are dependent on Doppler and angle profiles. Both resolutions are equivalent because the MMWCSAR is moving along a circular track with even angle spaces. Due to polar to Cartesian interpolation, the resolution of azimuth or elevation is higher around the rotation center and is lower at the edge. The azimuth/elevation resolution $\Delta l$ is defined as:

$$\Delta l = 2l \sin \frac{\pi (PRI) N_D}{\omega},$$  \hfill (5.10)
where \( l \) is the projection distance from the center. \( l \) has the limit from the center of origin to the azimuth/elevation edge, which is:

\[
0 \leq l \leq R \tan \frac{\theta_{FOV}}{2} + r,
\]

where \( R \) means the range at which we measure the azimuth/elevation resolution and \( \theta_{FOV} \) denotes the FOV angle of radar looking vision. Because the resolution depends on the location of the azimuth/elevation FOV figure, in general, the worst resolution in a azimuth/elevation FOV figure is used to judge the MMWCSAR system’s azimuth/elevation resolution. Therefore, the resolution for the azimuth/elevation is:

\[
\Delta l = 2(R \tan \frac{\theta_{FOV}}{2} + r) \sin \pi \frac{(PRI)N_D}{\omega}.
\]

Note that the resolution of azimuth/elevation is dependent on the range at which we measure the azimuth/elevation resolution.

### 5.2.5 Constraints of Parameters

In order to reconstruct a high quality image, more data should be acquired in the 3D datacube in the acquisition stage. However, the number of data have some limitations as below.

#### Constraints of Number of Doppler Bins

Because the radar is moving and targets are static within the sampling period, targets have relative velocities with respect to the MMWCSAR depending on their azimuth and elevation locations. If the target is far from the center of its range bin, the relative velocity increases. Accumulated Doppler bins define the resolution of the Doppler frequency based on targets’ azimuth and elevation locations. Hence, to
better reflect the relative velocities of targets in the datacube, more Doppler bins are needed. We need to have enough Doppler bins to cover all of the relative velocities within the FOV angle:

\[
\frac{N_D + 1}{2} \Delta v_D \geq \frac{4\pi r}{\omega} \sin \frac{\theta_{FOV}}{2},
\]

(5.13)

where \( \Delta v_D \) is the velocity interval between two adjacent Doppler bins. \( \Delta v_D \) is defined as:

\[
\Delta v_D = \frac{c}{2f_c(PR)/(N_D)},
\]

(5.14)

where \( c \) is the speed of the electromagnetic wave.

By choosing the appropriate number of Doppler bins, the system is able to capture all needed data for range-Doppler response within a limited given time period. This allows the system to capture enough frame data to form a range-Doppler-frames datacube. In this case, the frame data serve as the rotation angle. Therefore, the 3D datacube is consisting of range-Doppler-angle with the magnitude of targets. PRI also has a limit, which needs to cover all of the relative velocities with respect to radar to avoid blind speed:

\[
\frac{2\pi r}{\omega} \leq \frac{c}{2f_c(PR)}.
\]

(5.15)

**Constraints of the Number of Angle Bins**

However, choosing too many Doppler bins results in fewer range-Doppler responses per full-round scan. To increase the number of angle bins, one needs to decrease the time used for capturing fast-time-slow-time samples. Hence, the system needs to take less time in capturing per fast-time-slow-time samples, so that allows more angle
data (frames data) recorded through rotating. The sampling frame time for each fast-time-slow-time sample is:

\[ T_{Ch} = (PRI)(N_D) \].

(5.16)

The radar movement is relatively constant when collecting fast-time-slow-time samples. That is to say, to achieve each sample within 10° rotation, thus allowing 36 angle samples per full-round scan, we have another constraint:

\[ \frac{2\pi T_{Ch}}{\omega} \leq 2\pi \left( \frac{10^\circ}{360^\circ} \right). \]

(5.17)

Simplify the equation, and we get:

\[ \omega \geq 36(PRI)(N_D). \]

(5.18)

The maximum detectable Doppler frequency is restricted by Doppler compression:

\[ f_{d,\text{max}} = \frac{2\pi r 2f_c}{\omega c} \leq \frac{2}{PRI}. \]

(5.19)

The same, the minimum detectable Doppler frequency is restricted by:

\[ f_{d,\text{min}} = \frac{2\pi r 2f_c 1}{\omega c N_D} \leq \frac{2}{(PRI)(N_D)}. \]

(5.20)

Merging Equations (5.13)–(5.20), we get constraints with our system. A proper choice of parameters is:

\[ N_D = 25, PRI = 30 \times 10^{-6} \text{ s}, \]
\[ r = 0.2 \text{ m}, \omega = 0.6 \text{ s/round}. \]

In the later sections, simulation and experiment restrictions follow constraints discussed in this section. The different chosen parameters can result in different scan-
ning schemes. Thus, this depends on targets and detecting scenarios. For example, we want to see metal objects concealed behind people’s clothes [110]. We need to increase the swinging rate and improve the FOV resolution. Using our constraints, we reduce our frames and PRI in order to meet the criteria.

5.3 Radar Imaging Processing

In this section, we discuss range calibration, as well as the imaging processing for the receiving datacube.

5.3.1 Range Calibration

From Section 5.2.1, additional range calibration is needed. This is because: the fan-shaped range bin should be converted into a plane-shaped range bin; the range profile of the radar should match with the rotation position, as we assumed the rotation radius is zero. From Figure 5.2, for a single target $A$, the displacement vector from radar is the actual measured range. The assumed range is from the origin, $\vec{OA} = (R_1 \sin \beta_1 \cos \alpha_1, R_1 \sin \beta_1 \sin \alpha_1, R_1 \cos \beta_1)$, which is not changing throughout time. The radar location at different time is $\vec{OH} = (r \cos(2\pi t/\omega), 0, r \sin(2\pi t/\omega))$. Thus, the revised range is:

$$R_{\text{revised}} = |\vec{HA}| = |\vec{OA} - \vec{OH}|$$

$$= \sqrt{(R_1 \sin \beta_1 \cos \alpha_1 - r \cos \frac{2\pi t}{\omega})^2 + (R_1 \sin \beta_1 \sin \alpha_1)^2 + (R_1 \cos \beta_1 - r \sin \frac{2\pi t}{\omega})^2}$$

(5.21)
The range difference of the revised range and assumed range is:

\[ R_{\text{diff}} = R_{\text{revised}} - R_{\text{assumed}} = |\overrightarrow{HA}| - |\overrightarrow{OA}| \]

\[ = \sqrt{(R_1 \sin \beta_1 \cos \alpha_1 - r \cos \frac{2\pi t}{\omega})^2 + (R_1 \sin \beta_1 \sin \alpha_1)^2 + (R_1 \cos \beta_1 - r \sin \frac{2\pi t}{\omega})^2} \]

\[ - \sqrt{(R_1 \sin \beta_1 \cos \alpha_1)^2 + (R_1 \sin \beta_1 \sin \alpha_1)^2 + (R_1 \cos \beta_1)^2} \]

(5.22)

As \( |\overrightarrow{HA}| \) is only dependent on the time of the rotation, thus the range-Doppler-angle 3D datacube requires the calibration of range using \( R_{\text{diff}} \) at each range-Doppler planar data corresponding to time. Time in our MMWCSAR system is related to the rotation angle \( \theta \), which is the angle profile of the datacube. A time-indexed matrix \( V^{-1} \) is thus able to perform the range calibration in compressed sensing in Section 5.4.

### 5.3.2 Radon Transform and 3D Imaging Reconstruction

Radon transform and 3D image reconstruction are illustrated in many applications, i.e., computed tomography [86] and thermoacoustic tomography [111]. Obtaining tomographic image from projections data and conversion has been introduced in [86, 89–91, 112, 113]. The cross-sectional image of targets at each range bin consists of projection angle \( \theta \) and the projection distance \( \xi \). The IRT is performed for reconstruction tomography.
The Radon transform is an integral transform converts the 2D image to its projections, \( p(\xi, \theta) \). The Radon transform can be defined as [86]:

\[
p(\xi, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\xi - x \cos \theta - y \sin \theta) \, dx \, dy ,
\]

(5.23)

in which \( f(x, y) \) denotes the original 2D density distribution function indexed by \( x \) and \( y \); \( \xi \) is the projection distance from center; \( \theta \) is the projection angle; and \( \delta(\cdot) \) is the Dirac delta function.

Its inverse transform, IRT, is widely used for image reconstructions. IRT can reconstruct the image from the projection data by several techniques. Traditionally, people used the back-projection theorem to recover the inverse Radon transform [112]. A recent CT development inspired the central slice theorem (CST) [86], which is the simplest method conceptually compared to back-projection and iterative algebraic techniques [91]. The theorem states that the 2D Fourier transform (FT) of the original function \( f(x, y) \) is the function of 1D Fourier transforms of the projection slices in the order of angles. The 2D FT of the original function is:

\[
F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp[-j2\pi(ux + vy)] \, dx \, dy.
\]

(5.24)

\( P(\rho, \theta) \) is the 1D FT series of the projections \( p(\xi, \theta) \), which can be represent by:

\[
P(\rho, \theta) = \int_{-\infty}^{\infty} p(\xi, \theta) \exp(-j2\pi \rho \xi) \, d\xi.
\]

(5.25)

Hence, the 2D Fourier domain function \( F(u, v) \) can be obtained from the Fourier domain function \( P(\rho, \theta) \) by interpolation between polar and Cartesian coordinates:

\[
P(\rho, \theta) = F(u, v) \big|_{u=\rho \cos \theta, v=\rho \sin \theta}.
\]

(5.26)
Therefore, the spatial image can be obtained from projection slices by CST. In this MMWCSAR system, the Doppler-angle data can be converted by this algorithm to reconstruct FOV images of targets’ scene at different range bins.

3D Imaging Reconstruction and Point Spread Function

The profile of range data is independent of the Doppler-angle data after the geometric calibration of the range. The Doppler-angle planar data can be used to reconstruct the 3D image. For each range bin, Doppler-angle data are the projections of the 2D azimuth/elevation FOV figure in the back-projection domain [114]. The transformation of the Doppler-angle and 2D FOV figure is basically using the Radon-inverse Radon transform pair. This pair can be expressed as an invertible matrix of the Radon transform-IRT linear matrix with interpolation of the Cartesian and polar coordinates. Thus, by using the “2D + 1D” model [93], 2D azimuth-elevation FOV images are resolved at different range bins. These images are adding the range bin profile to reconstruct a volume FOV figure. This allows 3D imaging.

For the point spread function (PSF), it defines the response of an imaging system to a point source [115]. The MMWCSAR system azimuth-elevation FOV image is formed as the response from point scatterers with the following summation:

\[
MMWCSAR(x, y) = \sum_{i=1}^{N} A_i \delta(x - x_i, y - y_i) \ast h(x, y), \tag{5.27}
\]

where \(N\) represents the total number of point scatterers and \(A_i\) is the scattered field amplitude. \(x, y\) is the azimuth and elevation location, respectively. \(h(x, y)\) is the PSF of the MMWCSAR system. \(h(x, y)\) can be regarded as the system’s impulse response to any point scatterers of the target. The image can be expressed as the convolution of scatterers with the PSF. For our system, the PSF for one scatterer at a range of 5 m at the center of the radar view is shown in Figure 5.5.
From Figure 5.5, the side lobe level is $-13$ dB. The PSF has circular side lobes from the center. This is caused by IRT along all directions. The side lobes exhibit a sinc-function shape because of the finite bandwidth in the Doppler and angle domain. The common way to suppress the side lobes is to use windowing. In our MMWCSAR system experiment, we are using the Hanning window.

![2D PSF at range of 5 meters (amplitude in dB)](image)

Figure 5.5.: 2D point spread function (PSF) for the point scatterer located at the center $(0,0)$ at a range of 5 m (amplitude in dB).

### 5.4 Compressed Sensing for 3D Imaging Radar System

In terms of improving sensing results and boosting the rotating accuracy, compressed sensing is used in our MMWCSAR.
5.4.1 Compressed Sensing Review

The basic CS idea is reviewed below. Suppose that we have a two-dimensional (2D) image with a size of $A \times B$, which can be extract into a one-dimensional (1D) vector $\vec{f}$ with a length of $N = AB \times 1$. Any 1D vectors can be constructed by sampling the sensing basis $\phi = [\phi_1|\phi_2|...|\phi_M]$ [94]. $M$ is the number of measurements of the sensing basis $\phi$ and is smaller than the length of $\vec{f}$. Thus, the sampling vector $\vec{y}$ can be expressed as:

$$\vec{y} = \phi_k \vec{f}, \quad k = 1, 2, ..., M.$$  \hfill (5.28)

The restriction laying the $K$-sparse signal $\vec{f}$ should have $K < M < N$, which allows the signal recovery from the $M$ measurements.

CS requires the sparsity and incoherent sampling. For sparsity, we have $\vec{f} \in \mathbb{R}^N$, which is $N = AB$ pixels in the 2D image. For an orthogonal transform, i.e., discrete cosine transform (DCT), almost all of the pixels could have sparse expansion without much perceptual loss. This results in a representation basis $\psi = [\psi_1|\psi_2|...|\psi_N]$ [94], which allows the following representation:

$$\vec{f} = \sum_{i=1}^{N} \psi_i x_i.$$ \hfill (5.29)

For incoherence sampling, from [94], the required sensing basis and representation basis should have the following incoherence parameter:

$$u(\phi, \psi) = \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \phi_k, \psi_j \rangle|.$$ \hfill (5.30)

The reconstruction uses $l_1$-norm minimization [116–118]. The proposed solution $\hat{f}$ is constructed by $\hat{f} = \phi \hat{x}$, where $\hat{x}$ is the solution to the convex optimization:

$$\hat{x} = \min\{\|\vec{x}\|_1 : \vec{y} = \phi \psi \vec{x}\}.$$ \hfill (5.31)
Thus, the measurement \( \hat{x} \) is produced by the sparsest signal of the decoding model.

### 5.4.2 Compressed Sensing on Doppler-Angle Data

As the Doppler profile is in the frequency domain and the angle profile is in the spatial domain, slow-time data are used, and the pulse compression is done apart from pulse-Doppler compression. Consequently, we have a DFT matrix of \( U^{-1} \). \( U^{-1} \) is a 2D complex matrix, which has 1D DFT along fast-time bins, with slow-time and angle profiles repeating to match the whole 3D data size. \( U \) is then the transform from range frequency to fast-time samples with repetition of slow-time and angle profiles. The range calibration matrix can be expressed as \( V^{-1} \).

We also implement the RT and IRT matrix for the CS. The projections on each range profile are the projection domain data of the 2D azimuth-elevation FOV figure. We use a rectangular IRT matrix \( W^{-1} \) transform the slow-time-angle 2D data to the 2D azimuth-elevation FOV figure with the range profile repeating to match the whole 3D data size. The IRT matrix representation \( W^{-1} \) is produced combined with linear interpolation from polar coordinates to Cartesian coordinates. Thus, we have a conversion from the original 1D-reshaped range-slow-time-angle data \( \tilde{b} \) to the 1D-reshaped calibrated range-azimuth-elevation data \( \tilde{x} \):

\[
\tilde{b} = UVW \tilde{x}.
\]  

(5.32)

Fourier transform and Radon transform are linear transforms. Both of the transformation matrices are invertible. In addition, both \( U^{-1} \) and \( W^{-1} \) work on 1D-reshaped range-slow-time-angle data. Hence, CS requires solving \( \tilde{x} \) by using three linear mapping matrices \( W^{-1}V^{-1}U^{-1} \):

\[
\tilde{x} = W^{-1}V^{-1}U^{-1} \tilde{b}.
\]  

(5.33)
In this approach, the representation basis $\psi$ is represented as:

$$\psi = U V W. \quad (5.34)$$

For any condensed signal $\vec{y}$, the sensing of the range-slow-time-angle data $\vec{b}$ can be expressed as:

$$\vec{y} = \phi \vec{b}. \quad (5.35)$$

Therefore, the reconstruction is able to be implemented by using $l_1$-minimization on the sensing basis $\phi$ and the representation basis $\psi$ from Equation (5.34):

$$\hat{x} = \min \{\|\vec{x}\|_1 : \vec{y} = \phi U V W \vec{x}\}. \quad (5.36)$$

To summarize, $U^{-1}$ converts the 1D-reshaped fast-time-slow-time-angle original data into 1D-reshaped range-slow-time-angle data. $V^{-1}$ is the calibration matrix of the range bins. $W^{-1}$ transforms the range-slow-time-angle data into 1D-reshaped range-azimuth-elevation FOV figure.

In our simulation and experiment, we solve the convex optimization through the MATLAB primal-dual interior point method [119,120]. The sparsity of Doppler-angle planar data is exploited. Because angle data are limited due to PRI and the ADC sampling rate when MMWCSAR accumulates data, the Doppler-angle data form a sparsity signal (multiple sinusoidal shapes) from targets. Through the optimization, the sparsity of the original slow-time-angle signal can be exploited to recover from fewer samples than that of the Nyquist sampling theorem [94]. Therefore, the compressed sensing is able to be implemented to recover the 3D imaging from fast-time-slow-time-angle datacube.
5.4.3 Sensing Basis $\phi$ Selection

We first define the compressed sensing ratio (CSR) as $R_{CS}$, expressed as:

$$R_{CS} = \frac{N}{M}, \quad (5.37)$$

$M$ and $N$ are the number of rows and columns of the sensing basis, respectively. CSR is the ratio of the length of the expected signal $\vec{b}$ over the length of condensed signal $\vec{y}$. It specifies the reconstruction quality and size of the sensing basis and representation basis. The slow-time-angle data CS is done with compressing both slow-time and angle. These data are transformed on each calibrated range profile. Sensing basis $\phi$ can be expressed as follows:

1. Reduced rotation acquisition matrix:

   The matrix is expressed as:

   $$\phi_{k+h,k} = \begin{cases} 
   1, & k = 1, 2, \ldots, M & \& 0 \leq k + h \leq M \\
   0, & \text{otherwise}
   \end{cases} \quad (5.38)$$

   $h$ is the offset depends on the rotation span. This allows sensing the expected signal with reduced rotation angles. The sensing basis reconstructs the condensed signal into a full angular projections Doppler-angle signal along the angle profile. Different projections provide different IRT responses. Thus, this sensing basis enables the sampling at fewer angle bins, which reduces the swinging inaccuracy. This sensing matrix is also applicable to the Doppler profile, which also improves projections along angle profile.
(2) Reduced sampling matrix:

This matrix is expressed as:

\[
\phi_{k,\lfloor R_{CS} \rfloor (k-1)+1} = \begin{cases} 
1, & k = 1, 2, ..., M \\
0, & \text{otherwise}
\end{cases}
\] (5.39)

\(\lfloor R_{CS} \rfloor\) denotes the max integer smaller than \(R_{CS}\). This allows the sensing basis sensing the expected signal at its higher sampling rate. The sensing basis converts the condensed signal with more projection data along the Doppler or angle bins. This method avoids swinging at an inconsistent rate of the velocity. This allows the signal recovery at better constant sinusoidal-shaped Doppler-angle data.

(3) Gaussian or random matrix:

The Gaussian sensing basis is shown below:

\[
\phi_k = \exp\left[-(k - \mu)^2/(2\sigma^2)\right], \quad k = 1, 2, ..., M .
\] (5.40)

\(\mu\) is the expected value of the Gaussian, and \(\sigma\) is the standard variance of the Gaussian. The random matrix is also applicable. This method allows the sampling followed the normal compressed sensing procedure. This corresponds to some CS applications in MRI, i.e., [97]. The sensing basis converts the condensed signal in a smoother way. It gives the compensation to the signal. It allows the signal recovery with much more precise points along the Doppler-angle data.

The following simulation and experiment will provide the results of CS involved in the MMWCSAR system. The comparison will be provided with the IRT method.
5.5 Simulation and Experiment

5.5.1 Simulation Setup and Results

MMWCSAR is able to reconstruct the 3D image along each range bin. The following simulation provides a FOV figure of range of 5 m. The MMWCSAR system parameters can be found in Table 5.1. The four targets’ scheme parameters are shown in Table 5.2.

Table 5.1: MMWCSAR simulation setup. PRI, pulse repetition interval.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center frequency</td>
<td>$f_c$</td>
<td>76.5 GHz</td>
</tr>
<tr>
<td>Chirp starting frequency</td>
<td>$f_{start}$</td>
<td>76 GHz</td>
</tr>
<tr>
<td>Chirp end frequency</td>
<td>$f_{stop}$</td>
<td>77 GHz</td>
</tr>
<tr>
<td>PRI</td>
<td>$PRI$</td>
<td>$30 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Chirp duration</td>
<td>$T_P$</td>
<td>$20 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Fast time samples</td>
<td>$N_R$</td>
<td>400</td>
</tr>
<tr>
<td>Slow time samples</td>
<td>$N_D$</td>
<td>100</td>
</tr>
<tr>
<td>Frame samples</td>
<td>$N_{Ch}$</td>
<td>68</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>$f_s$</td>
<td>$20 \times 10^6$ Hz</td>
</tr>
<tr>
<td>Angular velocity</td>
<td>$\omega$</td>
<td>0.2 s/round</td>
</tr>
<tr>
<td>Rotation radius</td>
<td>$r$</td>
<td>0.6 m</td>
</tr>
<tr>
<td>Signal to noise ratio (SNR)</td>
<td>$SNR$</td>
<td>10 dB</td>
</tr>
</tbody>
</table>

Table 5.2: Four targets’ scheme setup. RCS, radar cross-section.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Target 1</th>
<th>Target 2</th>
<th>Target 3</th>
<th>Target 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>5 m</td>
<td>5 m</td>
<td>5 m</td>
<td>5 m</td>
</tr>
<tr>
<td>Azimuth location</td>
<td>−1.2941 m</td>
<td>1.7101 m</td>
<td>0 m</td>
<td>2.1131 m</td>
</tr>
<tr>
<td>Elevation location</td>
<td>0.8682 m</td>
<td>1.7101 m</td>
<td>0 m</td>
<td>−0.4358 m</td>
</tr>
<tr>
<td>RCS</td>
<td>1 m²</td>
<td>1 m²</td>
<td>1 m²</td>
<td>1 m²</td>
</tr>
</tbody>
</table>
As the MMWCSAR system is producing the radar transmitted LFM waveforms, the received signal together with produced Gaussian noise is analyzed with the IRT method and the involved CS method. Additionally, in order to process Doppler-angle data without the influence of the range profile, the range calibration is necessary in the simulation. From the 3D datacube, the 3D fan-shaped range bin is accumulated. From Equation (5.22), we process every Doppler-angle slice along the range axis with a calibration matrix $V$ to convert the fan-shaped range bin into a plane-shaped range bin.

The simulation results are shown in Figure 5.6 with FOV figures of 5 m seen from the radar panel. Figure 5.6a addresses the IRT method with the perfect setup without swinging inaccuracies. The velocity of the radar is changing evenly with direction and a constant magnitude. Figure 5.6b–g introduces all of the involved CS method by applying different sensing basis on slow-time-angle bins separately. For the experiment, we will use the Equation (5.38) matrix for the slow-time profile as in Figure 5.6d, as it provides high contrast and better scatterer indication.
Figure 5.6.: Simulation results (front view from radar panel at 5 m, the $x$ axis and $y$ axis are azimuth and elevation distance from the center, respectively). (a) Inverse Radon transform (IRT) method with the perfect setup; (b) compressed sensing (CS) applying Equation (5.38) to the angle profile with $R_{CS} = 1/2$; (c) CS applying Equation (5.39) to the angle profile with $R_{CS} = 1/2$; (d) CS applying Equation (5.40) to the angle profile with $R_{CS} = 1/2$; (e) CS applying Equation (5.38) to the slow-time profile with $R_{CS} = 1/2$; (f) CS applying Equation (5.39) to the slow-time profile with $R_{CS} = 1/2$; (g) CS applying Equation (5.40) to the slow-time profile with $R_{CS} = 1/2$. 
5.5.2 Experiment Setup and Results

With the equipment from the INRAS MIMO radar [121,122] and three metal ball targets, a basic system can be set up. The INRAS MIMO radar is a four-transmitter, eight-receiver MIMO radar. We use only one transceiver element, which satisfies the MMWCSAR configuration. The radar works at 77 GHz in range-Doppler mode. The ball targets are made from metal with a high RCS, which can produce high reflective beams compared to other targets' reflections. The parameters for the experiment can be seen in Table 5.3. The experimental setup scheme is shown in Figure 5.7a. The ball lineup with measurement is given in Figure 5.7b. The balls are set at a range of 1.21 m, 1.66 m and 2.16 m from the radar panel, respectively.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Target 1</th>
<th>Target 2</th>
<th>Target 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>1.21 m</td>
<td>1.66 m</td>
<td>2.16 m</td>
</tr>
<tr>
<td>Azimuth angle $\alpha$</td>
<td>$\approx -2^\circ$</td>
<td>$\approx -5^\circ$</td>
<td>$\approx 5^\circ$</td>
</tr>
<tr>
<td>Elevation angle $\beta$</td>
<td>$\approx 0^\circ$</td>
<td>$\approx 0^\circ$</td>
<td>$\approx 0^\circ$</td>
</tr>
</tbody>
</table>
The rotating angle per frame is around 15.3 degrees, based on the calculation of $360^\circ$ over the number of frames $N_{Ch}$ accumulated by the signal processing unit. Multiple swinging circles are recorded in order to reduce the impact of inaccuracies caused by hand. Additionally, in order to minimize the hand swinging inaccuracy on Doppler-angle data, the range calibration is necessary in the experiment. Otherwise, some ghost targets will be observed around detected targets. The calibration method is similar to that of the simulation part.

The results provide a 3D scattering plot of the FOV in front of the MMWCSAR. The indices of the figure are the range in meters, the azimuth angle and the elevation angle in degrees. The figures of the applying IRT and CS methods can be seen in Figure 5.8. We add the threshold to distinguish the metal ball targets from the table, the board and walls. The metal ball targets’ reconstruction using the IRT method (Figure 5.8a,b) gives a slightly larger size in azimuth and elevation directions compared to the actual size. Reconstruction using the CS method (Figure 5.8c,d)
provides a more precise scatterer indication, but lost the shape of the metal ball. The range resolution is 37.5 mm for both methods, as the range resolution is dependent on bandwidth only. The azimuth/elevation resolution based on Equation (5.12) is 36.2 mm for the 2.16-m range bin. The experiment measured azimuth/elevation resolution is 44 mm from the worst recognized 2.16-m target. The CS method provides a better scatterer indication, but the IRT method gives a more precise ball shape. In conclusion, the MMWCSAR system using both methods is capable of 3D imaging by moving the radar along a circular track.

Figure 5.8.: Experiment results. (a,b) IRT method; (c,d) CS applying Equation (5.38) to the slow-time profile.
5.6 Discussion

5.6.1 Millimeter Wave 3D Imaging Radar

For both the IRT and CS method, the 3D range-azimuth-elevation FOV figure is recovered. The proposed system can have the following advantages:

- Fast:

  To recover the 3D imaging, the proposed system needs to collect data when the system is moving. A full circular movement track is needed for the IRT method to collect data, while the CS method can greatly decrease the number of samples. It takes only seconds for MMWCSAR to recover an FOV figure.

- Portable:

  Due to its size, the proposed system can be wearable for 3D imaging. Both the IRT and CS methods produce convincing results from the simulation and experiment.

- High resolution:

  With other SAR imaging device shown in [92, 93], the resolution is around 0.1 m. Our MMWCSAR working on 77 GHz can achieve 37.5-mm range resolution and 36.2-mm azimuth/elevation resolution at a range of 2.16 m (from Equation (5.12) and experiment data), theoretically. Due to hand swinging inaccuracy and noise, the resolution is reduced to around 44 mm in experimental measurements.

  Due to limited access to resources, we rotate the platform manually and record the time we finish one rotation. Due to the rough estimation of the rotation, the accuracy and resolution are lower than the theoretical analysis. To compensate the errors, we set a calibration target, e.g., we use a corner reflector at a range bin without any other objects; we track the errors reading the Doppler-angle profile of the corner reflector. After compensation, the imaging still has decreased resolution. However, if we can
integrate an inertial sensor into the system to extract real-time velocity information, a more precise and accurate figure could come out.

5.6.2 Comparison with the Radon Method and the Compressed Sensing Method

The conversion of the 3D data matrix is linear, and the signal is sparse. The inverse transform of the matrix is accessible, thus allowing the involved CS method. Comparing to the IRT method, the CS method gives the following advantages:

- Flexible:

  From the simulation and experiment, it is clear that the involved CS MMWCSAR system is more flexible on data reconstruction. The calculations in the signal processing module are only involved in the matrix multiplexing and solving $l_1$-minimization. The acquisition is allowed more freedom. Based on the sampling of targets, the CS method is able to reconstruct a better image than that of the IRT method. Data accumulated are not limited to a full circle. Mistakes and errors can be eliminated from the CS method, as well. Besides, $R_{CS} = 1/2$ is used for both simulation and experiment. It is flexible to adjust the compressed sensing ratio to improve the MMWCSAR FOV image further.

- High-SNR:

  The peak of targets is more recognizable with a high decibel difference to the background compared to that of the IRT method. For example, in the simulation part, the CS provides a 43 dB peak identification to a 24 dB peak in the IRT method. The peak power is measured as $-32.6$ dB in the CS method compared to $-80.8$ dB in the IRT method. Besides, improvements on scatterer indication are also noticeable. In the experiment part, targets’ reflective scatterers are shown in more accurate locations. Targets are more easily recognized by the involved CS method of MMWCSAR.
• Fast-acquisition:

As matrix transformation is implemented in the CS method. Huge data convolution along different axes is accomplished. Fast implementation on acquisition is achieved at the cost of more time spent on signal processing. It is an advanced signal processing method used in large data matrices in modern imaging devices.
6. STRIP MAP SYNTHETIC APERTURE RADAR FOR AUTOMOTIVE RADAR

6.1 Background

The combination using strip map SAR technique FMCW automotive radar range-Doppler response (RD response) leads an effective imaging method besides target tracking and indication commonly used in automotive radars. The current autonomous driving research prefers using lidar for simultaneous localization and mapping (SLAM) as lidar provides a better resolution for localization surrounding areas. However, the road side conditions for autonomous driving cars are complicated. The mmWave imaging provides additional information surrounding the vehicle. The mmWave imaging system is capable of mapping the obstacles and targets. Compared to lidar, mmWave radars are low-cost, consistent on all weather conditions. Therefore, a mmWave imaging system is discussed in this chapter.

Many range-Doppler SAR imaging is introduced in [123–127]. In [123], Shao et al gave an overview and equations on squinted airborne SAR imaging. In [124], Deng et al introduced the extended RD algorithm for SAR using Legendre orthogonal polynomials. The method using mmWave imaging with single transceiver is introduced in [71]. Some terrain mapping techniques including lidar SLAM, mmWave SLAM and lidar-radar fusion mapping are provided in [128–132].

In this chapter, we demonstrate a novel way using RD response to form a SAR terrain mapping while the vehicle is driving. With the interpolation of vehicle speed and time from current autonomous driving test car based on robotic operating system
The structure of this chapter is as follows. In Section 6.2, the methodology of mmWave terrain mapping is introduced. In Section 6.3, experiments of different bandwidth using mmWave automotive radar terrain mapping in a parking lot are presented.

### 6.2 Terrain Mapping Methodology

Since only a single channel 77 GHz automotive radar is performed for SAR, the digital beamforming method and other multi-channel methods are not applicable. The vehicle velocity and location data are fused from ROS sensors of the autonomous driving car.

The procedures of our proposed system can be seen in Fig. 6.1. The 2D fast-slow time is accumulated with repeating FMCW chirps of the 77 GHz radar. After low
pass filter (LPF) and analog-to-digital converter (ADC), electromagnetic (EM) waves convert to digital signal based on the sampling frequency using onboard processing. The stretch processing, filtering, range fast Fourier transform (FFT) and azimuth FFT is processed by computer. The range cell migration correction method (RCMC) [124] is applied before sensors fusion.

The side looking SAR scheme is shown in Fig. 6.2. The RCMC method is briefly the calibration of the range cells of different Doppler bins. The transfer function can be expressed as:

$$H(\omega) = \exp \left[ j \frac{2\pi}{c} R_s f_r \sin^2 \theta \right]$$  \hspace{1cm} (6.1)

where $c$ is the EM wave speed, $R_s$ is the reference range from radar to the center of the scene, $f_r$ is the range frequency and $\theta$ is the squinted angle, which is related the azimuth (Doppler) frequency.
Table 6.1: 300 MHz bandwidth experiment setup

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center frequency</td>
<td>$f_c$</td>
<td>77 GHz</td>
</tr>
<tr>
<td>Chirp starting frequency</td>
<td>$f_{\text{start}}$</td>
<td>76.85 GHz</td>
</tr>
<tr>
<td>Chirp end frequency</td>
<td>$f_{\text{stop}}$</td>
<td>77.15 GHz</td>
</tr>
<tr>
<td>Pulse repetition interval</td>
<td>PRI</td>
<td>$80 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Chirp pulse width</td>
<td>$T_P$</td>
<td>$25 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Fast time samples</td>
<td>$N_R$</td>
<td>128</td>
</tr>
<tr>
<td>Number of chirps</td>
<td>$N_D$</td>
<td>500</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>$f_s$</td>
<td>6.67 MHz</td>
</tr>
</tbody>
</table>

Secondly, nodes subscribed from ROS is fused with the SAR data. The fusion of the ROS sensors providing radar with the live velocity of the vehicle and terrain mapping locations. Since the automotive radar is not ROS compatible, we need to align with the ROS sensors.

The terrain mapping is then done with interpolation and mapping from the fused radar/GPS data. The tomographic map is accumulated throughout the driving with different speed. The SAR mapping with emulating multiple transceivers along the track is then projected onto the final mmWave terrain map.

6.3 Experimental Results

6.3.1 Setup

As we stated above, we are implement range-Doppler response in SAR reconstruction model. The experiment consists of using two different bandwidth setups. The 300 MHz bandwidth parameters are introduced in table 6.1.

Thus the bandwidth of the automotive radar SAR is 300 MHz, range solution is 0.5 meters, the maximum range to detect is 64 meters, the range-Doppler acquisition
Table 6.2: 1 GHz bandwidth experiment setup

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center frequency</td>
<td>$f_c$</td>
<td>77 GHz</td>
</tr>
<tr>
<td>Chirp starting frequency</td>
<td>$f_{\text{start}}$</td>
<td>76.5 GHz</td>
</tr>
<tr>
<td>Chirp end frequency</td>
<td>$f_{\text{stop}}$</td>
<td>77.5 GHz</td>
</tr>
<tr>
<td>Pulse repetition interval</td>
<td>$PRI$</td>
<td>$80 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Chirp pulse width</td>
<td>$T_P$</td>
<td>$25 \times 10^{-6}$ s</td>
</tr>
<tr>
<td>Fast time samples</td>
<td>$N_R$</td>
<td>128</td>
</tr>
<tr>
<td>Number of chirps</td>
<td>$N_D$</td>
<td>500</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>$f_s$</td>
<td>6.67 MHz</td>
</tr>
</tbody>
</table>

interval is 0.04 seconds, Doppler resolution is 0.0484 m/s and the maximum speed allowed for vehicle is 24.2 m/s or 87 km/h.

Except the 300 MHz setup, we added 1 GHz bandwidth for comparison. Using higher bandwidth reduces the maximum range but improves the range resolution. The parameters are shown in Table 6.2.

Thus the bandwidth of the second test is 1 GHz, range solution is 0.15 meters, the maximum range to detect is 19.2 meters, the range-Doppler acquisition interval is 0.04 seconds, Doppler resolution is 0.0484 m/s and the maximum speed allowed for vehicle is 24.2 m/s or 87 km/h.

The vehicle is a 2009 Ford Escape Hybrid, and was professionally actuated. The interface for the vehicles control is primarily through the ROS, and various simulators and sensors that utilize ROS are also available. The mounted radar with testing car is shown in Fig. 6.3. The testing ground is the parking lot 3039 in front of Electrical and Computer Engineering Building of the University of Arizona. The satellite image provided by Google Maps can be seen in Fig. 6.4.
6.3.2 Results

We tested the radar platform with 300 MHz on the accumulation interval of 0.347 seconds per RD response (based on the radar accumulation laptop). The synthesizing
is based on the converting the ROS time to the universal time. Based on the SAR terrain mapping technique, we obtain the terrain mapping of 300 MHz in Fig. 5.

In Fig. 6.5, the Figures 6.5a, 6.5c and 6.5e provide the three samples of terrain mapping working at 300 MHz bandwidth with different driving behavior. The x-axis is the east side looking strip map SAR from the test vehicle. The y-axis is the azimuth axis which we drove from south to north. Figures 6.5b, 6.5d and 6.5f are the velocity interpolation versus time of Figures 6.5a, 6.5c and 6.5e, respectively. The mapping figures give the approximately the same targets indication. The three samples are working at different speed figures on the same track, thus convinced the results of how the SAR radar mapping on the terrain.

We tested the platform with 1 GHz on the accumulation interval of 0.339 seconds per RD response. Other settings are the same as the 300 MHz experiment introduced before. The comparison test of 1 GHz bandwidth results are shown in Fig. 6.6. In Fig. 6.6, the Figures 6.6a, 6.6c and 6.6e provide the three test samples of terrain mapping working at 1 GHz bandwidth. The x-axis is the east side looking strip map SAR from the test vehicle. The y-axis is the azimuth axis which we drove from south to north.

6.3.3 Discussions

For a typical 300 MHz terrain mapping (Fig. 6.5c), the car targets are shown in a point cloud map which caused by the EM waves travel from radar transmitter to the reflective surface of the car and come back to the receiver. It comes with a range of

$$R = \frac{cT}{2},$$

(6.2)

where $T$ is time of reflection of EM waves. Different cars reflect different peak power signals and thus provide a whole SAR terrain mapping. Except cars, other reflectors
Figure 6.5.: 300 MHz bandwidth SAR terrain mapping. (a)(c)(e) Using different velocity mapping from (b)(d)(f), respectively.

would also cause reflections like the trash can in the Fig. 6.5c. For example, the building targets at the azimuth of 60-80 meters and range of 40-50 meters are ob-
Figure 6.6.: 1 GHz bandwidth SAR terrain mapping. (a)(c)(e) Using different velocity mapping from (b)(d)(f), respectively.

served. The trash cans are shown at azimuth location of 80 meters and range location of 50 meters.
For a typical 1 GHz terrain mapping, with better range resolution, few car rows can be seen from the terrain mapping due to the limitation of range bins. However, the parking cars gap is hard to distinguish due to the azimuth resolution. In Fig. 6.6a, it is clear that the adjacent cars parked at two different rows are not easy to separate. Only shapes of cars can be observed. The differences between the 300 MHz and 1 GHz tests are the point cloud maps are more intensive with higher peaks. With different driving behaviors, the SAR terrain mapping results remain the same.
7. CONCLUSION

In this dissertation, a full workflow of radar signal processing is shown.

In Chapter 2, sensor fusion using mmWave radar and monocamera is achieved. The results are shown. The radar-camera fusion system is consisting fuse radar camera raw data with EKF for tracking multiple objects at the same time. The association is done with coordinates transformation and error bounds estimations. The noises result from sensors are correspondingly processed throughout separating covariance matrix to ensure the reliability of detections. The fusion system takes both sensors’ advantages and thus ensures error bounds minimizations from better resolutions on different perspectives from different sensors. The RMSEs of range, angle and doppler are shown in experiment results. Cross validating with camera’s vision and radar’s point clouds from raw data, the radar-camera fusion-EKF system provides a better and more reliable detections of targets. A low cost and reliable RTK sensor system is presented with detecting confidence.

The future work can use machine learning method to replace background subtraction algorithm to improve the vision extractions of targets. As well as use some classification algorithms like micro-doppler signatures from radar [133] to classify targets. The fusion-EKF provides a novel way to improve tracking and detecting reliability for radars in ADAS systems and autonomous driving.

In Chapter 3, MDS on human behavior monitoring is used for target classification and human behavior recognition by mmWave radars. A real-time human motion behavior monitoring system using mmWave radar is introduced. It can perform as a ward in an interested area to collect human motion behavior information. Future work can apply this application to automotive radars, which allows automotive radar
to classify car, pedestrian, and bicyclist. In addition, more human motion behavior can be added to the network for better recognition of the environment and the detection. The real-time human motion behavior monitoring system using mmWave radar can perform as a ward in interested area to collect human motion behavior information. The future work can apply this application to automotive radars, which allows automotive radar to classify car, pedestrian and bicyclist. In addition, more human motion behavior can be added to the network for a better recognition of the environment and the detection.

The Chapter 4 is supporting radar-camera sensor fusion for mmWave radar data clustering. The clustering algorithm REDBSCAN is shown with better detecting reliability on radar signals. In this chapter, the robust and adaptive REDBSCAN algorithm is mathematically presented and compared to conventional GMM and DBSCAN algorithms. The comparison shows REDBSCAN is not restricted by preset parameters and it can be fed with radar parameters. It is adaptive because it can work in different detecting scenarios like traffic monitoring or indoor people counting. It is robust because small and large targets like corner reflectors, cars and walls can be clustered correctly.

And additionally, MMWCSAR imaging system in Chapter 5 with compressed sensing is detailed discussed. With Doppler frequencies and inverse Radon transform, the 3D imaging can be reconstructed. A complete 3D circular SAR imaging radar is simulated and experiment tested. It can generate a high resolution 3D image by compressed sensing. The resolution and constraints of the MMWCSAR platform are discussed. We take a step further with the signal processing module of our system. We discussed the range calibration and 3D imaging reconstruction. In addition to using the IRT method, the involved CS matrix transformation of the original data to the final volume FOV image is shown.
The proposed MMWCSAR system is efficient, portable and fast for widespread use. The radar transceiver used in this design is more affordable than using an MIMO imaging or traditional SAR imaging. It avoids the large antenna, as well as the complex radar transceivers. A user just needs to move the radar along a circular track. A high resolution volume FOV figure can be extracted using our algorithms.

With help of IMU data, the mmWave radar’s ground stripmap SAR techniques are illustrated. In Chapter 6, a SAR terrain mapping for automotive mmWave radars. The whole SAR terrain mapping experiment comes out a positive effect. Our system is able to distinguish the static targets like cars, buildings and trash cans. Additionally, the single transceiver build is low-cost and easy to implement using mmWave radar. The point cloud map is achieved. However, the azimuth resolution is not able to distinguish the parking cars gap clearly. The future steps include improve the SAR radar terrain mapping azimuth resolutions on different range bins.

All of these are in fulfillment of my full radar signal processing research during my Ph.D. academic career.
MIMO array of mmWave radar emulates larger array than the physical array. The geometry of linear uniform array used in patch antenna of mmWave radar can be seen in Fig. A.1. The equation of solving MIMO array is:

\[ E(\theta) = E_0 \sum_{n=0}^{N-1} a_n \exp\left(j \frac{2\pi}{\lambda} nd \sin \theta \right) \] (A.1)

where

- \( E(\theta) \) : receiving voltage
- \( E_0 \) : transmitting voltage
- \( a_n \) : antenna weight
- \( \lambda \) : wavelength
- \( d \) : antenna separating width
- \( N \) : number of array antennas
- \( \theta \) : angle of arrival

When referring to radar’s range. The common used radar range equation is:

\[ P_r = P_t \frac{G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} \] (A.2)
where

\[ P_r : \text{receiving power} \]
\[ E_t : \text{transmitting power} \]
\[ G : \text{antenna gain} \]
\[ \sigma : \text{radar cross section} \]
\[ R : \text{range of scatter} \]

mmWave radar’s stretch processing is typically mixed receiving signal with transmitting signal. The output is mixture of time domain signals with frequencies and phases. When referring to fast-slow time signal, the deramping signal can be written as:

\[ r(t) = k \exp[j2\pi(\frac{2R}{c_0} \frac{BW}{PRF} + \frac{2vf_c}{c_0})t] + n(t) \quad (A.3) \]
where

\[ r(t) : \text{deramping signal} \]
\[ n(t) : \text{Gaussian noise} \]
\[ k : \text{power constant} \]
\[ c_0 : \text{EM wave speed} \]
\[ BW : \text{bandwidth} \]
\[ PRI : \text{pulse repetition interval} \]
\[ v : \text{Doppler velocity} \]
\[ f_c : \text{center frequency} \]

Two terms in the Equation (A.3) represent fast time samples and slow time samples, respectively. They are also referring to beat frequencies during stretch processing. Common method to solve range and Doppler frequencies is to use Fourier analysis.

These radar basics are used among chapters.
REFERENCES


[40] “Awr1642 evaluation module user guide.”


[46] “Yuv pixel formats.”


