A Dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Fusion of Optical Images with Radar Images Constructed using Pulsed Doppler Beamforming to Enhance Object Tracking

Creating a novel five layer image augmented with radial velocity and distance properties

Jiajia Shi

Autumn 2015

University of Technology, Sydney,
Faculty of Engineering and Information Technology
Center for Real Time Information Networks

Supervisor
Prof. Robin Braun

Co-supervisor
Dr. Bruce Moulton

Date of the graduation
TBA

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text. I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

My Contributions

Papers

- Braun, R., Shi, J.: Combining Computer Vision with 1 Dimensional Doppler Radar to provide a low cost robotic vehicle tracking system. In: Proceedings of the 1st Asia-Pacific Conference on Computer-Aided System Engineering, APCAST 2012, p.25
- Shi, J., Braun, R.: Crossed linear arrays using Doppler radar beam-forming for detecting single moving targets. In: Proceedings of the 2nd Asia-Pacific Conference on Computer-Aided System Engineering, APCASE 2014, p.118
- Shi, J., Braun, R.: Image construction using beam forming. In: Proceedings of the 2nd Asia- Pacific Conference on Computer-Aided System Engineering, APCASE 2014, p.121
- Shi, J., Braun, R.: "Crossed Linear Arrays Using Doppler Radar Beamforming for Detecting Single Moving Targets". Book series: Studies Computational Intelligence, Vol. 595, Grzegorz Borowik et al: Computational Intelligence and Efficiency in Engineering Systems, 978-3-319-15719-1, 334637
- Shi, J., Braun, R.: "Image Construction Using Low Cost Airborne Beamforming". Book series: Studies Computational Intelligence, Vol. 595, Grzegorz Borowik et al: Computational Intelligence and Efficiency in Engineering Systems, 978-3-319-15719-1, 334637

Patents

- Provisional Patent - Disclosure DISC-UTS-000433 Radar Enhanced Optical Images
- To be applied for - Sensor minimization using fusion of radar imaging and optical imaging

Contents

List of Figures

List of Tables

Abstract

Target tracking that detects the presence, color, contour, position, velocity, and acceleration is very important in intelligent systems. Single or multiple sensor fusion methods have been developed to improve the tracking quality. There are algorithms to fuse radar and optical images at a decision-making level. However, not many algorithms have been applied to radar image and optical image fusion. This thesis investigates the fusion of these two images. A new five-layer image format is proposed.

Beamforming, a radar signal processing method, has been applied to remote sensing. This suggests the idea that beamforming can be used to create images similar to optical images, with additional information such as range and radial velocity. The radar image can be fused with the optical image to enhance target tracking.

This thesis attempts to prove this proposition. First a tracking system with an optical imaging subsystem and a radar imaging subsystem was designed. Optical images were captured by a high-quality webcam, and radar images were captured by a system designed and built by the author. The radar subsystem is a short-range airborne acoustic radar which contains the transmitter and receiver board, FPGA board and a PC. All the channels need to work simultaneously. The data rate and computation load were very high when the number of sensors increased. For a low-cost development, the resources of FPGA are very limited. Hardware construction is central to this thesis.

Next was the development of a mathematical model of the range layer using pulsed beamforming. The performance with multiple targets was evaluated. The relationship between angle resolution and sensor number was investigated. Linear sensor array beamforming was extended to two dimensions to create the image. It was shown that the targets could be clearly extracted. The required resources increased linearly as resolution improved while on the contrary for a full rectangular array the required resources increased exponentially. Improved image resolution with fewer resources was investigated. Structures included two-line multiplication beamforming, crossed array beamforming, and other sparse structures. Their performances were tested with added noise. Results showed that the algorithms were robust. Sparse rectangular sensor arrays may have ambiguity. The pulsed mode was adopted to detect the distance. A matched filter was used to maximize SNR and minimize false alarms.

In addition to pulsed beamforming, Doppler Beamforming was adopted to provide radial velocity information, which is important for detecting moving objects.

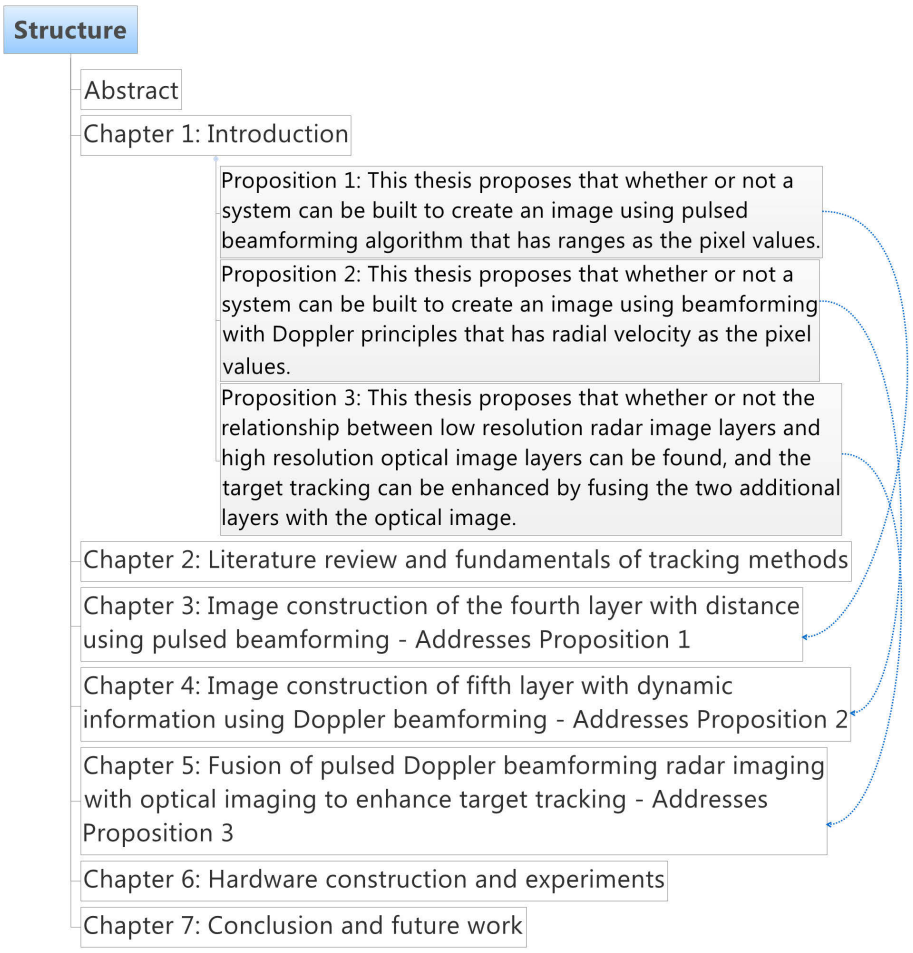
This can remove stationary "clutter" from the image. The performance of linear sensor array beamforming was analyzed for both single and multiple targets. Sensor spacing and angle of arrival were investigated. Then the beamforming was extended to two dimensions. Full rectangular array beamforming, two-line multiplication beamforming, crossed sensor array beamforming and sparse rectangular array beamforming and were studied. FFT filtering methods and filter bank methods were investigated. The processing algorithm and the example image were provided.

Five fusion scenarios for the proposed new image format are discussed.

- Scenario 1 is the fusion of video camera frames with Doppler beamforming radar to enhance moving object detection. The fusion of frame differencing with Doppler beamforming was studied.
- Scenario 2 is using imaging to reduce the ambiguity of crossed array beamforming. The template matching and frame differencing were used to process the region which may contain targets. Then a decision making strategy for reducing ambiguity was introduced.
- Scenario 3 is the fusion of optical imaging with Doppler beamforming radar to enhance tennis ball tracking. The region of interest was marked by the radar. The optical imaging was used to judge the in/out of the tennis ball.
- Scenario 4 is the fusion of optical imaging with pulsed beamforming radar to track a target in 3D position. The Kalman filter was studied and two fusion algorithms were researched.
- Scenario 5 is using Doppler beamforming radar to improve gait analysis. The micro-Doppler characteristics are used to analyze human movement. Then Doppler processing was combined with Doppler beamforming to relate the Doppler frequency to the body part, which can be projected into the optical image.

The experiments show that the creation of a range image layer and a radial velocity image layer are successful. The proposed two extra image layers can be combined with a traditional RGB image to enhance object tracking. The disadvantage is that the data rate is very high when the sensor number is increased. In two dimensional beamforming, there is ambiguity in some shapes of sensor arrays. A sparse rectangular array can be used to reduce the sensor number with fewer ambiguities. The ambiguities can be removed by fusion of different layers in the new image format. The new image format and the fusion algorithms can bring great value to the area of surveillance, sports, medical care and computer-human interaction.

1 Introduction



1.1 Background

1.1.1 Tracking system

A tracking system should be able to sense the environment and understand the meaning of the data, such as subtracting the targets from a background and fetching features of the targets [9, 10, 11]. This is like a mapping of the world into a set of much lower dimensionality. One type of sensor usually provides one aspect of the target [12].

Due to the diversity of environments, target characteristics and tracking purposes, many different types of sensors and signal processing algorithms have been developed. As different sensors can provide complementary information, various sensor fusion methods to enhance target tracking have been developed [12]. Prior research [13, 14, 15, 16, 17, 18, 19] suggested that performance of different fusion methods varies greatly in different applications. The goal of this research is to design a specified system with sensors and signal processing algorithms to enhance target tracking in some specified applications.

1.1.2 Benefits of sensor fusion

The advantage of using multiple sensors is as follows [20, 12]:

- Multiple sensors can provide redundant information about the same environmental characteristic. The redundant information can increase the robustness of the system when some sensors fail;
- Different sensors can provide complementary information;
- Data can be processed in parallel;
- Multiple sensors can reduce ambiguity;
- One sensor's information can be used to verify other sensor's information;
- Information can be obtained more accurately, as well as in less time and at a lower cost. With more timely information because more information is obtained, the system can reach a conclusion faster;
- Enable features which are not able to be perceived by individual sensors;
- Less costly information means a set of ordinary sensors can replace expensive sensors to fetch similar information.

1.1.3 Image fusion in tracking system

Image fusion is one type of sensor fusion. Without specification, an image usually refers to an optical image. In computer systems, an image is a two-dimensional

well-ordered scaled data matrix, which can reflect characteristics of real objects in the world. Each value in the matrix is a pixel. In this thesis, all two dimensional data matrices that can reflect real objects are called images.

1.2 Motivation of research

1.2.1 Using beamforming to create image layers

Beamforming is a radar signals processing algorithm which is widely used for the detection of targets in remote sensing applications. However, it is not widely used for the creation of images. In particular, optical imaging has become so widespread that other kinds of imaging, for other purposes do not gain much attention. In imaging, the creation of a two-dimensional representation of a "scene" can be understood in the field of view. This representation is a collection of values related to the scene at a particular location. In optical imaging, this is an ordered collection of values representing the light intensity being emitted or reflected from a portion of the scene. This is called a pixel.

The question arises as to whether or not measurements of factors other than light intensity would be useful. For example, if the imaging were to be part of a machine vision or robotic system, then the value of a pixel would be used to indicate the presence or absence of an object at that location. In such a case, a radar image may be more useful. The pixel in the radar image would indicate the intensity of the reflected microwave or acoustic signals at that location. There are many other factors about the object at the location of the pixel that may be very useful. In particular, it may be needed to know the radial velocity of that object or its distance from the receiver.

Optical imaging is not well adapted to provide that information. However, electromagnetic and acoustic signals reflected off the object have the potential to provide this kind of information. For example, Doppler radars can measure radial velocity.

Radar was developed as a means of observing an object in circumstances where it could not be seen. Perhaps it was too distant, or hidden behind clouds, or perhaps it was night time. It was not developed for the purposes of creating an image. All of the radar technology was optimized for detecting these remote objects. Even beam forming radar, which was developed so as to remove the need for a moving or rotating antenna was developed for this purpose.

In this thesis, new methods for using pulsed Doppler beamforming to create images were reviewed and described. By using radar technology, the images created can have new information, such as velocity and distance. Extensive work using ultrasonics has been done, and has been used as the basis of this thesis.

1.2.2 Fusion of radar and optical image

Both the optical imaging and radar imaging have similar steps: pre-processing, background subtraction, feature extraction and decision making. As they both detect the same object, the sensory information should have some relationships that can improve the performance of both.

Optical images can easily be understood by people. People can identify most of the targets by analyzing light intensity and color. However, in machine learning systems, some images, which are easily understood by humans, are very difficult for a computer to process. Also, optical images have these limitations:

1. Optical images highly depend on illumination, neither too high nor too low;
2. The horizontal and vertical distances and velocities can easily be measured by optical images, but radial distances and velocities can hardly be measured;
3. Optical images usually sample less than 30 frames per second, so fast moving objects cannot be detected;
4. Optical images cannot detect objects behind obstructions;
5. Optical imaging can easily be affected by shadows, backgrounds and the weather; and
6. Optical imaging is passive, which senses the environment by collecting illumination of the scene.

Compared with cameras which passively receive environmental light, radars actively transmit electromagnetic waves or acoustic waves to the environment and analyze the returned signals. Therefore, radars have many different characteristics. For instance:

1. Radars can sense the radial distance and velocity;
2. Radars have much lower resolution than cameras; and
3. Radars are not affected by the illumination changing.

According to these characteristics, the strengths of optical imaging and radar are complementary. There is much research on fusing radars with optical images at a decision making level. In the research, radar and optical imaging were processed separately and the results were collected for making a decision. As both cameras and radar can create images, this thesis examines the fusion of the radar imaging and optical imaging.

1.2.3 New image format

For the benefit of fusing the optical imaging and radar imaging, a new image format has been proposed. Radar can provide both range and velocity information. This

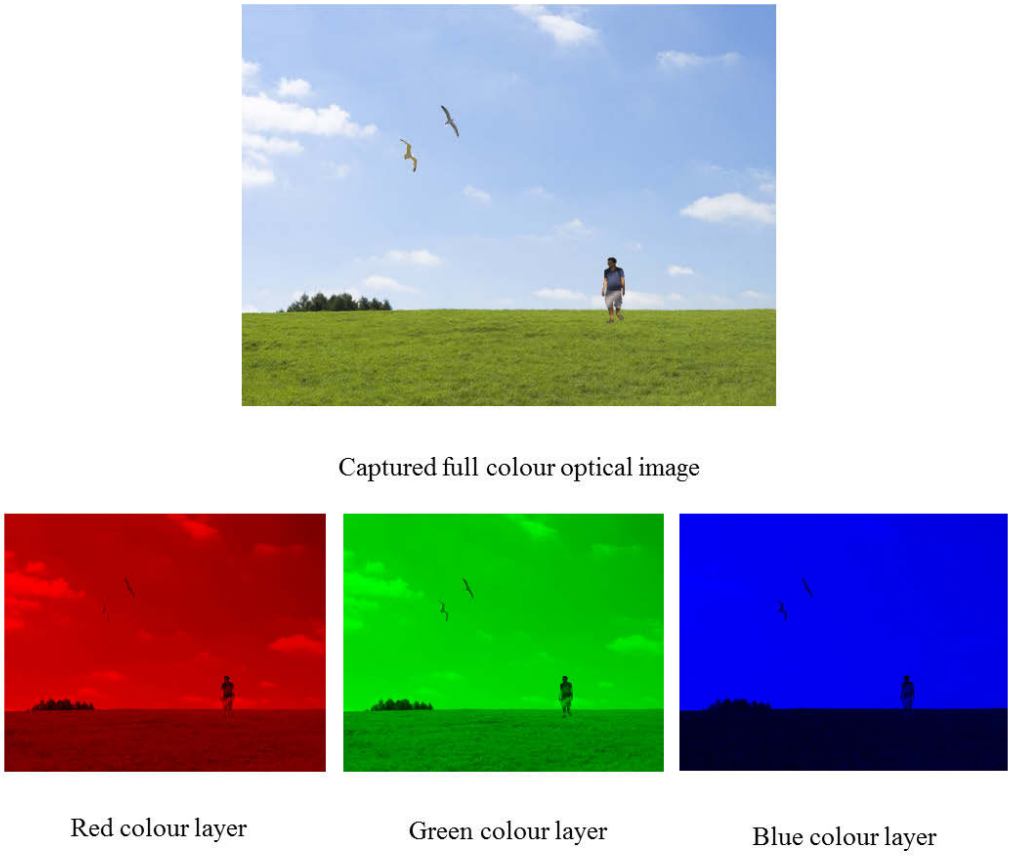


Figure 1.1: Example of RGB image layers

thesis studied the way to create hardware which can both work on range detection and radial velocity detection.

A typical full color image is constructed of red, green and blue light layers. Figure 1.1 is an example of an optical image. In the image, a person is walking on the grass towards the observer, and two birds are flying freely. These three objects can be considered as the foreground. Other objects can be considered as the background.

Figure 1.2 shows the proposed radar image. The radar used is a vision range airborne acoustic pulsed Doppler beamforming radar. In the radar image, two

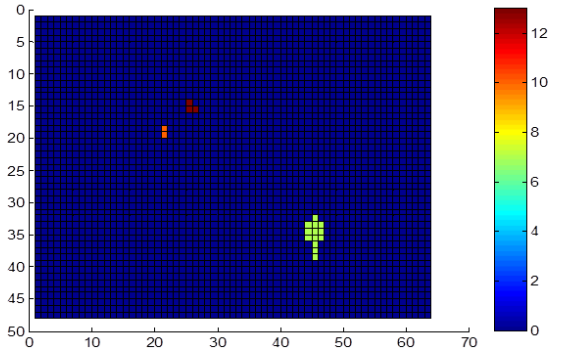
layers were created. The first layer is the distance layer. Only the subtracted targets have the values and the distances are obtained by a pulsed radar. The second layer is the radial velocity layer. Only moving objects can be subtracted and velocities can be obtained by the Doppler effect. To store radial distances and velocities with the light intensity values which range from 0 to 255, the radial distance and velocity values are scaled to integers and the scale can be adjusted by applications.

The light intensity in different colors, the object range and the object velocity reflect the features of a target. Both optical imaging and radar imaging have been used to detect the targets. This leads to the idea in this thesis of creating a new five-layer image format. The radar images usually have lower resolution than optical images. All the image data are stored in the same memory. The fusion method can be developed based on the new image format data. Also, if necessary, more layers can be developed; these extra image layers can improve the performance of the current optical imaging. Figure 1.3 shows the proposed five layer image.

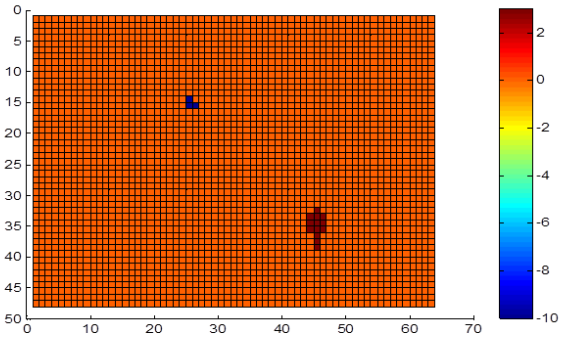
1.3 Applications

The proposed imaging method can be applied in these areas:

1. Ball and player tracking in sports: In some ball games such as tennis, the judgment system needs to track the ball and make an in/out decision. The velocity of a ball is usually too fast for optical imaging. In golf or soccer, the tracking of a ball trajectory can make the game more interesting and help the players to improve their performance. Also, the fusion of optical and radar imaging can enable gesture detection of players, and the results can be used to track the athlete's state and set up a training method for them. Correcting the gesture can improve the athlete's performance.
2. Human-machine interaction. Gesture analysis can help a computer to understand human activity meaning. This can make the games more entertaining. Microsoft has already done work on this, such as the Kinect. Image processing on its own is not sufficient for understanding people.
3. Medical care. Much research has been done on how to use radars or optical images to detect disease in human gestures.
4. Surveillance: In some environments such as banks, airports and military establishments, it is necessary to identify a potential threat no matter what time it is. Currently, the work is usually done by guards inspecting a video or some other manual operation. Automatic detection is not reliable.



Pulsed beamforming image



Doppler beamforming image

Figure 1.2: Example of radar image layers

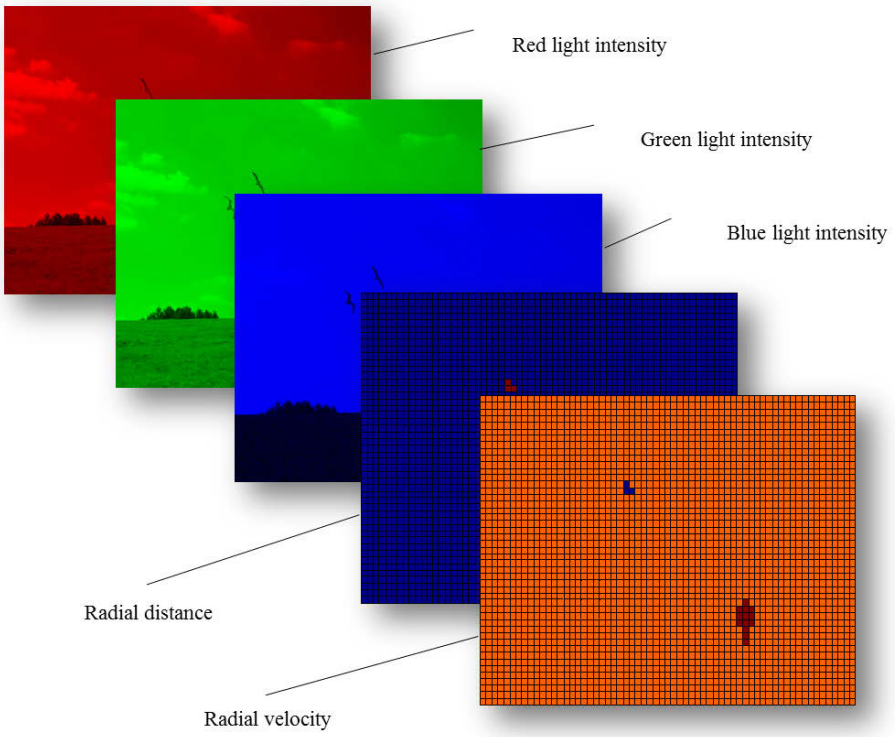


Figure 1.3: Proposed five image layers

1.4 Research scope

This research is concerned with the building of a robust and cost-effective vision-range airborne acoustic radar, and with the investigation of the method to fuse the captured pulsed beamforming and Doppler beamforming radar images with optical images. The scope of this study is in the following areas:

1. the investigation, application, and improvement of currently available vision range radar imaging methods for object tracking,
2. the investigation of various advanced digital signal processing techniques in radars in order to estimate the presence, range, velocity and Angle of Arrival (AoA),
3. the investigation of optical imaging for object tracking,
4. the numerical and experimental verification of the selected methods on the one-dimensional linear sensor array pulsed beamforming and Doppler beamforming,
5. the numerical verification of various types of two dimensional pulsed beamforming and Doppler beamforming,
6. the hardware construction for the verification of the idea, such as to verify the attenuation, accuracy, data rate, signal to noise ratio and algorithm performance and
7. the fusion algorithms of the proposed two-layer radar imaging with standard optical imaging. The algorithms were specially designed based on the applications.

1.5 Summary of contributions

The contributions of this study can be divided into these categories:

Preliminary contribution - System design and hardware construction

The structure of the system contains an optical imaging subsystem and a radar imaging subsystem. The optical imaging sub-system contains a standard web-cam. As radar systems vary with large differences according to environments, in this study, the radar is a low cost vision range airborne acoustic radar. The first step was to build such a radar that can be fused with optical imaging.

The author designed the radar system. It includes the circuit design and the Field-Programmable Gate Array (FPGA) programming. The Printed Circuit Board

(PCB) includes the transmitters, receivers, amplifiers, ADCs and DACs. The primary function is to transmit the signals generated by the FPGA and to receive the returned signals. The received signals were converted to digital signals and transmitted to the FPGA through a Serial Peripheral Interface (SPI). The prototype was designed, tested and improved. One difficulty was to choose appropriate transmitters and receivers. The sensor spacing between each receiver is $\lambda/2$ to ensure there were no ambiguities, but most transducers are larger than this size. MEMS ultrasonic microphones were selected. The sensitivities were tested. Then the PCB boards were designed and built. The difficulties were how to convert and limit the returned signal to the proper range and how to improve the Signal to Noise Ratio (SNR). The ADC needed to be as close as possible to the receivers and all the noise from the chips and power lines had to be limited to a small level. The sub-modules in the PCB should not interfere with each other. Testing and redesigning the radar PCB board design helped to fulfill the requirements.

The functions of the FPGA included:

1. Controlling the ADCs via a SPI,
2. Converting serial data to parallel data,
3. Controlling Static Random Access Memory (SRAM),
4. Controlling Dynamic Random Access Memory (DRAM), and
5. Controlling Flash RAM to store the data.
6. Pre-processing of the data,
7. Generation of the transmitting signals,
8. Setting up a Nios II CPU to control the system, and
9. Transmission of data to the PC through the USB.

The benefits of the FPGA is that all the channels of received data could be processed in parallel. No matter how many receivers were adopted in the system, the processing speed was not affected by the sensor number.

The main point of the design was

1. To increase the robustness of the system,
2. To optimize the resource due to the large data rate and limited FPGA resources, and
3. To check the time sequence to make the system work in real time without errors.

The radar sub system had been shown to be working properly for the system. The limitation of the system was that the data rate was too high for USB 2.0 to transmit all the data to the PC in real time. Two solutions were proposed to solve this problem. The first method was to finalize the algorithm and process most

the data inside the FPGA. However the algorithms also consume considerable resources. The second method was to use USB 3.0 to transmit the data. This will become the aim of future work.

Theoretical study on Pulsed Beamforming

The hardware design, the experiments and the theoretical study on beamforming have been done. The investigation of the difference between the time domain and frequency domain processing gave rise to an algorithm being proposed. The resolution, when there are multiple targets, was simulated. In two-dimensional beamforming, various sensor array structures were studied.

The performance of the full rectangular array beamforming, two-line multiplication beamforming, crossed array beamforming and sparse rectangular array beamforming was explored. The relationship between resolution and ambiguities was investigated.

Compared with a full rectangular array, a sparse rectangular array had ambiguity. By carefully arranging the array shape the ambiguities were reduced to a low level which allowed them to be distinguished from the true targets. Also, the pulsed mode of beamforming was investigated.

The relationship between the generated image and the optical image was investigated. In simulation, the new image layer with distance information was proposed.

Theoretical study on Doppler Beamforming

In addition to using pulsed beamforming radar to construct an image with distance information, another contribution of this research is to create another image layer with radial velocity information. The method is similar to the pulsed beamforming image. The method is to subtract and analyze the Doppler beamforming signals. Two-dimensional Doppler beamforming was also investigated.

Objects with different Doppler frequencies need to be separated first. FFT filters and FIR filter banks were investigated and compared.

To reduce the sensor number, the performance of the two-line multiplicative beamforming, the crossed array beamforming and other types of sparse rectangular array beamforming were investigated. The simulation showed that the new image layer with velocity information could be created.

A new five layers image format

This is based on the two proposed radar image layers. A novel five layer image format which contains traditional optical red, green and blue, and two extra radar

layers with distance and radial velocity was proposed. In the future, much more sensory information will be added into the traditional optical image to enhance the image processing. This image format can be a standard format in some applications.

Scenarios on the fusion of the proposed radar system with optical imaging

Based on the proposed image format, the last contribution of this research was to apply these image layers in a practical environment. In detail, it has been demonstrated that the new image layers can be used in these applications. The full rectangular array Doppler beamforming image can be fused with the optical image to enhance moving object detection. The crossed array Doppler beamforming image can be fused with the optical image to enhance golf ball tracking, and ambiguity can be verified and removed by the optical image. The full rectangular array Doppler beamforming image can be fused with the optical image to enhance tennis ball tracking. The full rectangular array pulsed beamforming image can be fused with the optical image using a Kalman filter to increase the moving object tracking in a three dimensional position. Finally, the Doppler beamforming radar was used to improve gait analysis, which can locate the position of the frequency and fuse with optical images.

1.6 Design constraints

The low cost radar imaging and optical imaging fusing system should have these properties:

1. Real-time: In a surveillance application, the system needs to track the target in real-time, the signals need to be processed continuously with minimum delay and alarms need to be reported simultaneously. In the sports scenario, the detection should always be able to follow the ball or the players.
2. Portable: Traditional radars used in military detection and weather prediction are usually not portable. To extend the application area to daily life, the designed radar should be able to take everywhere and be fixed easily.
3. Low cost: A radar system is usually expensive. A reliable real time optical imaging system is also expensive. The cost of the whole system should be limited to an affordable level, such as 1,000 Australian Dollars.
4. Simple structure: The system should be as simple as possible. The system with a simple structure usually has a high stability and a low cost.
5. Low computation load and less memory consumption: Both radar signal processing and optical imaging produce a high computation load and much

memory resource. By fusing radar imaging and optical imaging, the features which cannot easily be extracted by one sensor, may be extracted by another sensor with a low computation load and few memory resources.

6. High accuracy: This is the most important indicator for tracking systems. In most cases, tracking systems are designed for specific applications, and the system needs to be adjustable.
7. Autonomous: The system must be able to work independently of the operators. Machine learning algorithms need to be applied to make the judgments.

1.7 Research question, propositions and approach

The thesis does not follow the classical "science" model of a Hypothesis leading to a Methodology, leading to Experimentation that then proves or disproves the Hypothesis. Its contributions are rather embedded in its propositions that are given substance through emulation, simulation and experimentation. That is a more appropriate model for an engineering Ph.D. that describes the creation of a new product or technology.

In particular, the thesis is all about the "fusion" of the capabilities of very different remote sensing modalities. This fusion creates a new modality that has capabilities that are greater than the sum of the two constituent parts.

1.7.1 Research question

Given that the two types of sensor, optical imaging and pulsed Doppler Beamforming radar imaging, have very different properties, and that a single sensor's performance is not reliable and the cost is very high, then:

this research hypothesizes that their respective strengths can be fused to gain a big advantage in target tracking at a low cost.

1.7.2 Research propositions

- Given that an enhanced image with additional information would be very useful in robotic applications;
- given that optical imaging has become very inexpensive and effective;
- given that radar imaging can provide very different capabilities; and
- given that an acoustic radar image system is inexpensive and possible, and digital processing speed is high enough to process data in real time,

This research proposes that:

- **Proposition 1:** whether or not a system can be built to create an image using a pulsed beamforming algorithm that has ranges as the pixel values;
- **Proposition 2:** whether or not a system can be built to create an image using beamforming with Doppler principles that has radial velocities as the pixel values; and
- **Proposition 3:** whether or not the relationship between low resolution radar image layers and high resolution optical image layers can be found, and the target tracking can be enhanced by fusing the two additional layers with the optical image.

The whole thesis is structured around these propositions. Chapter 3 is to solve **Proposition 1**, chapter 4 is to solve **Proposition 2**, and chapter 5 is to solve **Proposition 3**.

1.7.3 Research approach

Step 1: Literature survey and solution proposed

Many algorithms have been used in object tracking and identifying object characteristics by vision and radar separately. These two methods contain several steps, usually known as pre-processing, segmentation, feature extraction and classification. In each step, the algorithms are similar both from the vision and the radar points of view. By comparing the characteristics of each algorithm, fusion methods can be proposed.

Step 2: Creation of a physical environment

The research environment is set up to test the performance. The whole system should contain a controller (computer or integrated board), one web-cam and an acoustic pulsed Doppler beamforming radar. Specifically, the acoustic pulsed Doppler beamforming radar should be designed and built by the researcher. The frequency, pulse width and voltage can be adjusted to get the best performance. Also, all the signals need to be sampled at a high sample rate. To improve the Signal to Noise Ratio (SNR), all the chips are low noise chips and the receivers have ADC chips very close to them. The control of the whole system and the communication with the computer is done by using an FPGA. Verilog was used to program the FPGA. The final PCB board is designed by Altium Protel DXP.

Step 3: Theoretical study and simulation

The theory should be researched in mathematics first. The proposed algorithm should be tested by simulation software such as Matlab.

Step 4: Experimental verification

The optical imaging sub-system and radar sub-system needed to be tested separately. Then the whole system needed to be tested in a real environment. If the performance is good, the algorithm can be integrated into the FPGA, which makes the system much smaller and easy to use.

1.8 Outline of thesis

This thesis contains seven main chapters:

Chapter 1 presents an introduction to the work, the motivation of research, the scope of the work and the contribution to knowledge. Three propositions were made and investigated in the following three Chapters.

Chapter 2 contains a literature review and the fundamentals of various tracking methods. First, images created by different types of sensors such as optical images, synthetic aperture radar, synthetic aperture sonar, acoustic camera and medical ultrasonography were investigated. The characteristics of these types of imaging were compared. Further, the principles of pulsed radar and Doppler radar were investigated. Also different types of radars which are used in tracking systems were studied. Examples are acoustic pulsed radar, electromagnetic Doppler radar and acoustic Doppler radar. The advantage and disadvantage of these methods were studied. Third, optical imaging methods and procedures in tracking systems were researched. Then sensor fusion methods, such as fusion layers were investigated. Finally methods to evaluate the system were introduced.

Chapter 3 investigates the construction and mathematical explanations of the fourth layer with distance using pulsed beamforming. This chapter aimed to solve **Proposition 1**. Specifically, the system structure was first described. Then the medium constraints such as velocity of sound, attenuation of ultrasonic waves and carrier frequency were studied. The far field assumption and narrow bandwidth assumption were investigated. With these two assumptions, the calculation was simplified without much error. Then the linear array beamforming was researched. To improve the main lobe to side lobe ratio, window functions were applied in beamforming. Further, the performance when there were multiple targets was compared. The relationship between angle resolution and sensor numbers was investigated. Then the beamforming was extended to two dimensions to construct images. The mathematics of two-dimensional beamforming was investigated and a simulation was created. The output was quite clear, and the targets were clearly extracted. A full rectangular array of sensors consume a considerable amount of resources such as hardware, computation load and memory. In order to increase the image resolution, some methods that can achieve similar results with fewer sensors were investigated. The structure included two-line multiplication beamforming, crossed array beamforming and other structures of two-dimensional beamforming.

The performance was also tested with added noise, and the results showed the algorithm was quite robust. Sparse rectangular sensor array beamforming may have ambiguities. Some solutions were proposed to solve this problem, which were further investigated in chapter 5. The next step was to detect the distance for beamforming. The triangular method was investigated both in one dimension and in two dimensions. Then the pulsed mode was adopted to detect the distance. The matched filter was also used to verify the signals and remove false alarms. Before fusion with optical images, the characteristics of the radar image, such as frame rate, resolution and coordinate were studied. To give an example of this kind of image, an image was manually created to demonstrate the principle. The final processing algorithms and example image were provided.

Chapter 4 investigates the construction of the fifth layer with dynamic information using Doppler beamforming. This chapter aims to address **Proposition 2**. First the benefits of Doppler processing and classification of Doppler radar were studied. Then the signal subtraction methods were investigated. A beamforming algorithm was applied for the subtracted Doppler signals. Specifically, frequency domain Doppler beamforming was studied. Then the performance of linear sensor array beamforming was analyzed when there were single or multiple targets. Then the sensor spacing and angle of direction were studied. To improve the performance, the zero frequency components needed to be removed from the subtracted Doppler signals. Then the beamforming was extended to two dimensions. Two line multiplication was used to distinguish different moving objects. Then crossed sensor array beamforming and full rectangular array beamforming were studied. Further, the creation of Doppler beamforming layers was discussed. Also, the FFT filtering method and the FIR filter bank were discussed. The final example of processing algorithms and example images were provided.

Chapter 5 describes five examples of the fusion of the proposed five image layers which aim to solve **Proposition 3**. First, image registration and radar and camera calibration were discussed.

- The scenario 1 is the fusion of video camera frames with Doppler beamforming radar to enhance moving object detection. The frame differencing and Doppler beamforming were studied separately. Then the fusion of these two methods was studied. The experiment and simulation were provided.
- The scenario 2 is using imaging to reduce the ambiguity of the crossed array beamforming. The limitation of control board computation load was studied and the template matching and frame differencing were researched. Then the decision making strategy and fusion for reducing ambiguity in crossed array Doppler beamforming radar were introduced.
- The scenario 3 is the fusion of optical imaging with Doppler beamforming radar to enhance tennis ball tracking. First, the Hawk-eye system was researched. Then the data fusion system and region of interest were discussed. Also, the image processing sub-system were discussed.

- The scenario 4 is the fusion of optical imaging and beamforming radar to track a target in 3D positions. The Kalman filter was studied and two fusion algorithms were researched.
- The scenario 5 is using Doppler beamforming radar to improve gait analysis. The micro-Doppler characteristics were used to analyze human movements. Then the Doppler processing was combined with Doppler beamforming to relate the Doppler frequency to the body part, which can be projected into optical images.

Chapter 6 describes the hardware construction. First the system structure was described. Then the ultrasonic transducers were introduced. Also the controller and programing were described. The construction of the radar image by pulsed Doppler beamforming radar was shown and the procedure was described. The difficulty was in the conflicts between image resolution and the FPGA resource. To solve this problem, methods to reduce the sensor number were investigated. The results show that the sensor number can be reduced while increasing the ambiguities of the target. Another conflict comes from the detection distance and beamforming signal to noise ratio. In frequency domain beamforming, if the FFT length was increased, the resolution improved.

Chapter 7 summarizes the work presented in this thesis and describes the limitations of the device and future work.

2 Literature review and fundamentals of tracking methods

2.1 Introduction

In this section, imaging from different types of sensors are first investigated and compared. Understanding of the principles of existing imaging methods and comparing the advantages and disadvantages in different application can help to build and optimize the design.

The hardware to capture optical images is quite mature. The format of the output is well standardized. In terms of radar, the situation is very different. As there is no standard radar hardware that can fit all applications, every researcher builds or configures their own radars. Therefore in this chapter, radars for different applications and the standard image processing methods for the applications are investigated and studied.

The fusion layers are further researched and some methods used in sensor fusion and image fusion are studied. Ultimately, the evaluation methods used in the proposed system are described.

2.2 Images created by different types of sensors

2.2.1 Optical imaging

An image usually refers to an optical image, which is a two dimensional artifact to depict visual perception. Digital cameras are the most common devices to capture images. Digital cameras capture data and use Analogue to Digital Convertors (ADCs) to convert and store the light intensity into a two dimensional well ordered scaled data matrix. This kind of data matrix can be directly processed by computers and other micro-controllers.

In color images, red, green and blue (RGB) light is captured separately. In digital images, a pixel is the combination of RGB values in a particular location. Then the ordered collection of two dimensional values are input to the memory of the processing unit and stored as an image file.

2.2.2 Synthetic aperture radar

Synthetic Aperture Radar (SAR) [21, 22, 23, 24, 25] is used to create long distance radar images. A SAR is usually carried by a moving object such as a plane and senses the surface of the environment. On the other hand, an Inverse Synthetic aperture radar (ISAR) is usually fixed and scans rotating objects. A SAR detects targets by comparing the motion between the radar and the targets. The combination of returned signal can be considered as a long antenna which can obtain a high spatial resolution. The data are processed by the Fast Fourier Transform (FFT). A very large computation load is involved and real-time processing is a big challenge. Usually, data are processed after capturing.

2.2.3 Synthetic aperture sonar

In the ocean, visible light transmission is weak, while sound has a higher propagation velocity and lower attenuation in water than in air. Therefore, sonar is often used in underwater environments to detect vehicles or fish, and to draw a picture of the bottom of the ocean. The principle is very similar to radar, but at a much lower frequency.

2.2.4 Acoustic camera

An acoustic camera is a sonar imaging system to locate the source of sound [26]. Usually, a linear phased microphone array is used to simultaneously sample the received signals. Two dimensional microphone arrays have been applied in fishing acoustics. To achieve higher resolution, higher frequencies are used which reduce the ranging to tens of meters. The results are very useful to study fish behavior.

There is also some research in Germany [27] using airborne acoustic cameras to scan the interior of cars. They also provide a customized design of a sensor array. The device can be applied to sound analysis, free field measurements and interior wind turbines.

The difference between acoustic cameras and the work of this thesis is that acoustic cameras do not transmit signals. They passively receive signals and locate the source of the sound.

2.2.5 Medical ultrasonography

Acoustic radars have been widely applied in medical diagnosis to scan body structures. Piezoelectric transducers or capacitive micromachined transducers are laid in a line which generate an ultrasonic frequency between 2-18 MHz.

The direction of the sound is usually controlled by beamforming. The received echo is digitized to determine:

Table 2.1: Comparison of different types of imaging

	Medium	Emit Signals	Waves	Cost	Distance
Optical image	Air	Passive	Light	Low	Medium
Synthetic aperture radar	Air	Positive	Radio wave	High	Far
Synthetic aperture sonar	Air	Positive	Sound	High	Far
Acoustic camera	Ocean	Passive	Sound	Medium	Short
Medical ultrasonography	Human body	Positive	Sound	Medium	Very short

1. The time gap between when the sound is transmitted and received,
2. The focal length for the phased array,
3. How strong the echo is.

Then an image is created from this information. The Doppler effect detection can also be included in some applications to provide a colored ultrasonic image.

2.2.6 Comparison of different types of imaging

The comparison of different types of imaging is shown in Table 2.1.

2.3 Principles of radar

A radar emits and receives waves to detect the presence, angle, distance and velocity of objects [28].

By the difference in carrier, radar detection methods can be divided into electromagnetic wave based and acoustic wave based. By the difference in radar working mode, they can be divided into these modes: pulsed radar, Doppler radar, pulsed Doppler radar, beamforming radar, SAR radar and step radar (which sends signals at periodically increasing frequencies - infinitely small steps implies “chirp radar”).

The beamforming radar was investigated in detail in chapter 3 and in chapter 4.

2.3.1 Principles of pulsed radar

Pulsed radars measure the Time of Flight (ToF), which is the time the signal takes for the reflection to return [28]. The distance is half of the time multiplied by the

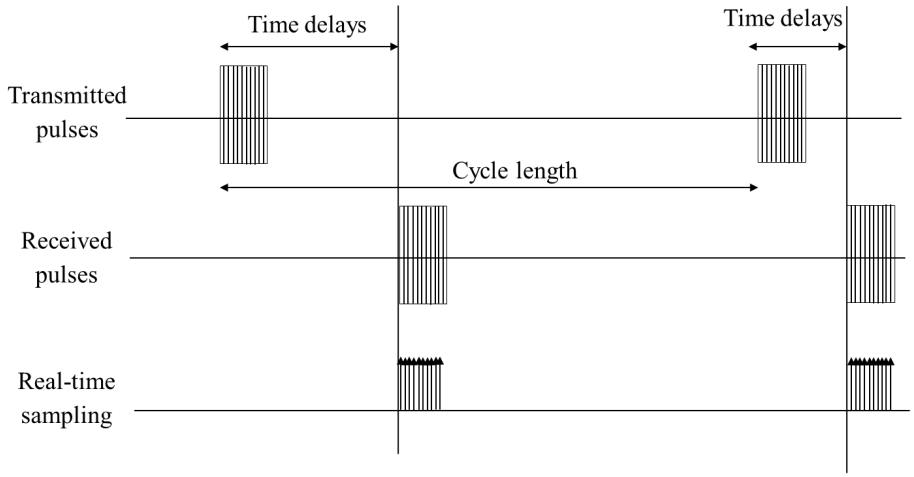


Figure 2.1: Principle of Pulsed radar [1]

velocity of the signal. The longer the gap between pulses is, the longer distance the pulsed radar can measure. The velocity of an electromagnetic wave is much faster than the velocity of an acoustic wave; the attenuation of an electromagnetic wave is much lower than the attenuation of an acoustic wave; therefore the detection range of an electromagnetic pulsed radar is longer than an acoustic pulsed radar. Figure 2.1 shows the block diagram of a pulsed radar.

The distance is given as follows:

$$D = v \cdot t_{ToF}/2 \tag{2.1}$$

where t_{ToF} is the time taken from the start to the return of the signal, and v is the velocity of the sound or radio wave in the medium.

2.3.2 Principles of Doppler radar

Doppler radar uses the Doppler effect to detect moving objects. The theory of the Doppler effect was proposed by Christian Doppler. The theory states that

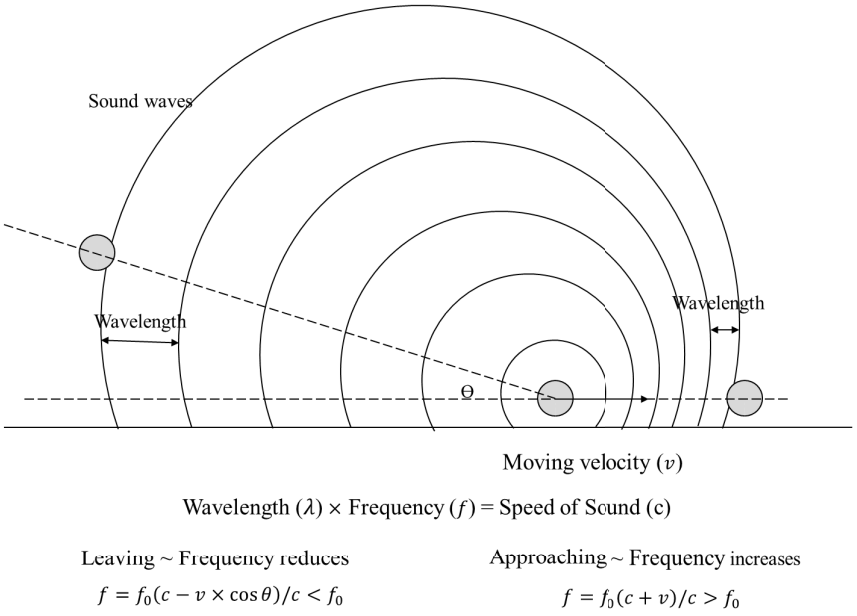


Figure 2.2: Doppler effects

the received frequency is higher than the emitted frequency when the transmitter approaches the receiver, and lower than the emitted frequency when the transmitter recedes from the receiver [29, 30, 31]. The effect exists both in acoustic and electromagnetic waves.

The Doppler shift frequency is given as follows:

$$f_d = f_0 \times \frac{v}{c} \cos\theta \tag{2.2}$$

where f_0 is the transmitter frequency, c is the velocity of the carrier, v is the velocity of the target, θ is the angle between the beam and the target's path.

If $\theta = 90°$, there is no Doppler shift, while if $\theta = 0°$, there is a Doppler shift and $f_d = f_0 \times v/c$, which gives the maximum Doppler shift.

Figure 2.2 is the diagram of the Doppler effects.

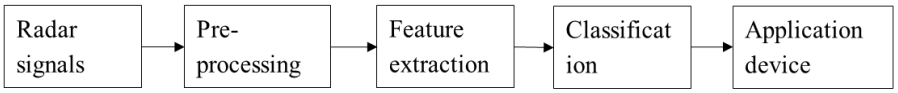


Figure 2.3: Procedure of radar signal processing

2.4 Radar used in tracking system

Prior research used both electromagnetic radar and acoustic radar to track targets. The procedure of radar signal processing is shown in Figure 2.3.

2.4.1 Acoustic pulsed radar

In the last decades of the 20th Century, some researchers investigated how to use radars to recognize moving objects. Weir and Childress. [32] found a method that used ultrasound ranging to measure the distance between a base unit and a transponder worn by the subject at the approximate body center. The computer connected to the base unit recorded the real time data of the distance and calculated the horizontal velocity which can be considered as a pictorial representation of that person's gesture. This could be used in clinical gesture analysis to find out whether or not the person would fall down and surgery was needed to correct the gesture.

Sabatini ajd Colla [33] proposed a method for sonar based recognition of walking people. The method of sensory information processing implemented in the device is a blend of grid based mapping and a wavelet based multiple resolution analysis. The wavelet based multiple resolution analysis allows the detection of the features that were peculiar to walking people in the range measurement sequence extracted from a single sonar sensor.

Karhik et al. [34, 35] found that the motion of some disabled or injured people who use crutches to walk can be treated as four-legged locomotion. In this way the gesture pattern can be described as a sequence of the states displaying the contacts of both feet and both crutches with the ground. The unstable states were defined as passive phases where the body is only supported in two contact points and where the center of the body is gravity and inertia driven in the direction of progression.

Peremans et al. [36] provided a design and a calculation of a high-resolution sensor based on tri-rural perception. It showed that within limits the sensor system

can discriminate between different types of reflectors, based on their radius of curvature. Also, a noise model, explaining the measured variations of the arrival times, was used to derive limits on the resolution of the results provided by the sensor.

2.4.2 Electromagnetic Doppler radar

Hewish et al. [37] and Nebabin et al. [38] distinguished humans and different kinds of vehicles in battlefields by using the Doppler effect induced by radial movements and the results showed that the false alarm rate (FAR) of their methods were 20%.

Geisheimer et al. [39] used the fully coherent, continuous-wave (CW) radar operating near 10.5 GHz to record the radar signature corresponding to the walking human. The received signal is the sum of Doppler shifted signals reflected from the various parts of the moving body. By the short-time Fourier transform and the chirplet transform, various parameters of the human gesture can be extracted from the signals.

Van Dorp et al. [40] used a kinematic walking model developed by Thalmann [41]. Van Dorp [40] proposed a model-based approach of estimating human walking parameters from radar measurements and visualizing the results as a human walking scene in virtual reality.

Sevgi et al. [42] used single-channel, airborne, synthetic aperture radar (SAR) for human target detection and identification. Human targets were differentiated from other detected slow-moving targets by analyzing the spectrogram of each potential target. The human spectrogram is unique, and can be used not just to identify targets as human, but also to determine features about the human target being observed, such as size, gender, action, and speed.

Otero [43] proposed a continuous wave radar for human gesture recognition. Compared with Geisheimer [39] who used a continuous chirplet transform, they found a new FFT method. The frequency and harmonic content of the individual moving body components were displayed on the horizontal or x-axis, their corresponding velocities on the vertical or y-axis, and their radar cross-section (RCS) on the intensity scale or z-axis.

The School of Electrical and Computer Engineering of Georgia Tech Research Institute had done much work on detecting humans based on radar. Gurbuz et al. [44, 45] had the non-linear phase approximation of a real human body by a Thalmann kinematic walking model [41]. They compared the performance of the FFT-based matched filter with the optimized nonlinear and linear phase detectors.

Researchers in the University of Pennsylvania used through-wall noise-like radar to detect human activities [46, 47]. Instead of traditional methods which use FFT or STFT to detect the Doppler shift, the Hilbert-Huang transform was proposed. Hilbert-Huang transform [48] is a signal analysis technique based on the combination of the Hilbert spectral analysis and empirical mode decomposition.

2.4.3 Acoustic Doppler radar

Kalgaonkar [49] proposed an acoustic Doppler Sensor (ADS) based technique for human recognition. The device was built at very low cost using off-the-shelf 40kHz acoustic devices and low-frequency audio-range sampling. They showed that the proposed mechanism can provide highly accurate human recognition. They used a 36kHz oscillator to mix the signal and used a sound card to capture the signal. In the signal processing, they used a FM demodulation to extract the Doppler signal and used a Bayesian classifier to classify the signal. The results showed that this method determined correctly well over 90% and the accuracy could be improved to 100% if they coupled classification readings with reflected energy contours.

Zhang [30, 31] also developed an acoustic continuous micro Doppler-radar to detect the gesture signatures of both human and four-legged animals in indoor and outdoor environments. To compare with the real situation, an accelerometer attached to the leg was used. Short-time Fourier transform was performed using MATLAB with a 50ms Hamming window and 1/2 overlap between adjacent transform windows. The results suggested that each person had a unique signature.

Georgiou [50] proposed a design and a collection of a multi-model data corpus for cognitive acoustic scene analysis. In the system, people walked or stayed stationary wearing speakers to let a microphone detect their position. 3D and 2D cameras were also used for system calibration. They tested the system in an anechoic chamber, a large classroom and an outside courtyard. The results showed that the localization cues were recorded properly: one could perceive the sound source moving around.

2.5 Optical imaging in tracking system

Digital cameras can sample the light intensity and store the data in a two-dimensional array. The scaled values of array elements, which are integers, are called pixels. Common image processing technologies include image coding, image enhancement, image restoration, image segmentation and image analysis. A series of continuous image frames can construct a video. By analyzing each frame, dynamic information of the objects can be obtained. The procedure of imaging is shown in Figure 2.4.

2.5.1 Image pre-processing

The purpose of image pre-processing is to reduce noise. There are several types of typical noise in optical imaging:

1. The Gaussian noise which comes from electric circuits, low intensity of illumination and high temperature changes of sensors.

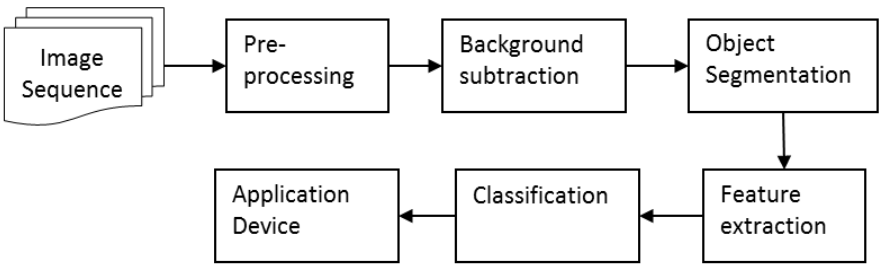


Figure 2.4: Procedure of optical imaging in tracking

2. The salt and pepper noise which looks like randomly distributed pepper and salt powder particles in images. The noise is mainly caused by the image segmentation or the domain transform.

3. Additive noise which is introduced by channels in the transmission.

Generally speaking, if the noise is additive random noise, mean filters, median filters and Gaussian filters can be used to reduce the noise thus improving the signal-to-noise ratio. The mean filter is widely used when the noise distribution is plain and the peak value is not high. The median filter has a good performance to filter the impulse noise and to highlight image edges and details. The Gaussian filter works well when filtering Gaussian white noise.

2.5.2 Background subtraction

The purpose of moving target detection is to locate the target and get the contour profile. Whether or not the contour profile can accurately display the target greatly affects the quality of results. Due to illumination change, complexity of background and shadows, detecting moving targets is still a challenge. Basically, several methods have been used to detect a moving target [51].

2.5.2.1 Background subtraction

The simplest method is to compare the difference of the current picture with a static background.

$$|I_{current}(i,j) – I_{background}(i,j)| \quad > \quad Threshold \tag{2.3}$$

The static background may change with illumination, camera oscillations and tiny unwanted moving objects such as moving tree leaves, so the difficulty lies in the quality of the background.

2.5.2.2 Frame differencing

The other method is frame differencing. The background is calculated according to the previous n frames [52, 53]. This method consumes very large memory resources.

2.5.2.3 Running average

The running average of the background can save memory resources.

$$B_t = \alpha I_t + (1-\alpha) B_{t-1} \tag{2.4}$$

where I_t is the pixel's current value, α is the weight. α determines the updated speed. Typically, α is 0.05. This method can deal with the background changing problem.

The selectivity of the running background average:

$$\begin{aligned} B_{t+1}(x,y) &= \alpha F_t(x,y) + (1-\alpha) B_t(x,y) \;\; if\, F_t(x,y)\; is\, background \\ B_{t+1}(x,y) &= B_t(x,y) \;\; if\, F_t(x,y)\; is\, foreground \end{aligned} \tag{2.5}$$

2.5.3 Image segmentation

Image segmentation is used to subtract targets from the background or separate different targets. Image segmentation does not only compare data, reduce the storage, but also greatly simplify subsequent analysis and processing steps.

• Histogram threshold segmentation: the grey histogram is the relationship between the grey pixels and the grey level. If both the distributions of foreground objects and the background are even, the grey histogram would have two peaks. The bottom of the two peaks can be chosen as the threshold. Because the histogram does not contain the location information of the object, the location should be determined by the content of an image.

• Guaranteed convergence particle swarm optimizer (GCPSO): this uses the overall grey difference between target objects and the background to set up the threshold.

• Regional growth : combines similar pixels around the target in the target area.

• Morphological segmentation: the goal is to make the domain of the moving object more complete.

– Erosion is to eliminate the object boundary and remove objects smaller than structural elements. If two objects have small connection, when the structural elements are big enough, they can be separated by corrosion.

– Dilation is to add the background targets. If two objects are very close, dilation could make the two objects appear connected together. It is very useful for removing the holes and cracks of images.

2.5.4 Feature extraction and presenting

The feature extraction methods can be divided into two categories: model-based and appearance-based. Model-based methods detect objects based on the target structure; these methods have higher accuracy and can deal with the occlusion problem, that is some objects block the objects behind them. The main shortcoming of model-based methods is the high computation load. Appearance-based methods calculate characteristics and analyze the information in a statistical way. The computation load is lower but the accuracy is lower, too. The occlusion problem is hard to address in appearance-based methods.

2.5.4.1 Appearance-based method

This method does not require an a-priori model and only need to statistically analyze the space-time model generated by target movements in image sequences. This method defines the predicted observed characteristics to extract and analyze the target. This method is often used in low resolution image sequences. A number of appearance-based approaches to target recognition have been investigated, such as silhouette baseline algorithm [54], silhouette mapping [55, 56], Energy image [57], Contour [58, 59, 60] and key Fourier description.

2.5.4.2 Model-based method

Model-based methods create models to observe targets, such as human joint angle changes, height, leg length and stride. The key to tracking these characteristics is to accurately find the speed of each part of the body. The advantage of a model-based approach is that a model can handle noise and occlusion and directly fetch signatures from model parameters. Model-based methods can also reduce the dimensionality. The disadvantage is that the computation is very high as it needs to continuously match and search.

The most used models are usually presented by blobs, ribbons or skeletons. Wren [61] proposed a method describing humans as a set of blobs which represented people's head, body, hands and legs.

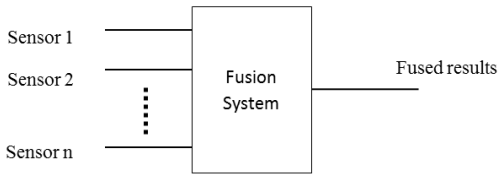


Figure 2.5: Fusion system

2.6 Principle of sensor fusion

The purpose of sensor fusion is to make sensing more accurate, complete and dependable. The fusion system inputs data from different sensors and outputs the more informative results, which is shown in Figure 2.5. The image fusion is one kind of sensor fusion, which is the projection of two or multiple images into a new image. The new fused image contains more useful information than any of the source images.

2.6.1 Sensor fusion levels

The methods to fuse the sensor information vary in different applications. Generally, sensor fusion can be performed at three different levels: the data level, the feature level and the decision making level. In an image fusion system, the three levels are pixel level, kinematic feature level and decision making level.

The lowest fusion level is the data level. Only the same types of sensors can be fused in the data level. If different sensors are used to create images, and if the image is the same reflection of the objects, images can be fused at the pixel level. Fusion at the pixel level needs re-sampling of image data to a common pixel scaling and map projection. At the pixel level image fusion, as the resolutions of different images are different, a pixel in an image may relate to a region of pixels in the other image. So this category can further been divided into pixel-to-pixel, pixel-to-region and region-to-region fusion.

The middle fusion level is the feature level. Features need to be subtracted from captured objects. In most tracking systems, only the moving objects are of interest. If the target is considered as a rigid body, the target can be subtracted from the background and only the shapes and the kinematic information is kept for further

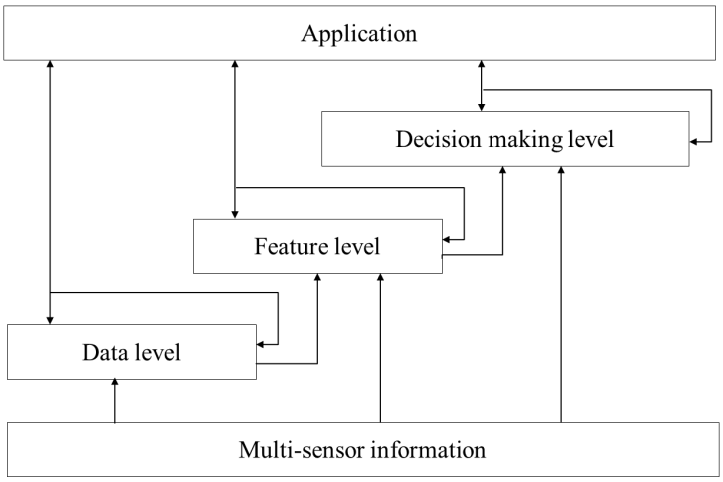


Figure 2.6: Sensor fusion levels

processing. The kinematic information calculation is based on Newton's laws of motion. Some filters can be used to minimize the measurement errors by fusing the kinematic information from different sensors. This kind of method can be described as feature level fusion.

The highest fusion level is the decision making level fusion. The methods include fuzzy logic, voting and some other statistical methods.

For the fusion system in Figure 2.5, the inputs could be data from different sensors either at the same level, different level or multiple levels. In most applications, the information is fused at the same level. The outputs could be further fused with other sensor's results at other levels. Sensor fusion usually integrates information from lower levels to higher levels. Sometimes, higher fusion levels can also send feedback to lower fusion levels. In [17], one sensor provides high level information and uses the raw data from other sensors to verify and improve the accuracy of the information. The structure of fusion levels is shown in Figure 2.6.

2.6.2 Sensor fusion methods

Much research has been done on senor fusion [62, 63, 64, 16, 14]. The fusion algorithms are the key to providing highly flexible and robust tracking and identification in real environments. Basic methods in sensor fusion include: the weighted

average, the Bayesian estimation, the Kalman filter and the statistical decision theory.

1. Weighted average is the simplest and most intuitive method. Weighted average is usually used in low-level dynamic data processing. The computation load is low compared with other methods.
2. Bayesian estimation is often used to fuse multi-sensor data in a static environment. Bayesian estimation can reduce the uncertainty such as in applications with additive white Gaussian noise.
3. In linear systems when the observed noise is Gaussian noise, Kalman filtering is used to fuse dynamic low-level redundant sensor information.
4. In non-linear systems, extended Kalman filtering (EKF) or decentralized Kalman filtering (DKF) are often employed.
5. Statistical decision theory is used in fusing the same type of data from multi-sensors, such as image data.

2.6.3 Image fusion methods

There are many popular image fusion methods such as the high pass filtering based method, the weighted average, the principal component analysis (PCA) and the wavelet transform. Basically, they can be divided into two categories: spatial domain fusion and transform domain fusion. In spatial domain fusion, images are fused by pixels, while in transform domain fusion, some transforms are required before fusion.

2.6.3.1 Spatial domain fusion

Weighted average is a popular method which calculates the average of corresponding pixels of source images. To get the weights, the principal component analysis (PCA) can be used. Also the weights can be adaptive.

Nonlinear methods separate the source image into low and high pass components. The low pass component fusion can enhance the local quality by a linear mapping, while the high pass component fusion can enhance the local contrast by a weighted average.

2.6.3.2 Transform domain fusion

Pyramid method is a popular method in transform domain fusion. This method divides the image at different scales. There are many methods to transform the image, such as the Laplacian transform and the discrete wavelet transform.

2.7 Evaluation of the system

Fusion evaluation can be divided into two categories: manual and automatic. In manual methods, the fused objects are manually observed and evaluated by people. This kind of method is expensive, time consuming and inconvenient. Also the result cannot be checked repeatedly as people have bias. The other category is objective methods which predict human perception. There are two types of objective methods. One requires a reference image, while the other does not. The method which requires a reference image is used in this thesis to analyze the performance of the proposed system.

2.7.1 Evaluation measure for image pixels

The way to measure the errors between the fused image and the true image is to calculate the root mean square errors

$$E_{rms} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|\hat{I}(i,j) - I(i,j)|^2} \tag{2.6}$$

where $I(i,j)$ is the pixel in the true image, $\hat{I}(i,j)$ is the pixel in the fused image, and the image size is M by N.

2.7.2 Evaluation measure for position, velocity and acceleration

If the targets are separated from the image separately and the sensor data in the feature level are fused, the root mean square position error (RMS-PE), velocity error (RMS-VE) and acceleration error (RMS-AE) between the fused results and the true state can be used for evaluation of the system.

$$RMS-PE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(p_{xt}(i)-\hat{p}_x(i))^2 + (p_{yt}(i)-\hat{p}_y(i))^2 + (p_{zt}(i)-\hat{p}_z(i))^2}{3}} \tag{2.7}$$

where, N is the sample number, $p_{xt}(i)$, $p_{yt}(i)$ and $p_{zt}(i)$ is the true position while $\hat{p}_x(i)$, $\hat{p}_y(i)$ and $\hat{p}_z(i)$ is the estimated position. For the velocity error and acceleration error, the equation has a similar expression.

$$RMS-VE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\left(v_{xt}\left(i\right)-\hat{v}_x\left(i\right)\right)^2+\left(v_{yt}\left(i\right)-\hat{v}_y\left(i\right)\right)^2+\left(v_{zt}\left(i\right)-\hat{v}_z\left(i\right)\right)^2}{3}} \quad (2.8)$$

$$RMS-AE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\left(a_{xt}\left(i\right)-\hat{a}_x\left(i\right)\right)^2+\left(a_{yt}\left(i\right)-\hat{a}_y\left(i\right)\right)^2+\left(a_{zt}\left(i\right)-\hat{a}_z\left(i\right)\right)^2}{3}} \quad (2.9)$$

3 Image construction of the fourth layer with distance using pulsed beamforming - Addresses Proposition 1

3.1 Introduction

The aim of this chapter is to provide a solution to create the fourth layer of the proposed image. The image provides the Angle of Arrival (AoA) and the range.

Radars have many varieties, and most of them send pulses or modulated continuous waves to measure the Time of Flight (ToF) to detect the presence and the range of the target. To detect the AoA, two popular methods are used. Some radars slowly steer the narrow beams toward the detection area. The mechanism looks like scanning in the space. In each direction, if there is an object being detected, the AoA of the object is the direction of scanning. Some other radars use sensor arrays and compare the phase difference between adjacent receivers to detect the AoA. The phase differencing can also be applied when sensor arrays are used to transmit signals. The principle is similar to narrow beam steering radars, but the beam is electronically steered without mechanical movement.

The aim of this research is to build a portable, airborne, vision range low cost radar. Beam steering radars are not suitable as the mechanical parts increase the cost and reduce the robustness. If real time radar processing is required, the beam steering speed should be very high. With the development of hardware on signal processing, phased sensor array radar is suitable for this application.

In terms of processing methods of phased sensor array radars, there are many specified methods in different applications. By comparing different methods, beamforming is a suitable and reliable method to apply.

Specifically, beamforming which is also called spatial filtering, is a radar signal processing algorithm which uses an array of sensors to enhance the amplitude of signals from a specified direction while reducing the amplitude of signals from other directions. Beamforming can be done on either the transmitter side or receiver side.

This method has been widely used for the detection of targets in remote sensing applications. However it is not widely used for the creation of images. Also, the beamforming cannot detect the distance of the objects. In this chapter, the theory and mathematics of pulsed radar and beamforming radar were investigated. To simplify the model, the medium constraints were investigated and two approximations were made. Based on the theory, the linear array beamforming was investigated in both the time domain and frequency domain and the performance of different methods were compared. Then the linear array beamforming methods were extended to two dimensions and the performance of different structures were compared. The principles of the pulsed radar and the matched filter were also investigated to improve signal to noise ratio and minimize false alarms. Finally a two dimensional acoustic pulsed beamforming radar was designed and the algorithm was described. By adopting this radar, a new image layer with the information of the target range was created. This image can be used in combination with traditional RGB imaging to improve image processing accuracy and speed.

3.2 System structure

In chapter 6, the reason why the hardware was chosen and how the hardware was built are described. Generally, PCs or control boards were used to control the FPGA board and process the data. In beamforming processing, a large amount of sensor data is needed to be processed simultaneously. The processing algorithm of each sensor is very similar. A FPGA is ideal and simple for duplicating parallel processing, therefore, no matter how many sensors are used as receivers, if there are enough resources in the FPGA, the increasing of sensor number will not significantly affect processing speed. For research purposes, all the data were transferred from FPGA to computers and the software such as Matlab and Visual C++ on windows were used to process the data. In industrial manufacturing, a PC is not necessary for processing the raw data. A control board with proposed algorithms can control the FPGA and process the data. The control system can output the results to a PC or other device for further action.

The basic structure of the proposed airborne pulsed and Doppler beamforming radar is shown in Figure 3.1

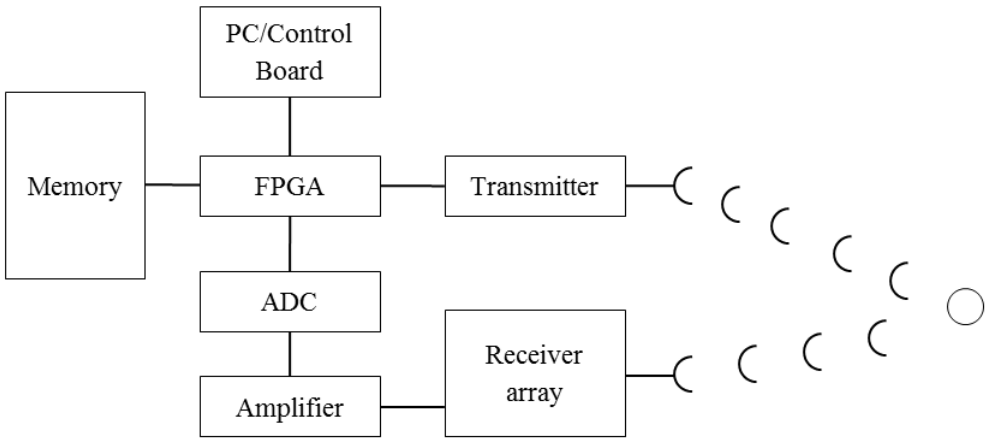


Figure 3.1: System structure of the proposed radar

3.3 Medium constraints and assumptions

3.3.1 Comparison of airborne ultrasonic radar with sonar and radar

Acoustic radar has a much lower carrier frequency than electromagnetic wave radar while the wave length is similar. Acoustic radar has the advantages that the structure is simple, the system is robust and the signal processing load is low. For research purposes, all the received data can be captured and transferred into PCs without pre-processing.

Sonar is a type of acoustic radar which is used under water, as the attenuation of sound in water is much lower than electromagnetic waves. The velocity of propagation can affect the frequency, transmitting range, pulse duration and repetition. The requirements of a sonar are different from a radar. In Table 3.1, the differences between airborne acoustic radar, sonar and radar are compared.

3.3.2 Velocity of sound

The sound velocity c is determined by the Newton-Laplace equation:

Table 3.1: Comparison of airborne acoustic radars, sonars and radars [6]

Property	Airborne acoustic radar	Sonar	Radar
velocity of propagation	$330m/s$	$1.5 \times 10^3 m/s$	$3 \times 10^8 m/s$
frequency range	$10 - 1000kHz$	$0.1 - 10kHz$	$300 - 30000MHz$
typical pulse duration	$0.0001 - 0.1s$	$0.01 - 1.0s$	$0.1 - 10\mu s$
typical pulse repetition rate	$20 - 1s^{-1}$	$0.2 - 0.01s^{-1}$	$2000 - 200s^{-1}$
typical wavelength	$0.08m$	$0.3m$	$0.03m$
Doppler shift	$\pm 10\%$	$\pm 2\%$	$\pm 0.0003\%$

$$c = \sqrt{\frac{K}{\rho}} \tag{3.1}$$

where ρ is the density, K is a coefficient of the bulk modulus of elasticity.

In dry air with 0% humidity, when the temperature is near 0 °C,

$$c_{\text{air}} = (331.3 + 0.606 \cdot \vartheta)\, m \cdot s^{-1} \tag{3.2}$$

where ϑ is the temperature in degrees Celsius (°C).

The temperature can be measured to adjust the results. In some applications, if the accuracy requirement is not very high, the system is assumed to work at 18°C, and then the speed of ultrasonic waves is approximately $342m/s$.

3.3.3 Attenuation of ultrasonic waves

The attenuation of sound in a medium depends on the following equation:

$$A = A_0 \cdot e^{-\alpha Z} \tag{3.3}$$

$$\alpha = \frac{0.1151}{v} U_t \tag{3.4}$$

where A_0 is the unattenuated amplitude of the wave, α is the attenuation coefficient of the wave, z is the distance, v is the velocity of sound, and U_t is in decibels per second.

Some experiments have been done and the results have demonstrated that for a 40kHz acoustic radar, after 20 meters, which is 40 meters in total, the received signals are still strong enough to be detected and the signal to noise ratio (SNR) is high. The measured attenuation is also affected by the sensor's sensitivity. As most ADCs need to sample at 0-5 Volts, op-amps are used to amplify the signals to fit the ADC working voltage. Also a circuit should be used to limit the amplitude to protect the ADCs.

In this application, 10bit ADCs were used, so the resolution was 1024 steps, which qualified it for both long distance and short distance object detection. If 12bit ADCs were used, the resolution could be even higher, but for the beamforming processing, the performance of 12bit ADCs was tested and found to be similar to 8bit ADCs, and the memory consumption was much higher.

To avoid noise from the transmitter directly to the receivers, the signals representing the objects with range lower than a meter were omitted and all the data in this range were set to zero.

3.3.4 Carrier frequency

For the cost consideration, the hardware needed to work for both pulsed beamforming and Doppler beamforming. High frequency is required for adequate Doppler-based measurement of low velocity objects [29]. On the other hand, processing radar signals required a great amount of calculation especially when the frequency was high. This is part of the reason why radar is so expensive. Also, according to the attenuation equation in the last subsection, the high frequency acoustic radar has short transmitting distance. Therefore, to reduce the cost and make the structure much simpler, the acoustic radar with relatively low frequency was found to be suitable for the application.

Acoustic airborne radars usually work from 2 Hz to 2 MHz. As noise needed to be avoided in human hearing range, the frequency was set at 40kHz, which was a popularly used frequency and the transducers could be easily bought. The wavelength of 40kHz acoustic wave is similar to the wavelength of a 35GHz electromagnetic wave. This bandwidth is often used for a Doppler radar. The wavelength is the minimum size of objects which can be detected, which is called the spatial resolution of the radar. So the 40kHz ultrasonic sine wave radar has a similar spatial resolution to the 35GHz electromagnetic wave radar.

$$\begin{aligned} \lambda_{sound} &= V_{sound}/f_1 = (340m/s)\,/40kHz = 8.5mm \\ \lambda_{light} &= C/f_2 = (299,792,458m/s)\,/35GHz = 8.5mm \end{aligned} \tag{3.5}$$

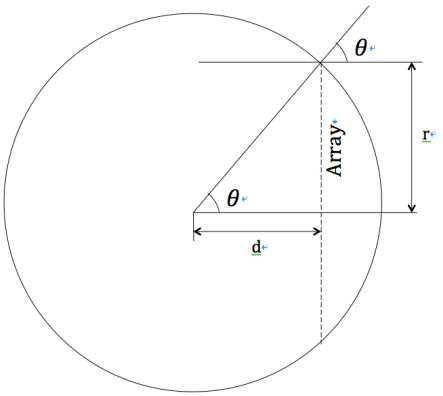


Figure 3.2: The far-field approximation [2]

3.3.5 Far field assumption

When the size of the sensor array is much smaller than the estimated object distance from the array to the object and the desired accuracy of the angle of arrival (AoA) is not very high, the direction of propagation for each sensor in the array is approximately the same, which can be seen in Figure 3.2.

In Figure 3.2, signals are transmitted from the center of the circle. The length of the linear array of sensors is $2r$. The perpendicular distance between transmitter and sensor array is d. It is assumed that $d = 10m$ and $r = 0.03m$. So θ is approximately $\arctan(0.03/10)/\pi * 180 = 0.1719$. If the angle resolution is 2^{o}, the angle difference between the farthest sensors can be ignored. The approximation can be made that all the received signals come in parallel.

3.3.6 Narrow bandwidth assumption

Based on the far field assumption, the AoA of all the sensors are approximately the same. In Figure 3.3, $\overrightarrow{p}$ is a unit-length vector in the direction of propagation. This unit-length is equal for each sensor. Also $\overrightarrow{d}$ is the vector from the first sensor to the second sensor. If all the sensors are laid in a line with even spacing, $\overrightarrow{d}$ is equal for every adjacent sensor. So the time delay between adjacent sensor is

$$\tau = \frac{\overrightarrow{d} \cdot \overrightarrow{p}}{c} = |\overrightarrow{d}| \cdot \sin\theta / c \tag{3.6}$$

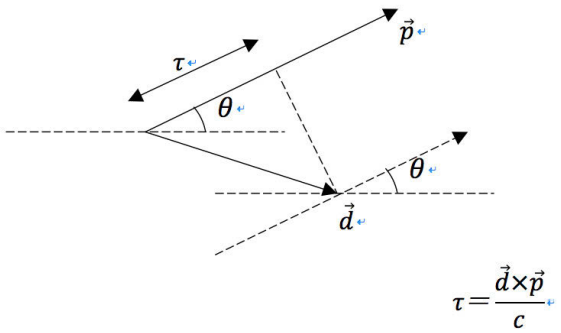


Figure 3.3: Narrow band approximation [2]

where c is velocity of sound. The signal received by the first sensor is supposed to be $x_1(t) = A(t)\, e^{j\omega t}$, so the signal received at the second element is $x_2(t) = x_1(t-\tau) = A(t-\tau) \cdot e^{j\omega(t-\tau)} = A(t-\tau) \cdot e^{j\omega t} \cdot e^{j\omega\tau}$. The narrow band system means a channel in which the bandwidth is roughly smaller than the coherence bandwidth. For a narrow band system, when the delay τ is small, the approximation can be made that $A(t) = A(t-\tau)$. Therefore, $x_2(t) = x_1(t-\tau) = A(t-\tau) \cdot e^{j\omega t} e^{j\omega\tau} = x_1(t)\, e^{j\omega\tau}$. The time-delay can be converted to the phase shift.

3.3.7 Noise type

Noise can incur in different stages of the signal processing. Basically, noise can come from the background, the system process and the measurement errors. Noise can be divided into additive noise and non-additive noise. Additive noise includes white noise, Gaussian noise and salt-and-pepper noise in image. Non-additive noise includes multiplicative noise and poison noise. The Gaussian noise is the statistical noise with the probability density function equal to a normal distribution. The salt and pepper noise is the noise which often affects black and white TV. The additive noise is not usually related to real signals, so it can be mostly removed by filters. Other types of noise are a bit harder to remove.

3.4 Linear array beamforming

In this research, to simplify the structure, only the case in which the beamforming was used on the receiver side was studied.

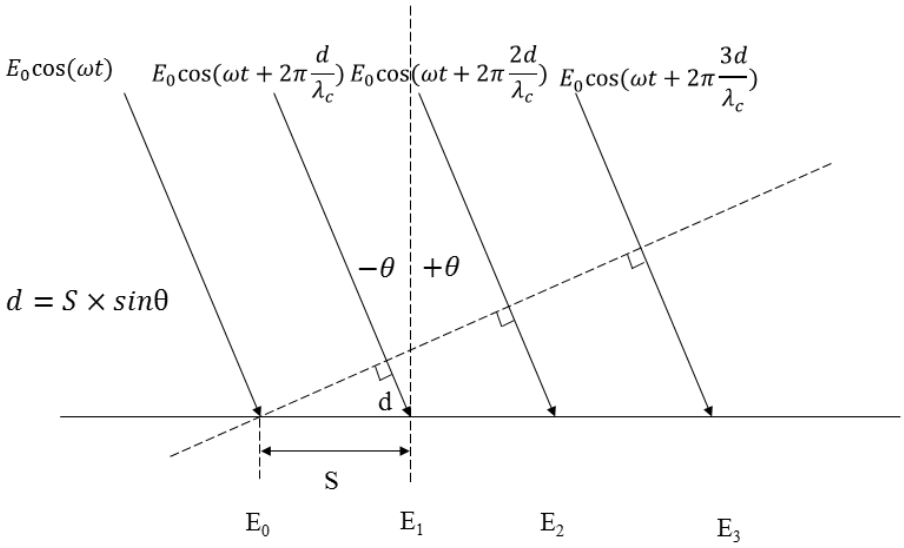


Figure 3.4: Beamforming calculation

The basic principle of beamforming processing is that the radar electronically steers beams by adding delay to each sensor according to the direction.

The structure can be seen from Figure 3.4.

3.4.1 Sensor spacing

The hardware in this thesis was designed for both pulsed beamforming and Doppler beamforming and the sensor spacing could not be changed after the radar was built. Therefore the sensor spacing needed to be suitable for both pulsed beamforming and Doppler beamforming radar.

To make the calculation easier, for a linear array of sensors, the spacing between adjacent sensor is equal as shown in Figure 3.4. All the spacings are equal to S. According to the assumption provided before, the time delay t_d can be expressed both by AoA and phase difference.

$$t_d = S \cdot \sin(-\theta)/c \tag{3.7}$$

$$t_d = \frac{\phi}{2\pi} \cdot (\lambda_c/c) \tag{3.8}$$

where $\theta \in [-\pi, \pi]$ is the AoA, $\phi \in [-\pi, \pi)$ is the phase delay, λ_c is the wavelength of the carrier frequency and c is the velocity of sound. Combining these two equations, the AoA can be obtained from the phase difference.

$$\theta = -\arcsin\left(\frac{\phi}{2\pi} / \left(\lambda_c / S\right)\right) \tag{3.9}$$

If θ can be detected by ϕ without ambiguities, for any ϕ, there is only one unique value of θ. According to the characteristic of the inverse sine function, inverse sine is monotonic when $\frac{\phi}{2\pi} / \left(\lambda_c / S\right) \in [-1, 1]$, as $\phi \in [-\pi, \pi)$, then $\frac{1}{2\pi} / \left(\lambda_c / S\right) \in \left[-\frac{1}{\pi}, \frac{1}{\pi}\right]$, therefore,

$$S \leqslant \lambda_c / 2 \tag{3.10}$$

When the sensor spacing is less than $\lambda_c/2$, there is no ambiguity for pulsed beamforming. For Doppler beamforming, according to Equation 4.1, the Doppler frequency is lower than the carrier frequency if the radial velocity of the target is lower than the velocity of sound. On this condition, if the sensor spacing is less than $\lambda_c/2$, there is no ambiguity in Doppler beamforming either. However, if the sensor spacing is not changed, the relationship between adjacent sensor phase shifts and AoA changes, in Doppler beamforming. The relationship between AoA and Doppler signal phase shifts will be discussed in chapter 4.

To achieve high resolution, in pulsed beamforming, set $S = \lambda_c/2$, and the explanation is in subsection 3.7.1.

3.4.2 Time domain beamforming

In Figure 3.4, a linear array of sensors are allocated facing the same direction with the same height. A transmitter is sending acoustic waves towards the target. According to the far field assumption, the returned signals come back parallel. According to the narrow bandwidth assumption, they can be written as follows:

$$\begin{aligned} s_q(t) &= E_q \cos\left(\omega_c t + q \times 2\pi d / \lambda_c\right) \\ &= E_q \cos\left(\omega_c t - q \times 2\pi \cdot S \cdot \sin\theta / \lambda_c\right) \\ &= E_q \cos\left(\omega_c \left(t - q \times 2\pi \cdot S \cdot \sin\theta / c\right)\right) \end{aligned} \tag{3.11}$$

where, $s_q(t)$ = received signals, E_q = maximum amplitude of received signal of qth sensor after attenuation in air, ω_c = angular frequency of carrier waves, S = space between sensors, λ_c= wavelength of carrier waves and ω_c, f_c are the angular frequency and frequency of the carrier wave.

The time delay of sensor q is defined as $\triangle t_q = -q \times 2\pi \cdot S \cdot \sin\theta/c$.

$$s_q(t) = E_q \cos\left(\omega_c \left(t + \triangle t_q\right)\right) \tag{3.12}$$

A computer can only process digital sampled signals, therefore, the signals need to be sampled first. And f_s is the sampling frequency. According to the Nyquist sampling theorem, if the original signals need to be recovered, f_s should be at least 2 times the highest frequency in the signal to be sampled, that is $f_s \geqslant 2 \cdot f_c$. Then $t = nT_s$, $n = 0, 1, 2, 3, ...$, and $T_s = T_c \cdot \frac{f_c}{f_s}$

$$s_q[n] = E_q \cos\left(\omega_c \left(nT_s - q \cdot 2\pi \cdot S \cdot \sin\theta/c\right)\right) \tag{3.13}$$

As high angle resolution is required, S is set to be $\lambda/2$,

$$\begin{aligned} s_q[n] &= E_q \cos\left(\omega_c T_s \left(n - q \cdot \frac{f_s}{2f_c} \cdot \sin\theta\right)\right) \\ &= s_0\left[n - q \cdot \frac{f_s}{2f_c} \cdot \sin\theta\right] \end{aligned} \tag{3.14}$$

The basic idea of beamforming is adding delay to each sensor and summing the received signals.

If time delay $-q \cdot \frac{f_s}{2f_c} \cdot \sin(-\varphi)$ to sensor q is added, the sum of the data with supposed delays from all sensors is

$$y[n, \varphi] = \sum_{q=0}^{Q-1} E_q \cos\left(\omega_c T_s \left(n - q \cdot \frac{f_s}{2f_c} \cdot \sin\theta - q \cdot \frac{f_s}{2f_c} \cdot \sin(-\varphi)\right)\right) \tag{3.15}$$

If $\varphi = \theta$,

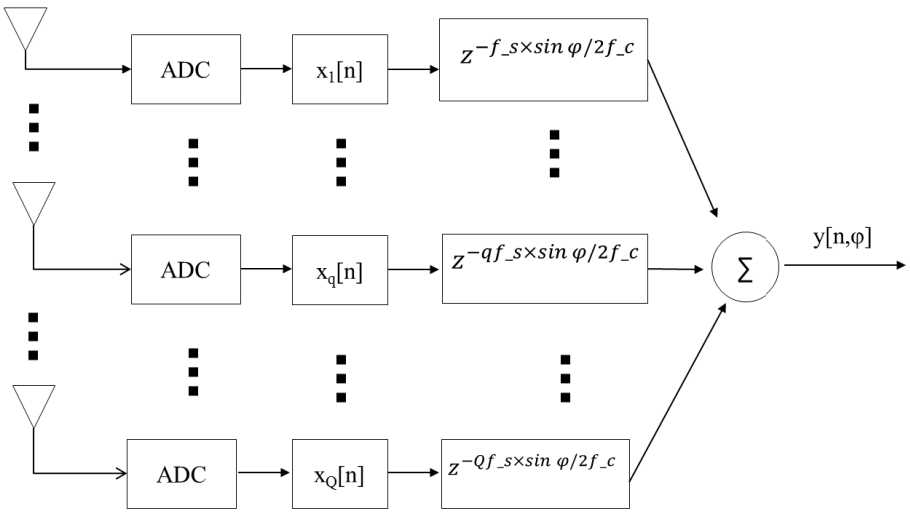


Figure 3.5: Time domain beamforming procedure

$$y\left[n,\varphi\right] = \sum_{q=0}^{Q-1} E_q \cos\left(\omega_c T_s n\right) \tag{3.16}$$

$$= \sum_{q=0}^{Q-1} s_q\left[n\right] \tag{3.17}$$

In this case, the power of $y\left(n,\varphi\right)$ achieve the maximum.

The flowchart of time domain beamforming is shown in Figure 3.5.

In the time domain, the added delay must be an integral number. It is assumed that $q \cdot \frac{f_s}{2f_c} \cdot \sin\varphi = N$, where N must be an integer. It can be found that the number of possible values of N is the beam number. Therefore,

$$\varphi = -\arcsin\left(N\frac{2f_c}{q \cdot f_s}\right) \tag{3.18}$$

Therefore, $\left|N\frac{2f_c}{q\cdot f_s}\right| \leqslant 1$, and for any value of q, this equation must be fulfilled, so

$$\left|N\frac{2f_c}{f_s}\right| \leqslant 1 \tag{3.19}$$

According to Equation 3.19, if $f_s = 2f_c$, Ncan be -1,0,1, so that 3 beams can be detected, and if $f_s = 4f_c$, N can be -2,-1,0,1,2, so that 5 beams can be detected.

If the sensor spacing is fixed at $S = \lambda/2$, the way to increase beam number is by increasing the sample frequency. In this application, to create a beamforming image, more than 30 beams were required. If sample frequency is increased, the device cost will increase sharply and data loads will expand quickly. In particular high speed ADCs are very expensive.

3.4.3 Interpolation beamforming

There are several improved methods to increase beam number without changing the hardware, such as interpolation beamforming. In interpolation beamforming, an interpolation factor D is used which is an integer value. The interpolation filter is used to digitally increase the sample rate.

The method adds a vector of $(D-1)$ zeros between 2 samples, then a low pass filter is used to generate the final sample series. The procedure is shown in Figure 3.6.

Then, the sample time of interpolation beamforming equals $1/(f_s \cdot D)$. This method increases the sample frequency, thereby increasing the beamforming number. However, this method still could not solve the problem when the system required more than 30 beams as D could not be very high.

3.4.4 Frequency domain beamforming

In the frequency domain, according to the Nyquist sample theorem, if $f_s/f_c \geqslant 2$, the whole signal can be recovered. Compared with time domain beamforming, the sample frequency does not need to be increased to increase beam numbers. In discrete Fourier transform, the time delay can be transformed into a phase shift. That is

$$x[n-\Delta n] \quad \leftrightarrow \quad X[k]\cdot e^{-j2\pi\Delta nk/N} \tag{3.20}$$

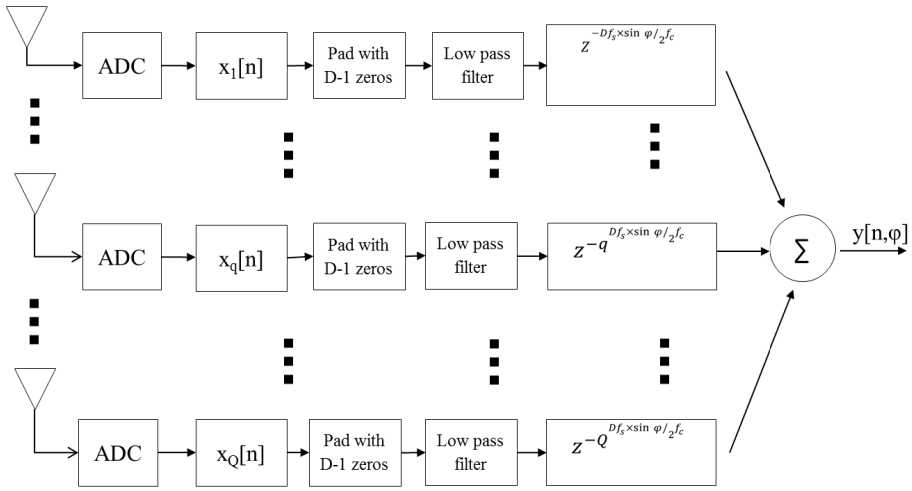


Figure 3.6: Interpolation beamforming procedure

So $s_q[n]$ can be transformed into the frequency domain. The sample length is assumed to be N.

$$S_q[k] = \sum_{n=0}^{N-1} s_q[n] \cdot e^{-j2\pi nk/N} \tag{3.21}$$

The supposed time delay in direction φ of sensor q is $\Delta n_{\varphi,q} = -q \cdot \frac{f_s}{2f_c} \cdot \sin\varphi$. So the delay and sum of the received signals $Y[k,\varphi]$ is

$$y[n,\varphi] = \sum_{q=0}^{Q-1} ifft\left(S_q[k] \cdot e^{-j2\pi\Delta n_{\varphi,q}k/N}\right) \tag{3.22}$$

If $\varphi = \theta$, the absolute value of $y[n,\varphi]$ achieves the maximum. The flowchart of frequency domain beamforming is shown in Figure 3.7.

The difference between time domain beamforming and frequency domain beamforming is shown in Table 3.2.

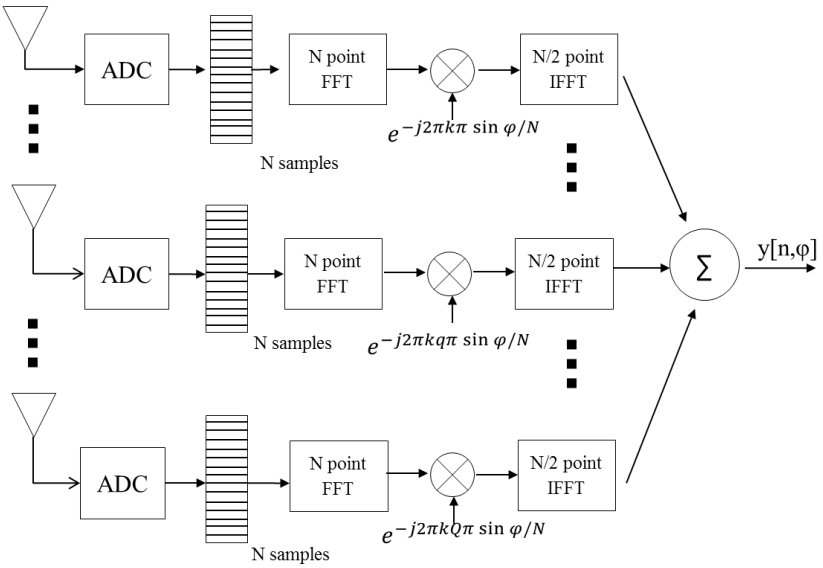
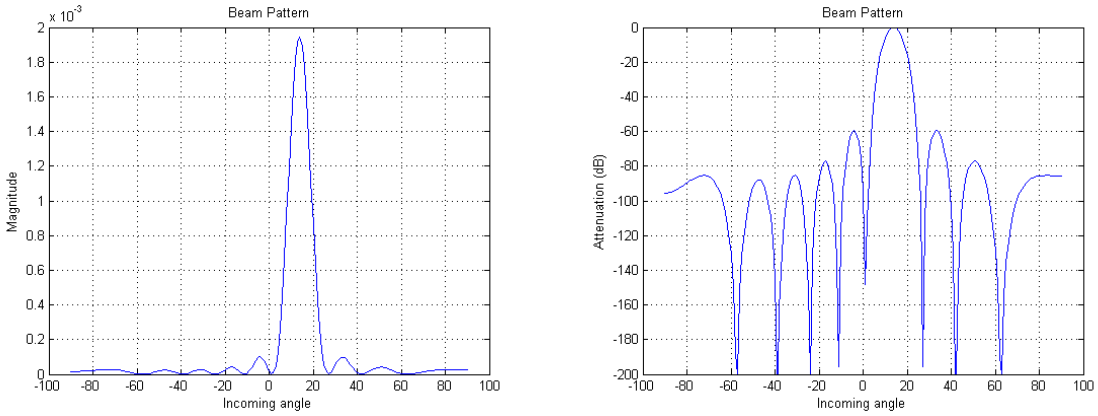


Figure 3.7: Frequency domain beamforming

Table 3.2: Comparison time domain and frequency domain [7]

Time domain	Frequency domain
The number of beams is limited by the sampling frequency	The number of beams formed is not limited by the sampling frequency if the sampling frequency is higher than twice the carrier frequency.
A low sampling rate results in fewer number of beams	Minimum number of beams is equal to the number of sensors
Data storage requirements are high because of high sampling rates	Data storage requirements are low because sensor data of a particular time instance is required.
High sampling rates, more number of beams, high data storage results in increased size and cost of the overall system.	Minimize system size and cost but require more processing power.
Flexible for arbitrary array geometries but with fewer number of sensors.	Flexible for linear and circular arrays and require relatively less hardware when processing sensor arrays with 100 or more sensors.



(a) Beam pattern of linear array beamforming

(b) Beam pattern of linear beamforming (dB)

Figure 3.8: The beam pattern of frequency domain beamforming

The simulation is done in Matlab. The radar is allocated in the origin of the Cartesian coordinate system in the plane, and the position of target is (2,7).

The propagation velocity is 340 m/s, the carrier frequency is 40kHz, the sensor spacing is 0.00425 meters, the number of sensors is 9, and sample number per cycle is 4, the FFT sample length is 256 samples. The simulation results are shown in Figure 3.8.

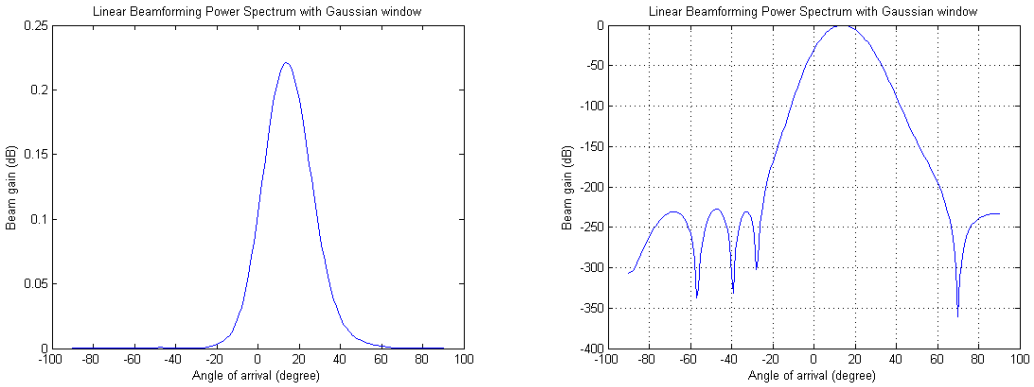
According to the simulation, the target angle is around 17^o which represent the AoA of the target and the main lobe to side lobe ratio is very high.

3.4.5 Window function

There are many types of window function. Some typical window functions include Rectangular window (no window function), Gaussian window, Hamming window, Triangular window and Dolph-Chebyshev window.

Gaussian window is described in Equation 3.23

$$w(n) = e^{-\frac{1}{2}\left(\frac{n-(N-1)/2}{\sigma(N-1)/2}\right)^2}, \sigma \leq 0.5 \tag{3.23}$$



(a) Gaussian window **(b)** Gaussian window (dB)

Figure 3.9: Linear frequency domain beamforming with Gaussian window

The simulation of Gaussian window function is shown in Figure 3.9. The parameter setting is similar to previous simulations.

Dolph-Chebyshev window is described in Equation 3.24 and Equation 3.25.

$$W_0(k) = \frac{\cos\left(N \arccos\left(\beta \cos\left(\frac{\pi k}{N}\right)\right)\right)}{\cosh\left(N \cosh^{-1}(\beta)\right)} \tag{3.24}$$

$$\beta = \cosh\left(\frac{1}{N} \cosh^{-1}(10^{\alpha})\right) \tag{3.25}$$

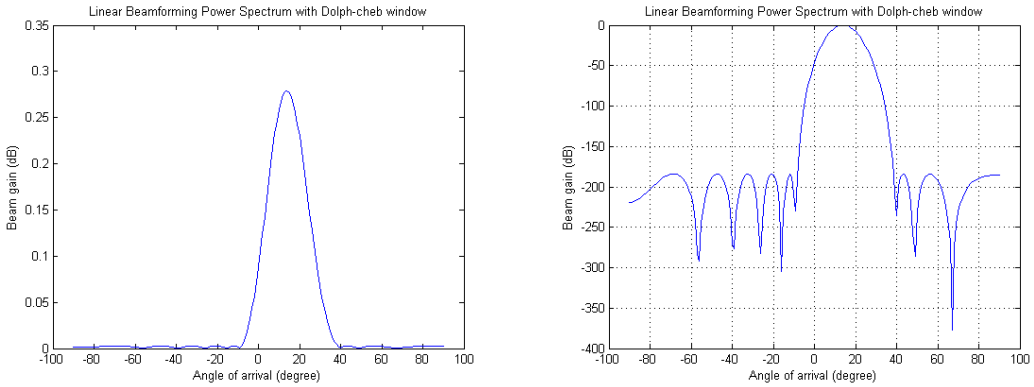
where side lobes level in dB is -20α

The simulation of Dolph-Chebyshev window function is shown in Figure 3.10a; the main lobe to side lobe ratio is set to be 40 dB, that is $\alpha = 2$.

According to the simulation the beamwidth is wider while the main lobe to side lobe ratio is much higher.

3.4.6 Multiple target tracking using frequency domain beamforming

In most applications, there is more than one object. In this section, the multiple target tracking using frequency domain beamforming is studied.



(a) Dolph-Chebyshev window

(b) Dolph-Chebyshev window (dB)

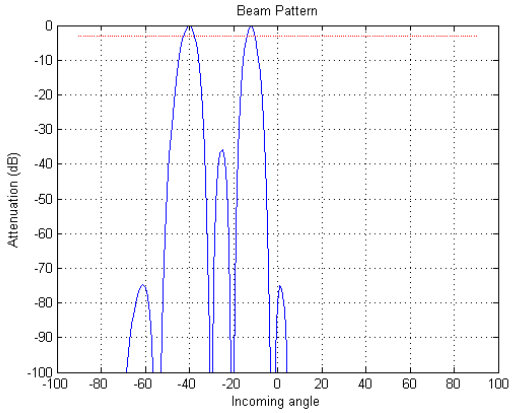
Figure 3.10: Linear frequency domain beamforming with Dolph-Chebyshev window

In delay-sum beamforming, all the sensors are added a phase delay which represents the time delay of sensors. That makes the phases of all the sensors approximately the same. In the case with two objects, the result is roughly the sum of two separate single object tracking results. If the phases of the signals from the two objects to the first sensor in the linear array are similar, the beam patterns interfere with each other; otherwise, the beam pattern of the two objects are separated clearly. This can be shown in Figure 3.11 and Figure 3.12. The number of sensor is 13 and the positions of the two targets are (9,10), (2,10).

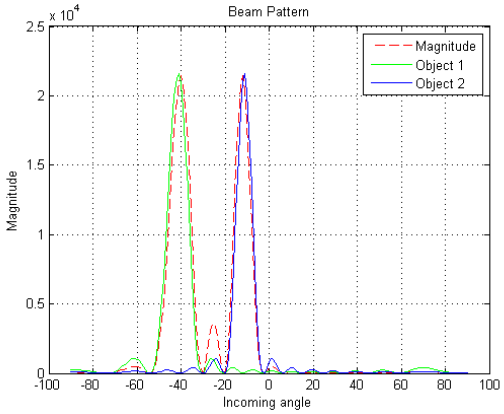
In Figure 3.11, the simulation of the two targets tracking is approximately the sum of the two simulations of single target tracking. There is a small peak in the middle of the two beam peaks. If the position of one target is slightly changed, that is in position (9,10)and (2,10.5), these two targets will not affect each other; instead, they enhance the main lobe to side lobe ratio.

3.4.6.1 Resolution of two objects tracking

The resolution of the two objects tracking depends on the position of the objects. According to the simulations above, in some positions, if the phases of the signals from the objects to the radar are similar, these two objects affect each other, and if the phases of the signals from the objects to the radar is inverse, they do not

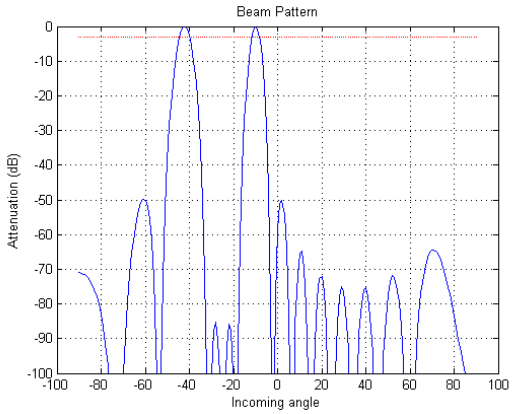


(a) Two objects tracking with interference

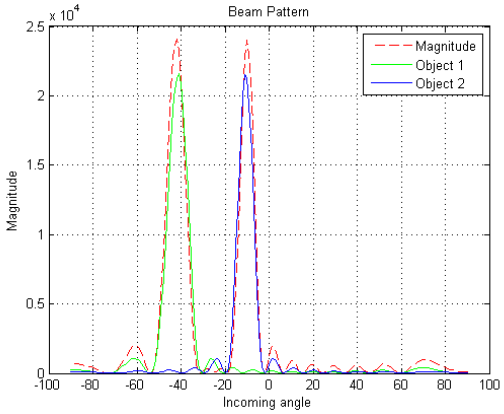


(b) Two objects tracking with interference

Figure 3.11: Two objects tracking with interference

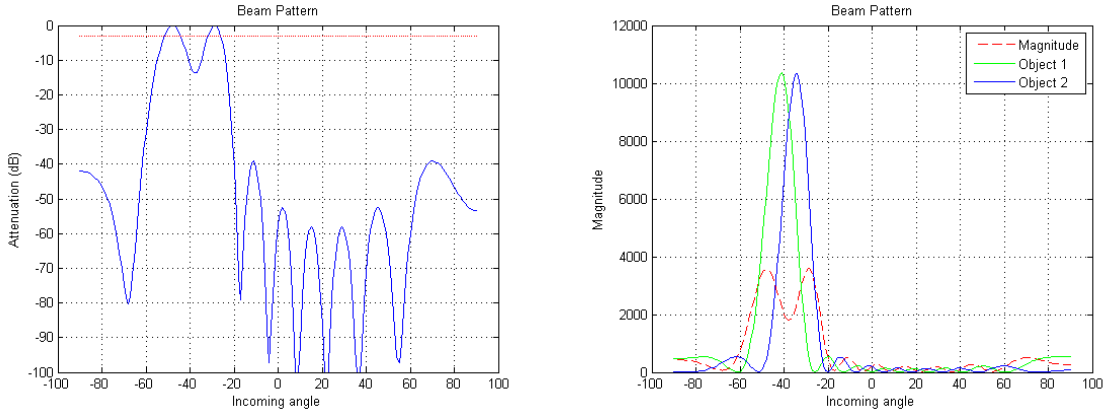


(a) Two objects tracking without interference (dB)



(b) Two objects tracking without interference

Figure 3.12: Two objects tracking without interference



(a) Resolution of two object tracking 1 (dB)

(b) Resolution of two object tracking 1

Figure 3.13: Resolution of two object tracking 1

interfere with each other. When they interfere with each other, the resolution is low and vice versa.

Figure 3.13 and Figure 3.14 shows the results of the two objects in the position (9, 10), (7,10) and (9,10), (7,10.5), separately.

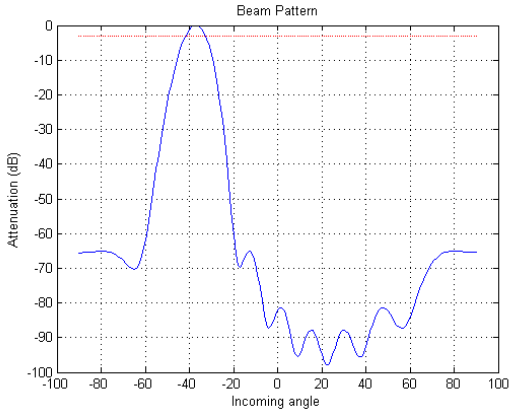
If the two object are in position (9, 10) and (7,10), these two objects can be separated, but both the positions are shifted by about 5^{o}.

If the two object are in (9,10), (7,10.5), these two objects cannot be separated, which is shown in Figure 3.14.

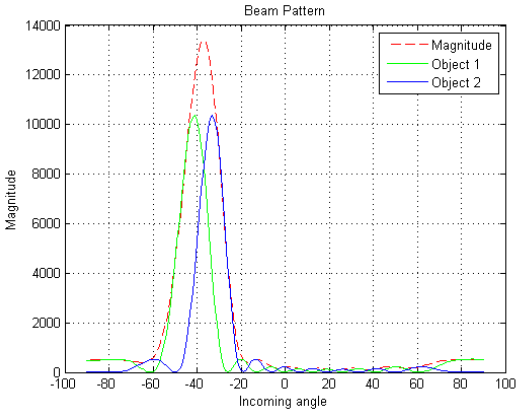
The way to solve the problem is to increase the sensor numbers. The simulation shows that when the sensor number is increased to 17, these two objects in (9,10), (7,10.5) can be separated with no angle shift, which can be shown in Figure 3.15.

3.5 Two dimensional beamforming

The aim of this research is to build a two dimensional pulsed beamforming and Doppler beamforming radar to generate distance and velocity image layers. The

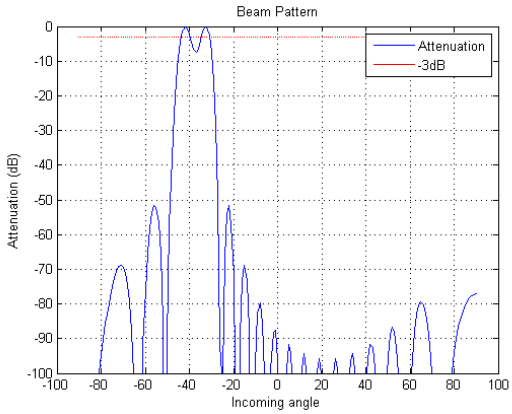


(a) Resolution of two object tracking 2 (dB)

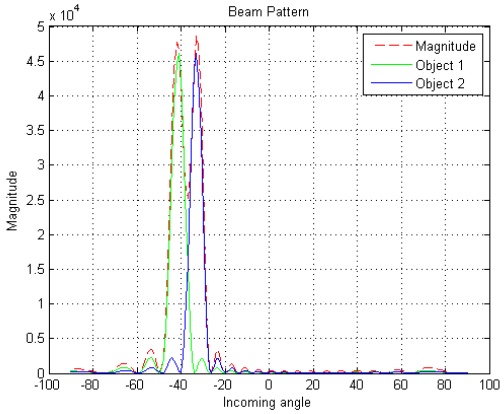


(b) Resolution of two object tracking 2

Figure 3.14: Resolution of two object tracking 2



(a) Resolution of two object tracking 3 (dB)



(b) Resolution of two object tracking 3

Figure 3.15: Resolution of two object tracking 3

principles of the linear array sensor beamforming can be extended to two dimensional beamforming when the linear array sensor layout is extended to two dimensions. In this case, not only the AoA in the horizontal direction, but also the AoA in the vertical direction can be detected. In contrast to one dimensional beamforming, there are many shapes of the sensor array layouts which can be adopted.

3.5.1 Full rectangular array beamforming

The basic sensor array shape is the full rectangular sensor array. Similar to the linear array beamforming, a two dimensional array with M by N sensors is as follows:

$$\begin{bmatrix} q_{0,0} & q_{0,1} & \cdots & q_{0,n} & \cdots & q_{0,N} \\ q_{1,0} & q_{1,1} & \cdots & q_{1,n} & \cdots & q_{1,N} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ q_{m,0} & q_{m,1} & \cdots & q_{m,n} & \cdots & q_{m,N} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ q_{M,0} & q_{M,1} & \cdots & q_{M,n} & \cdots & q_{M,N} \end{bmatrix} \tag{3.26}$$

According to Figure 3.4, the phase delay of sensor $q_{m,n}$ compared with the first sensor in the first line comes from two parts, that is the phase delay between $q_{0,0}$ and $q_{0,n}$, and the phase delay between $q_{0,n}$ and $q_{m,n}$. The AoA in the horizontal direction is assumed to be α and the AoA in the vertical direction is β. Therefore, the total phase delay of $q_{m,n}$ is

$$\theta_{m,n} = n\frac{S_h \omega_c \sin\alpha}{c} + m\frac{S_v \omega_c \sin\beta}{c} \tag{3.27}$$

To achieve the best resolution and simplify the calculation, the sensor spacings in both vertical and horizontal directions are set to a half wavelength, that is $S_h = S_v = \frac{\lambda}{2}$.

$$\theta_{m,n} = n\pi \sin\alpha + m\pi \sin\beta \tag{3.28}$$

Therefore, the signals received by sensor $q_{m,n}$ can be written as

$$s_{m,n}(t) = E_{m,n} \cos(\omega_c t + n\pi \sin\alpha + m\pi \sin\beta) \tag{3.29}$$

If the signals are sampled,

$$\begin{aligned} s_{m,n}[q] &= E_{m,n}\cos\left(\omega_c q T_s + n\pi\sin\alpha + m\pi\sin\beta\right) \\ &= E_{m,n}\cos\left(\omega_c T_s\left(q - n\cdot\frac{f_s}{2f_c}\cdot\sin\alpha - m\cdot\frac{f_s}{2f_c}\cdot\sin\beta\right)\right) \\ &= s_{0,0}\left[q - n\cdot\frac{f_s}{2f_c}\cdot\sin\alpha - m\cdot\frac{f_s}{2f_c}\cdot\sin\beta\right] \end{aligned} \tag{3.30}$$

The N length Discrete Fourier Transform of $s_{m,n}[q]$ can be written as

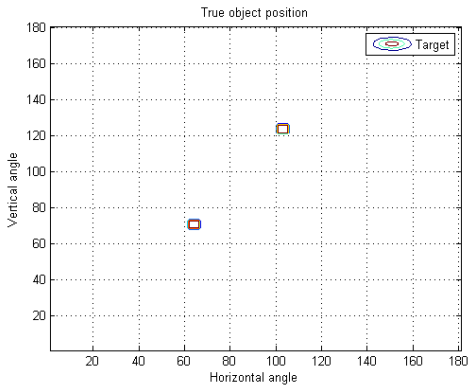
$$S_{m,n}[k] = \sum_{q=0}^{Q-1} s_{m,n}[q]\cdot e^{-j2\pi nk/N}$$

The supposed time delay in the horizontal direction with AoA at ϕ and in the vertical direction with AoA at φ of sensor q is $\Delta n_{m,n,\phi,\varphi} = -n\cdot\frac{f_s}{2f_c}\cdot\sin\phi - m\cdot\frac{f_s}{2f_c}\cdot\sin\varphi$. So the delay and sum of the received signals $y[q,\phi,\varphi]$ is

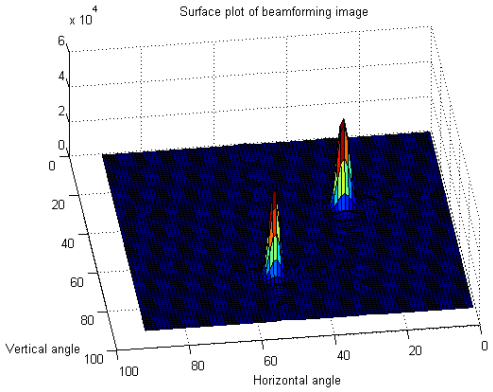
$$y[q,\phi,\varphi] = \sum_{q=0}^{Q-1} ifft\left(S_q[k]\cdot e^{-j2\pi\Delta n_{m,n,\phi,\varphi}k/N}\right) \tag{3.31}$$

Simulation is shown in Figure 3.16. The propagation velocity is 340 m/s, the carrier frequency is 40kHz, the sensor spacing is 0.00425 meters, the number of sensors is 9, the sample number per cycle is 4 and the FFT sample length is 256 samples.

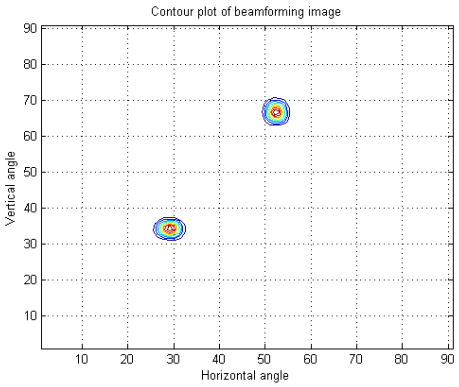
The simulation shows that the background is very clear and the object is easy to be subtracted.



(a) True object position

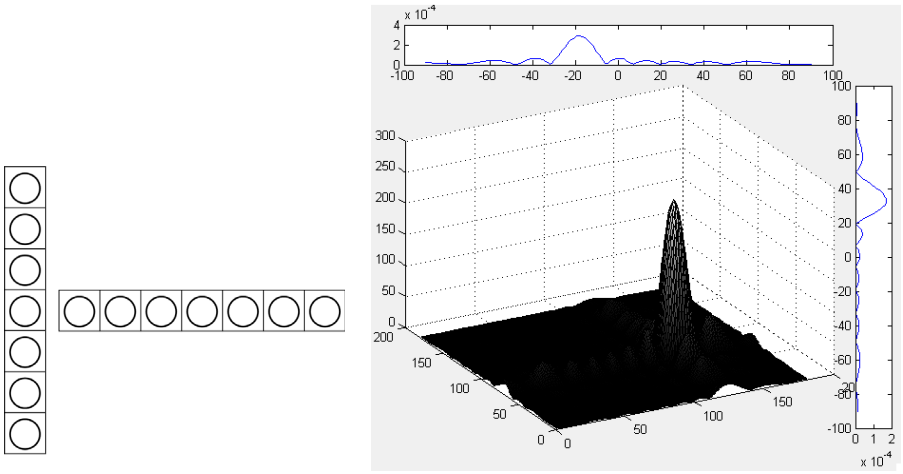


(b) Surface plot of beamforming image



(c) Contour plot of beamforming image

Figure 3.16: Beamforming full rectangular array simulation



(a) The sensor array layout **(b)** 2 dimensional beamforming

Figure 3.17: 2 dimensional beamforming

3.5.2 Two line multiplication beamforming

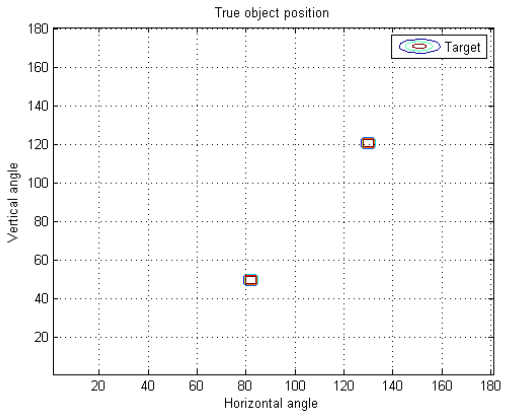
The full rectangular sensor array requires many sensors, especially when the resolution is high. One simple method to reduce the number of sensors is that two lines of sensors are used, where the horizontal linear sensor array detects the horizontal AoA, and the vertical linear sensor array detects the vertical AoA of the target.

The two lines of sensor array beamforming are processed separately and the results are obtained separately. By multiplication of the results of these two arrays, a two dimensional picture can be displayed, which is shown in Figure 3.17. The number of sensors in the simulation is 7 plus 7.

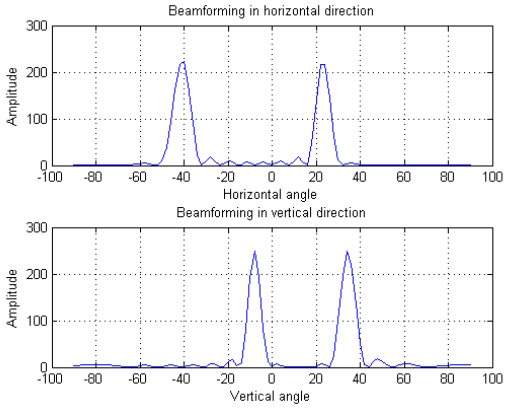
The advantage of this method is that the main lobe to side lobe ratio is very high, and the target can be easily subtracted from the background. The disadvantage is that when there is more than one object, the system has ambiguities. See Figure 3.18. In addition there is no way to detect the true objects just by the radar image.

3.5.3 Crossed array beamforming

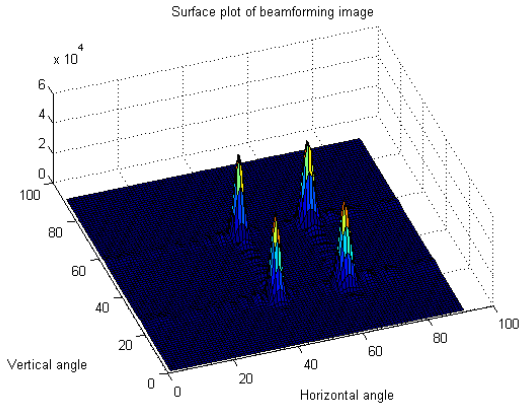
Another method is to use a crossed sensor array to detect the vertical and horizontal position. These two lines of sensors are in the shape of a cross and share



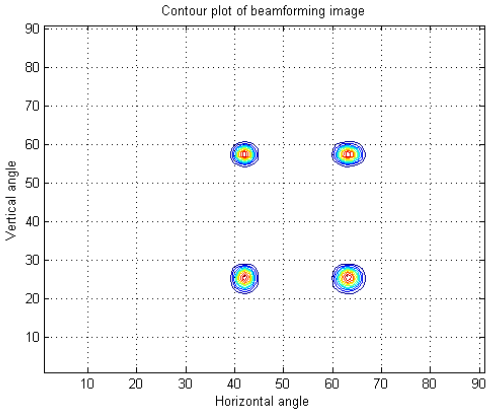
(a) True object position



(b) Beamforming in two directions

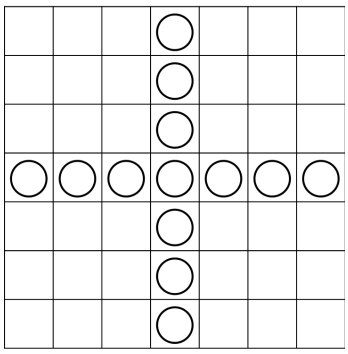


(c) Surface plot of beamforming image

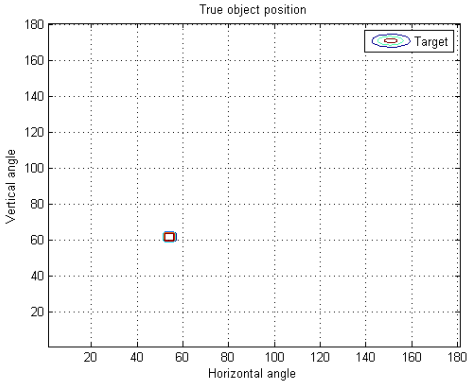


(d) Contour plot of beamforming image

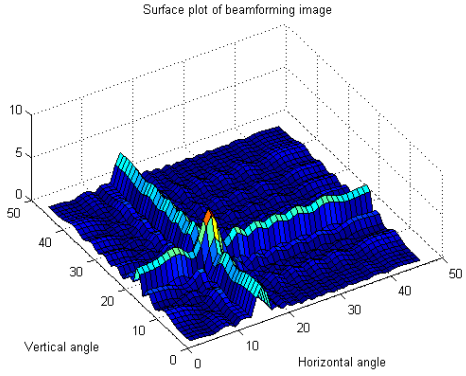
Figure 3.18: Two line multiplication beamforming



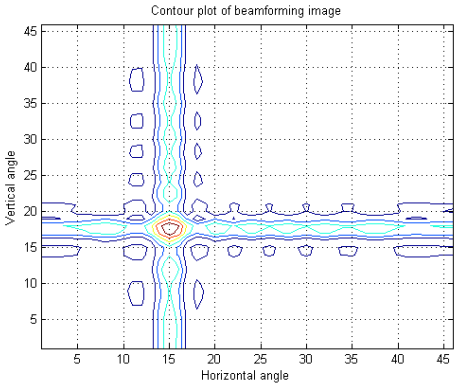
(a) The sensor array layout



(b) True object position



(c) Surface plot of beamforming image



(d) Contour plot of beamforming image

Figure 3.19: Crossed array beamforming, one object

one sensor in the middle.

This method works well when there is only one object which is shown in Figure 3.19. Compared with two line multiplication beamforming, the main lobe to side lobe ratio is not very high and there is some noise. The noise is in a similar shape to the layout of the receiver arrays.

If there are two objects, ambiguity also exists, as shown in Figure 3.20. The Cartesian coordinate system is used and the radar is allocated in the origin. The positions of the two targets are (5.5, 7.7, 9.9) (-9.9, -2.0, 9.9). All the targets in

the beamforming image have the similar side lobe shape with the layout of the sensor arrays.

If the position of the second target is shifted a small amount, there is a change to the phase from the second target to the center of the crossed sensor array. The new positions of the two targets are (5.5, 7.7, 9.9), (-9.9, -2, 10.9). There are no ambiguities, which is shown in Figure 3.21.

However, this method is not reliable because whether or not there are ambiguities depends on the phase difference of the targets to the center receiver sensor.

3.5.4 Other structure of 2D beamforming radar array

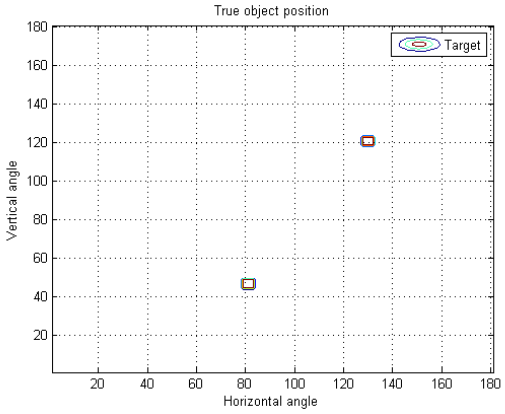
There are many ways to reduce the complexity. For example, the sensor arrays can be in the shape of "T", "+", "◇","X" and different type of checks. Here the performance of these structures of sensor arrays is tested.

In Figure 3.22, the sensor array is in the shape of a rectangle. There are still ambiguities. The performance is better than crossed sensor arrays but this method requires more sensors. The shape of beamforming results is similar to the shape of the sensor array layout.

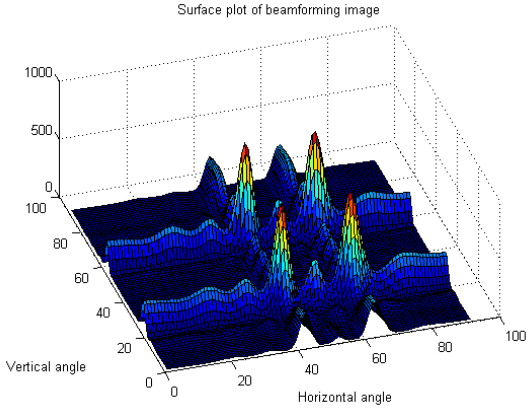
In Figure 3.23, the structure with 2 by 2 cells is tested. This structure is actually the combination of a crossed sensor array with a rectangular sensor array. The rate from the high peak to the low peak is at least three times; this can easily be subtracted from the background. So this situation is considered as there are no ambiguities.

3.5.5 Window function to reduce side lobe

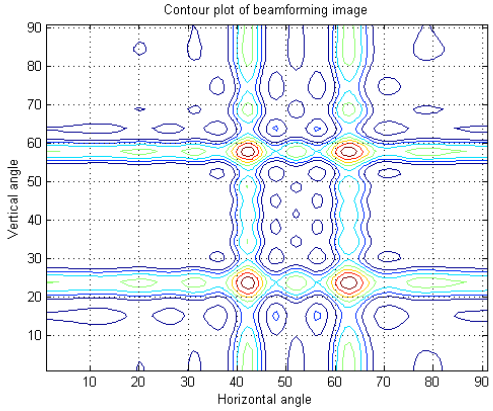
Similar to the window function in the linear sensor array, the one dimensional window function can be extended to a two dimensional window function to reduce the side lobes. According to experience in this study, Gaussian windows and Dolph-Chebyshev windows can be used. Figure 3.24 and Figure 3.25 show the simulation of crossed sensor array beamforming with Gaussian windows and Dolph-Chebyshev windows, separately. In Dolph-Chebyshev windows, the main lobe to side lobe ratio is set at 40.



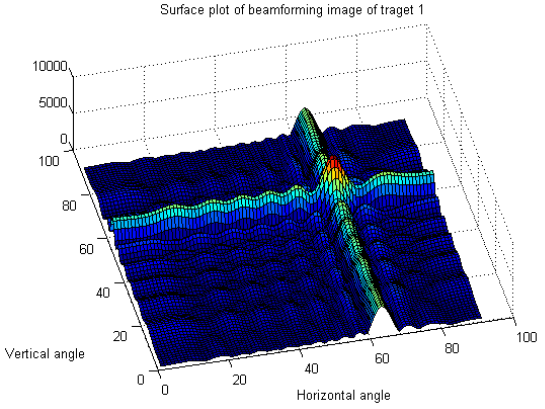
(a) True object position



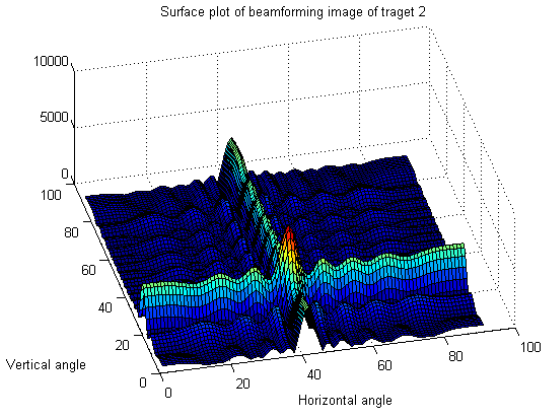
(b) Surface plot of beamforming image



(c) Contour plot of beamforming image

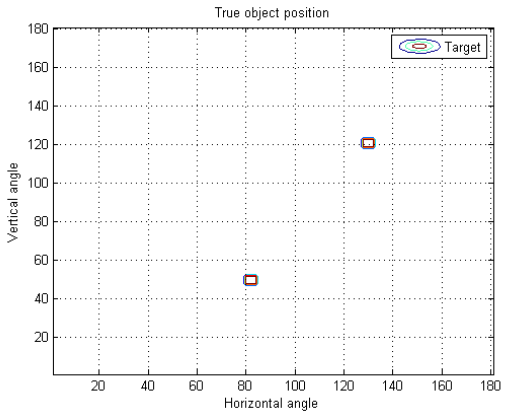


(d) Surface plot of beamforming image of target 1

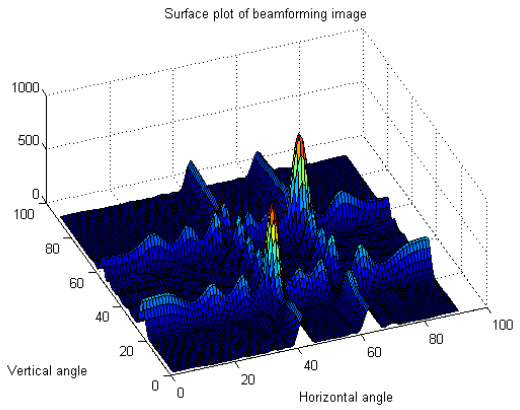


(e) Surface plot of beamforming image of target 2

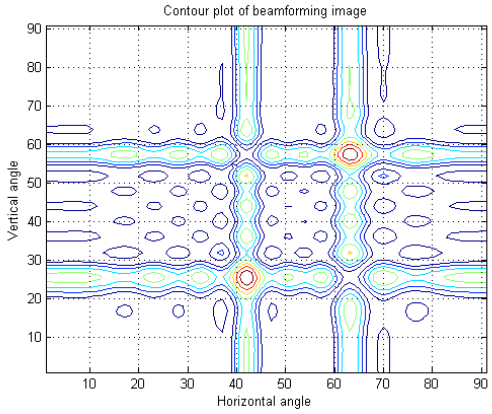
Figure 3.20: Two objects with ambiguities



(a) True object position

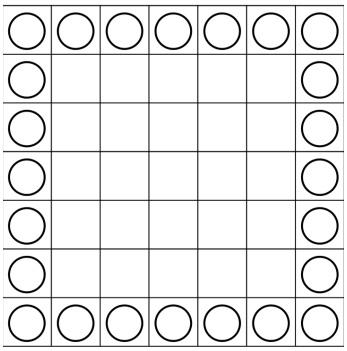


(b) Surface plot of beamforming image

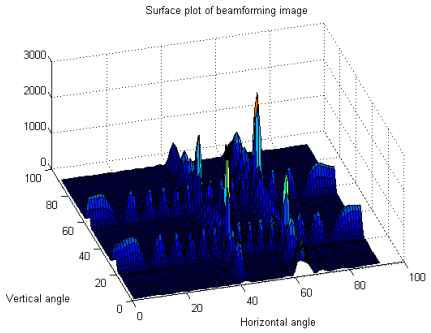


(c) Contour plot of beamforming image

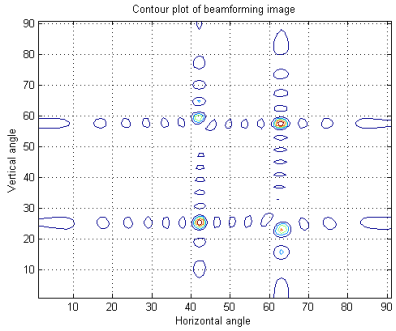
Figure 3.21: Two objects without ambiguities



(a) The sensor array layout

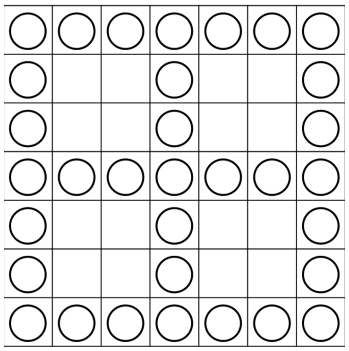


(b) Surface plot of the beamforming image

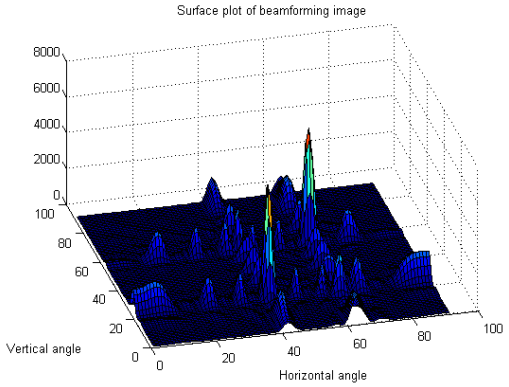


(c) Contour plot of the beamforming image

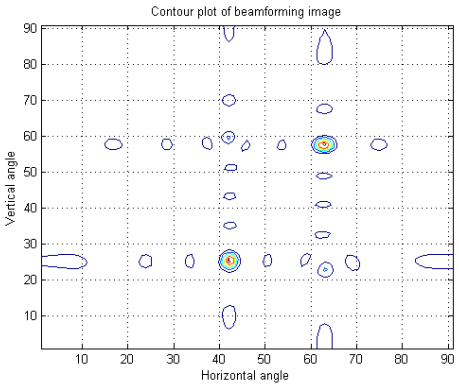
Figure 3.22: Rectangular sensor arrays



(a) The sensor array layout

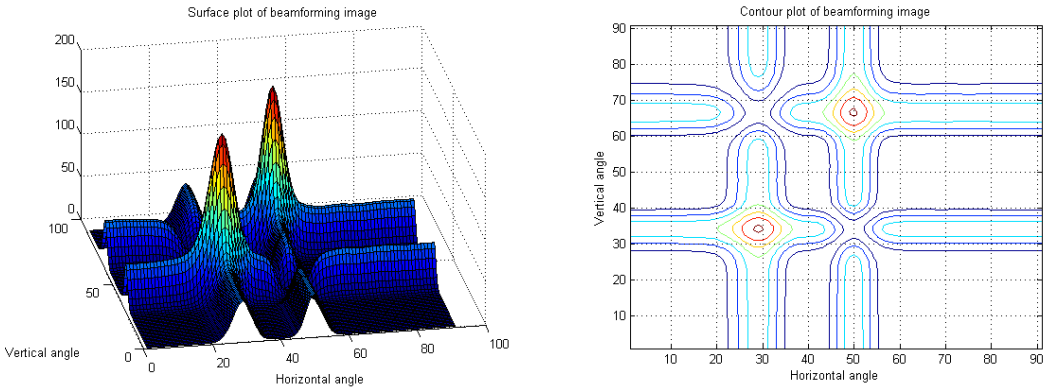


(b) Surface plot of beamforming image



(c) Contour plot of beamforming image

Figure 3.23: 2 by 2 cells arrays



(a) Surface plot of the beamforming image

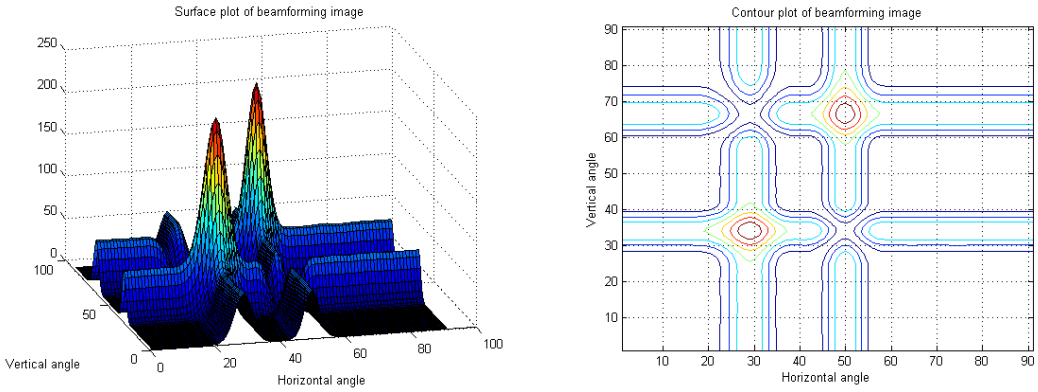
(b) Contour plot of the beamforming image

Figure 3.24: Crossed array beamforming simulation with Gaussian window

To make the picture more clear and reflect the optical image, the picture is from -60º to 60º in the horizontal and -45º to 45º in the vertical.

3.5.6 Performance when adding noise

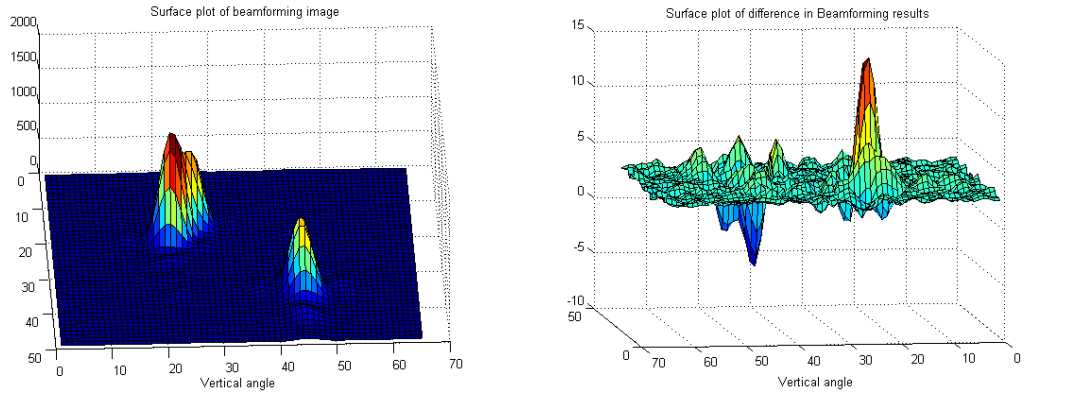
To test whether the system is robust or not, white Gaussian noise is added to the received signals of each channel. Then the beamforming images with and without noise are calculated to find out the root mean square errors (RMSE). The results are shown from Figure 3.26 to Figure 3.29.



(a) Surface plot of the beamforming image

(b) Contour plot the beamforming image

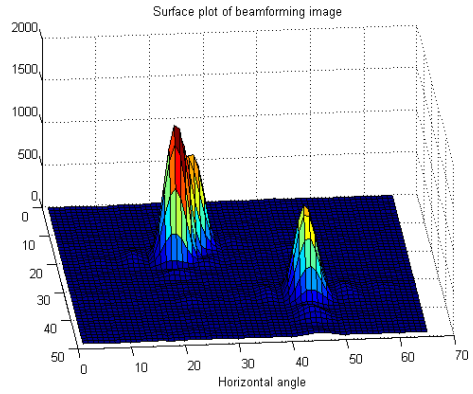
Figure 3.25: Crossed array beamforming simulation with Dolph-Chebyshev window



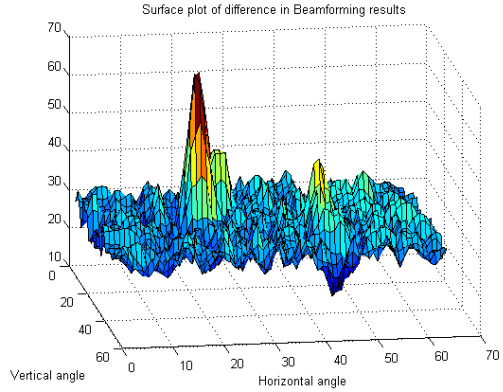
(a) Surface plot of the beamforming image

(b) Surface plot of the difference in beamforming results

Figure 3.26: Noise analysis, SNR = 10dB, left figure is the result and right figure is the difference

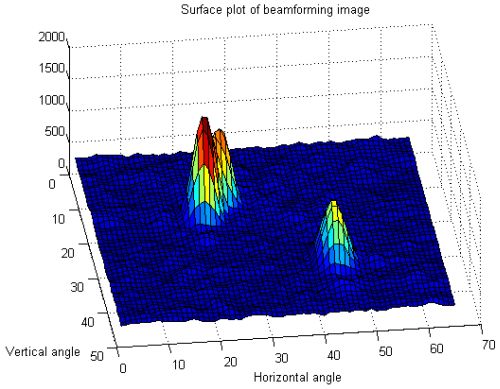


(a) Surface plot of the beamforming image

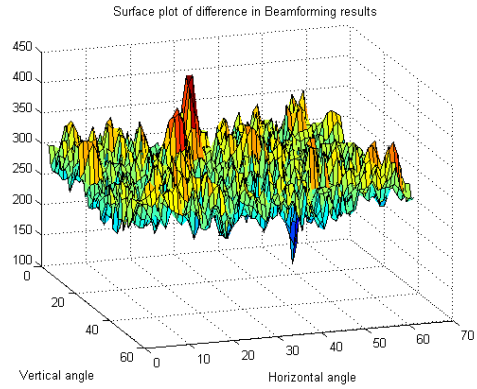


(b) Surface plot of the difference in beamforming results

Figure 3.27: Noise analysis, SNR = 0dB, left figure is the result and right figure is the difference

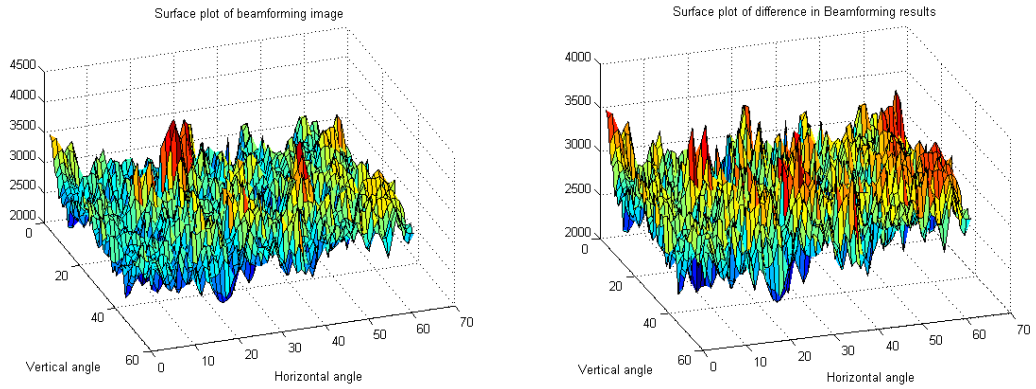


(a) Surface plot of the beamforming image



(b) Surface plot of the difference in beamforming results

Figure 3.28: Noise analysis, SNR = -10 dB, left figure is the result and right figure is the difference



(a) Surface plot of the beamforming image

(b) Surface plot of the difference in beamforming results

Figure 3.29: Noise analysis, SNR = -20 dB, left figure is the result and right figure is the difference

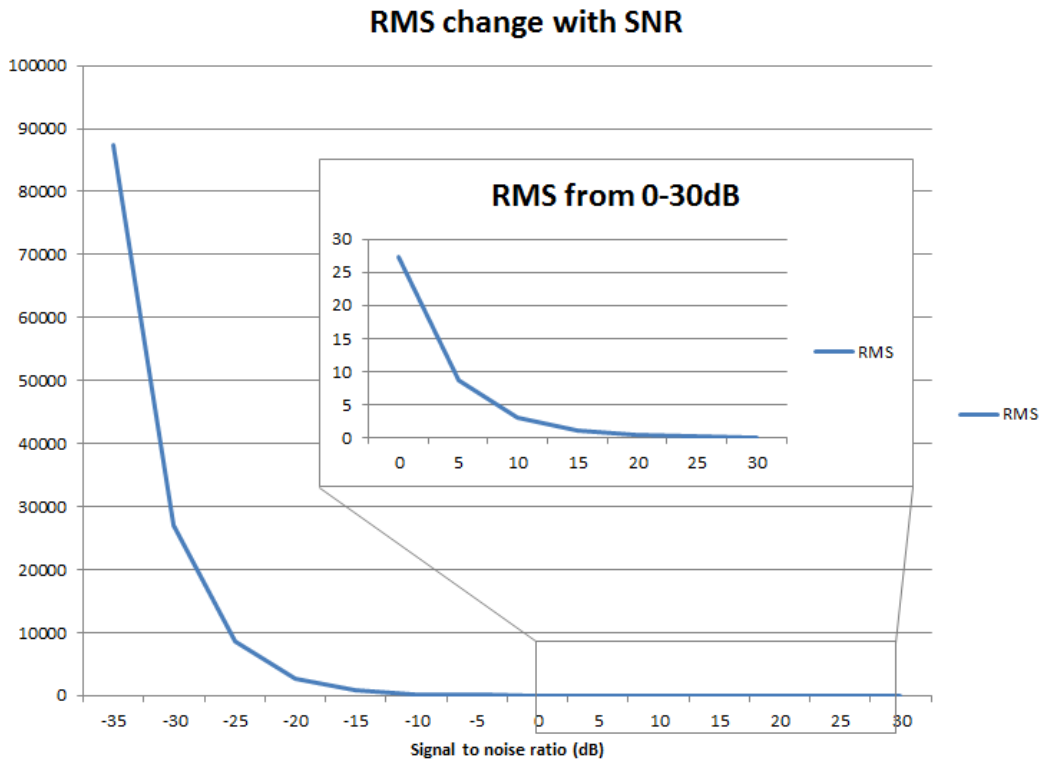


Figure 3.30: RMS error analysis with SNR

Here the beamforming results are quite robust even when the signal to noise (SNR) is lower than -10 dB. When the SNR is -20dB, the target cannot be detected. How the RMSE changes with the SNR is also tested which is shown in Figure 3.30. When the SNR is higher than -10dB, the RMSE is very small.

3.5.7 Removal of ambiguities

There are many ways to remove the ambiguities: 1) the pulsed beamforming method which will be described in the following section can separate the targets when the ranges of the objects are different. 2) the Doppler beamforming method which will be described in chapter 4 can separate the targets when the velocities of objects are different. 3) the captured radar image can be fused with the optical image which will be studied in chapter 5. A decision making algorithm can be applied to determine the possibility that there is an object or not.

3.6 Distance detection

3.6.1 Triangular method

Beamforming can detect the AoA of the targets. To detect the distance between the targets and radars, some methods have been studied. The first method is the triangular method. Two radars can be laid with some distance apart and facing the target. The principle of triangulation can be used to estimated the distance according to the position difference and the angle difference. This method is in popular use, such as in a sports playground. The radar needs to be settled in advance and the position of the radar should be known. The principle is shown in Figure 3.31.

It is supposed that the radar position is (r_{x1}, r_{y1}), (r_{x2}, r_{x2}) and target true position is (t_x, t_y) in Cartesian coordinates. Therefore $\tan\theta_1 = \frac{t_y - r_{y1}}{t_x - r_{x1}}$ and $\tan\theta_2 = \frac{t_y - r_{y2}}{t_x - r_{x2}}$. Then

$$\begin{aligned} t_x &= \frac{r_{y1} - r_{y2} - r_{x1}\tan\theta_1 + r_{x2}\tan\theta_2}{-\tan\theta_1 + \tan\theta_2} \\ t_y &= \frac{r_{y1}\tan\theta_2 - r_{y2}\tan\theta_1 - r_{x1}\tan\theta_1\tan\theta_2 + r_{x2}\tan\theta_1\tan\theta_2}{-\tan\theta_1 + \tan\theta_2} \end{aligned} \tag{3.32}$$

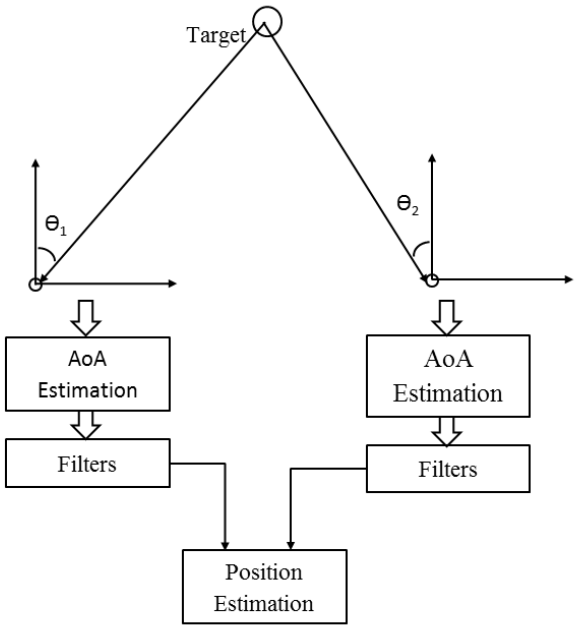


Figure 3.31: Triangular calculation

In other ways the equation can be presented as follows

$$\begin{bmatrix} -\tan\theta_1 & 1 \\ -\tan\theta_2 & 1 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} r_{y1} - r_{x1}\tan\theta_1 \\ r_{y2} - r_{x2}\tan\theta_2 \end{bmatrix} \tag{3.33}$$

In the format

$$Ax = B \tag{3.34}$$

as A and B are known,

$$x = A^{-1}B \tag{3.35}$$

In a three dimensional location system, at least three radars are required. Especially, one radar should be able to detect both the horizontal and vertical AoA.

The shortcoming of this method is that the distance accuracy highly relates to the accuracy of the AoA. Also, if there is more than one target, in each sensor, more than two objects can be detected, and it is difficult to relate the target in two radars. So, there are ambiguities.

3.6.2 Pulsed mode to detect distance

The other method is to use the pulsed beamforming method, which is shown in Figure 3.33. To simplify the structure, 1 transmitter and 4 receivers can be taken as an example. The transmitter periodically sends a pulse which contains a series of sinusoidal waves. The receivers sample the signals and count the time delay which is the Time of Flight (ToF) of the signals. According to the beamforming processing settings in this thesis, the adjacent sensor space is $\lambda/2$, the phase differences between adjacent sensors are less than one cycle.

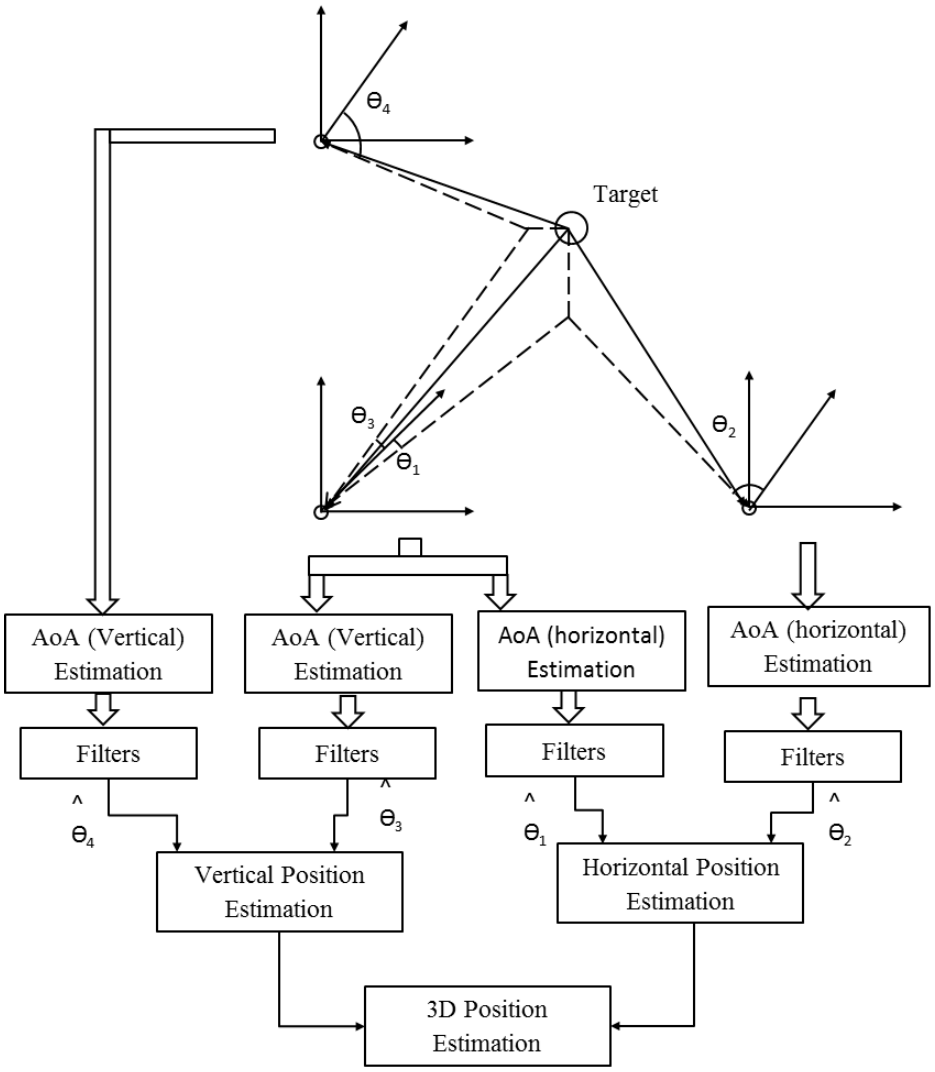


Figure 3.32: Three dimensional triangular calculation

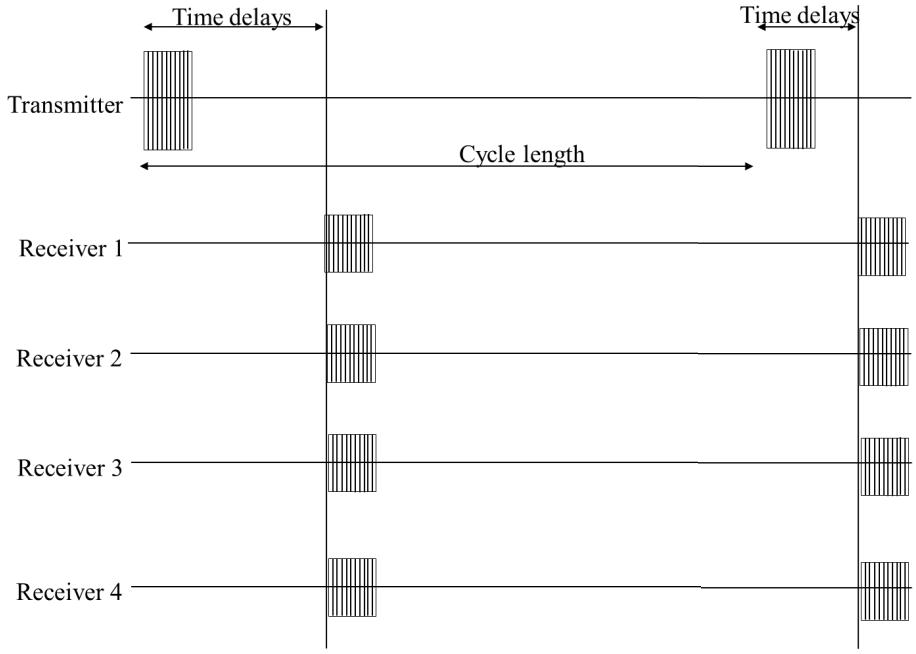


Figure 3.33: Pulsed radar

The range of the target can be calculated by

$$D = \left(c \cdot \left(t_1 + t_2 + t_3 + t_4\right)/4\right)/2, \tag{3.36}$$

where c is the speed of sound. t_1, t_2, t_3 and t_4 are the time delays of each receivers.

If there are many sensors, to save computation time, only the time delays of the first and last receivers are calculated:

$$D \quad = \left(c \cdot \left(t_1 + t_4\right)/2\right) \quad /2 \tag{3.37}$$

When there is a pulse, the beamforming processing starts. The result is a frame of radar image. As the beamforming calculation requires to sample N time series, the samples may have length as 128, 256 samples or more for FFT processing. When the range difference between two targets is too small, they are shown in the same frame. In this situation, they can be roughly considered in the same range, and the distance resolution is

$$r \quad = \quad c \times N/f_s \tag{3.38}$$

where c is the speed of sound, f_s is the sample rate, and N is the FFT series length. In the proposed radar in this thesis, the sample rate is 160kHz, if the N is 256 samples, the resolution of distance is 0.544 meters; this means if the targets with radial distance difference less than half a meter, they are in the same frame.

So the maximum frame rate f_{frame} is

$$f_{frame} \quad = \quad f_s \cdot Q \cdot l_s \tag{3.39}$$

where f_s is the sample frequency, Q is the number of samples per second, l_s is cycle number in each pulse.

Table 3.3: Possible outcomes for the Detection Algorithm

Detection	Truth	False
	No Signal Present	Signal Present
No signal present	Correct decision	Missed detection
Signal present	False alarm	Correct decision

3.6.3 Signal verification

The characteristics of sensors used as receivers in this thesis are similar to the characteristics of microphones. The sensors have lower attenuation at 40kHz and higher attenuation at other frequencies. Therefore the sensor can automatically filter some noise. However, when the hardware was tested in the real environment, some sound sources could be captured by the receivers. The noise could cause a false alarm, which is shown in Table 3.3.

One way to solve this problem is by using matched filters to verify whether the received signal is the transmitted signal or not. The noise can be filtered or reduced to a very low amplitude after using the matched filter.

The matched filter uses a known signal as a template to detect the presence of signals similar to the template in an unknown signal. It is a linear filter when SNR reaches the maximum.

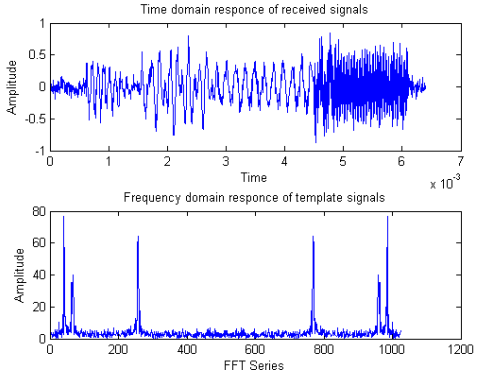
The matched filter expression is:

$$H(f) = S*(f) \tag{3.40}$$

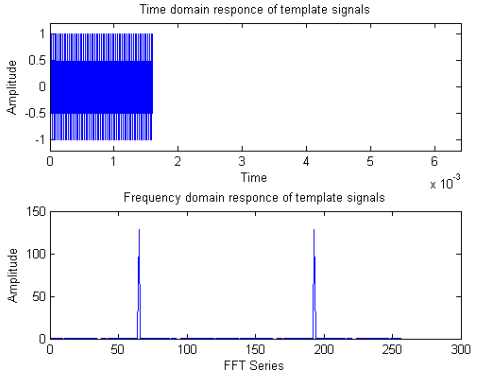
The matched filter is simulated in Figure 3.34. The results show that this filter works well.

3.7 Relationship with optical image

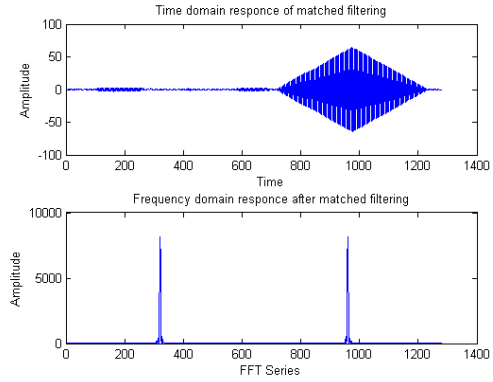
In the previous section, the method which can create a pulsed beamforming image is described. The purpose of this image layer is to be fused with the optical image to enhance target tracking. Before fusion with the optical image, the relationship between the proposed radar image and the optical image needs to be investigated.



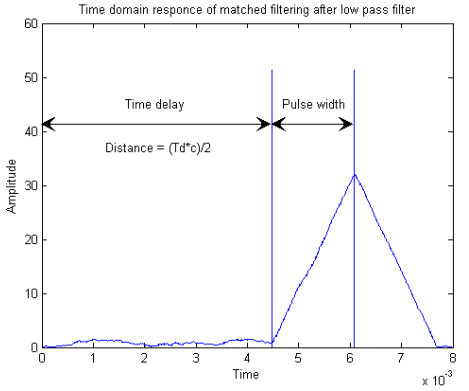
(a) Spectrum of received signals



(b) Spectrum of the template signals

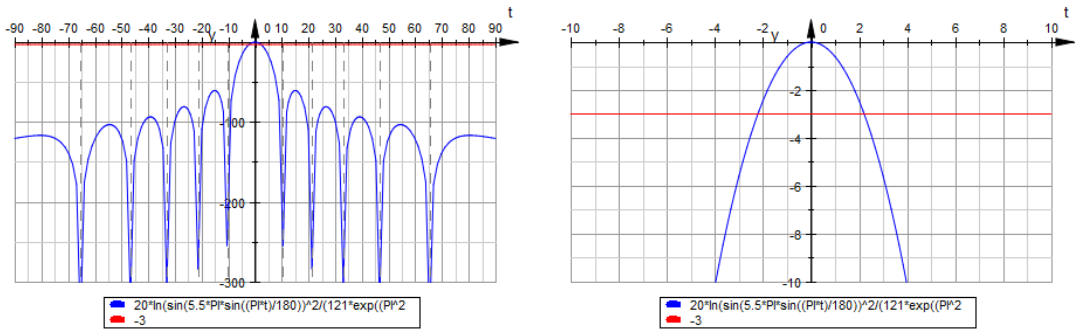


(c) Results of matched filtering



(d) Time domain response of matched filtering after low pass filter

Figure 3.34: Simulation on the matched filter



(a) Beam width **(b)** Beam width in detail

Figure 3.35: Beamforming width

3.7.1 Relationship of sensor number and resolution

In signal processing, bandwidth usually equals -3dB, which is half the power of the main lobe. A one dimensional beamforming linear sensor array can detect objects from 0^o to 180^o. If the beamwidth is lower than 4^o, more than 45 horizontal pixels can be detected in a radar image.

If the carrier frequency is a fixed value, in a Discrete Fourier Transform, the resolution is only related to the number of sensors . In Figure 3.35, Matlab is used to calculate the beamwidth and is set at -3 dB as the edge of the beamwidth. In Table 3.4, the relationship between number of sensors and the resolution can be found. The angle is roughly $R(n) = 3 * R(3)/n$, where n is the number of sensors. According to the table, when the beamwidth is lower than 4^o, 13 sensors are required, and when the beamwidth is lower than 2^o, 25 sensors are required.

In vertical one dimensional beamforming, the relationship of sensor number and resolution is the same.

The more sensors that are used in the beamforming processing, the more accurate the results are, and if the bandwidth is lower and the main lobe to side lobe ratio is higher. However, the hardware cost and the computation and memory consumption will also increase sharply.

Table 3.4: Relationship between sensor number and resolution

Number of sensors	Resolution (Degree)	Number of sensors	Resolution (Degree)
3	16.1	4	12.1
5	9.7	6	8.1
7	6.9	8	6.0
9	5.4	10	4.8
11	4.4	12	4.0
13	3.7	14	3.5
15	3.2	16	3.0
17	2.85	18	2.69
19	2.54	20	2.42
21	2.30	22	2.20
23	2.10	24	2.01
25	1.93	26	1.86

3.7.2 Frame rate

In an optical image, the most popular frame rate used is 24, 25 or 30 frames per second. In this system, frame rate varies with system modes and tracking requirements.

In frequency domain beamforming, if the processing speed is high enough, the frame rate is related to the sample frequency and sample series length. In the continuous wave beamforming case, the sample length for FFT determines the maximum frame rate.

$$F = f_s/N \tag{3.41}$$

where f_s is the sample frequency and N is the FFT series length. If a 40kHz ultrasonic wave is transmitted and the real time data is captured, and sample series length is 512, the frame rate is around 79fps, which is relatively high for detecting moving objects. The sample frequency can be increased or the Fourier series length can be reduced to increase the frame rate.

If the pulsed mode beamforming is set, the frame rate is determined by the pulsed cycle length. Inside each frame, the beamforming can be calculated continuously, and the time delay determines the range.

3.7.3 Pixel reflection

The VGA resolution is defined as 640 by 480. For an APS-C camera, the relationship between angle of view and focal length is shown in Table 3.5.

Table 3.5: Relationship between angle of view and focal length

Lens focal length	Angle of view in APS-C camera
10mm	107
18mm	74
50mm	30
100mm	15
250mm	6

A surveillance camera usually has a wide angle of view which means a short focal length. To compare the optical image pixels with the radar image pixels, the horizontal angle is supposed to be 128º and the vertical angle is supposed to be 96º. If 24 by 24 sensors are used, the radar image resolution is 64 by 48. That means around 10 by 10 pixels in an optical image is related to a pixel in a radar image.

3.7.4 Coordinate system

Many coordinate systems have been used to present object position. The Cartesian coordinate system is the most widely used system in which three mutually perpendicular planes are chosen and the three coordinates of a point are the signed distances to each of the planes. Other types of systems include the Polar coordinate system. In a radar system with beamforming processing, the AoA is used to present the location of objects both in the horizontal and the vertical directions. Therefore a non-conventional coordinate system [65] with the AoA as the main parameters is used. The system is shown in Figure 3.36.

In this coordinate system, the position of point P on the spherical surface is characterized by angle α, β and distance R. α is the angle between the projection of the vector R to the plane $x0z$ and axis $0z$, while β is the angle between the projection of the vector R to the plane $y0z$ and axis $0z$, and R is the directed distance from the origin of the coordinate system to the point P. The transform from $P\,(x_p, y_p, z_p)$ to $P\,(\alpha, \beta, R)$ is given as follows.

$$\alpha = \arctan \frac{x_p}{z_p} \tag{3.42}$$

$$\beta = \arctan \frac{x_p}{z_p} \tag{3.43}$$

$$R = \sqrt{x_p^2 + y_p^2 + z_p^2} \tag{3.44}$$

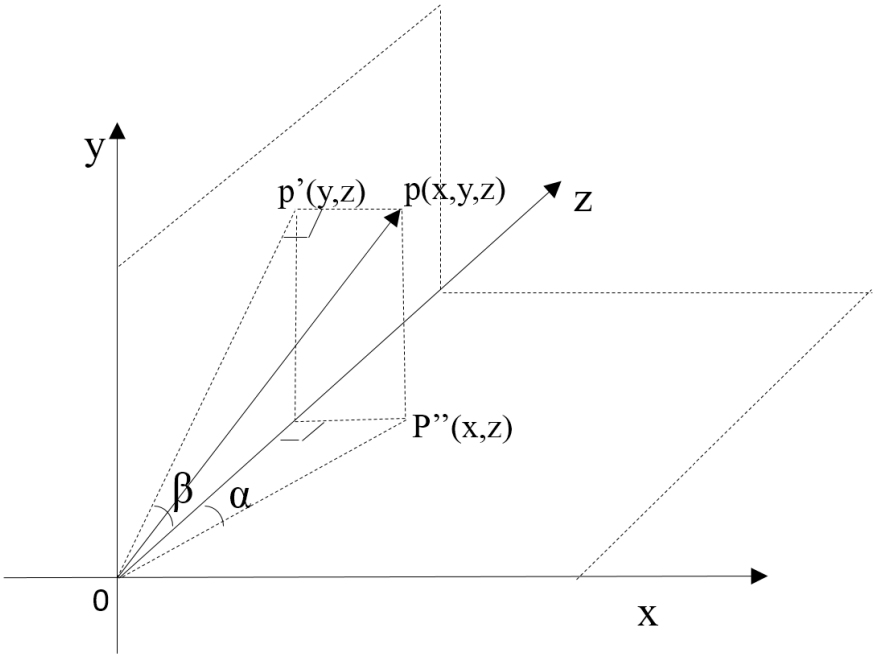


Figure 3.36: The coordinate system

Inversely, the transform from $P(\alpha, \beta, R)$ to $P(x_p, y_p, z_p)$ is given as follows.

$$x_p = \frac{R \tan \alpha}{\sqrt{1 + \tan^2 \alpha + \tan^2 \beta}} \tag{3.45}$$

$$y_p = \frac{R \tan \beta}{\sqrt{1 + \tan^2 \alpha + \tan^2 \beta}} \tag{3.46}$$

$$z_p = \frac{R}{\sqrt{1 + \tan^2 \alpha + \tan^2 \beta}} \tag{3.47}$$

3.8 Example of fourth image layer

3.8.1 Simulation setting

In the simulation, one person and two birds are supposed to be captured by the designed radar. The supposed optical image is shown in Figure 3.37

The resolution of the image is 640 by 480. Cartesian coordinate system can be used to measure the position of objects.

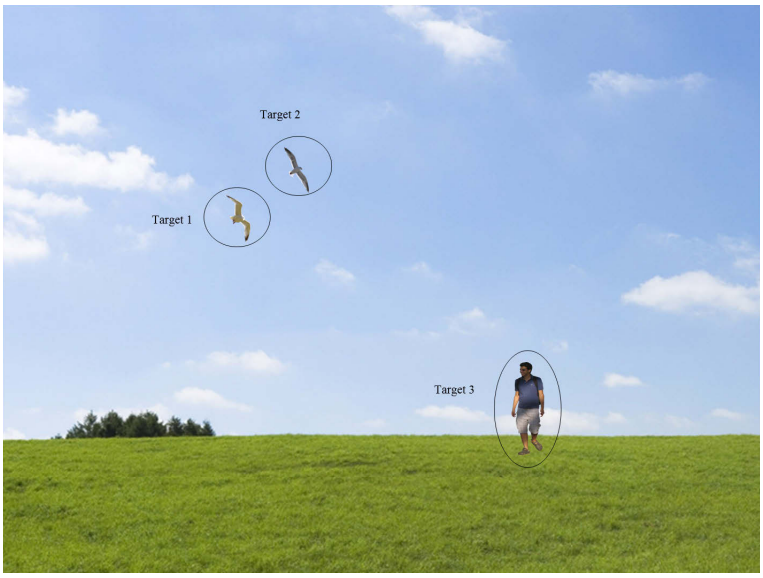


Figure 3.37: The supposed optical image

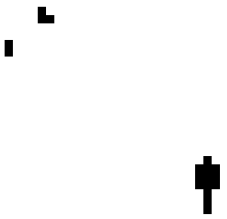


Figure 3.38: Image in radar resolution

The left and top of the image is set to be the origin.

The distance of the human is supposed to be 7 meters, and the distance of the left bird and right bird is 10 meters and 13 meters, respectively. So the centroid position of the person is (3.26, 2.71, 5.92), the centroid position of the left bird is (-4.32, -2.02, 9.00), and the centroid position of the right bird is (-3.19, -4.58, 11.91).

According to the simulation in the section above, the supposed minimum detection angle is 2^{o}. If the optical image has the same resolution, the image should look as in **??**

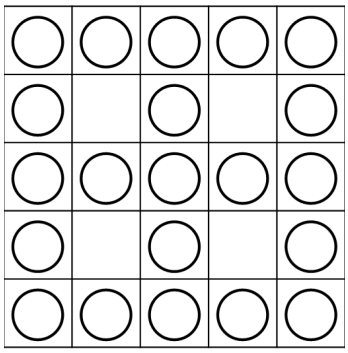
The resolution of **??** is 64 by 48 and only 18 pixels can reflect acoustic signals. In the image plane, the position of the reflection point is shown in Table 3.6.

3.8.2 Simulation on sparse rectangular array

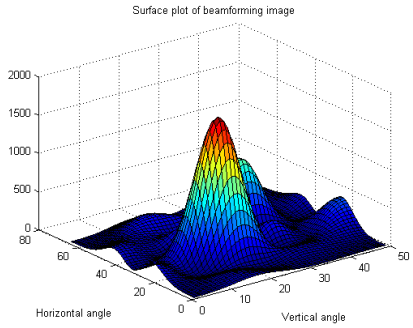
In the above sections, the sparse rectangular arrays have been shown to have the best balance between the sensor number and ambiguities. Here some types of sparse rectangular sensor arrays are used to test the resolution according to the sensor number and the attenuation ratio between the target and the ambiguities. In order to compare the relationship between the sensor array structure and the ambiguity, as well as the relationship between the sensor number and the resolution, all the targets are supposed to be in the same range. The results are shown from Figure 3.39 to Figure 3.49:

Table 3.6: Reflection points in Figure 3.38

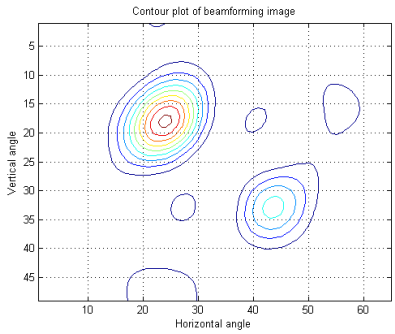
Object	Reflection point in image plane	Distance	Position in Cartesian coordinates
Left bird	(21,18)(21,19)	10m	(-3.91 -2.25 8.97)(-3.91 -1.91 9.04)
Right bird	(25,14)(25,15)(26,15)	13m	(-3.36 -4.66 11.72)(-3.36 -4.23 11.87) (-2.92 -4.23 11.98)
Human	(44,33)(44,34)(44.35) (45,32)(45,33)(45,34) (45,35)(45,36)(45,37) (45,38)(46,33)(46,34) (46, 35)	7m	(2.73 2.05 6.16) (2.73 2.28 6.09)(2.73 2.51 6.02)(2.96 1.81 6.13)(2.96 2.05 6.07)(2.96 2.28 6.00)(2.96 2.51 5.92)(2.96 2.73 5.84)(2.96 2.96 5.75)(2.96 3.18 5.65)(3.18 2.05 5.97)(3.18 2.28 5.90)(3.18 2.51 5.82)



(a) The sensor array layout

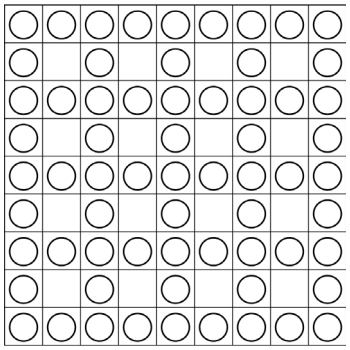


(b) Surface plot of the beamforming image

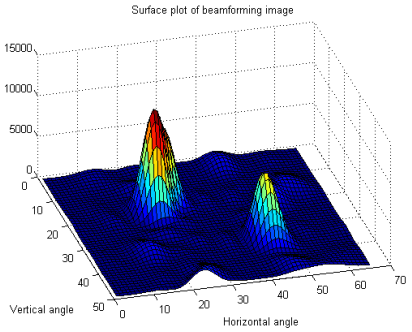


(c) Contour plot of the beamforming image

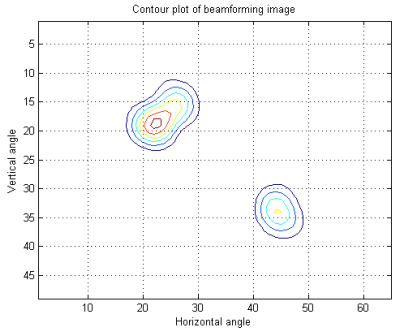
Figure 3.39: 5 by 5, hole size is 1 by 1



(a) The sensor array layout

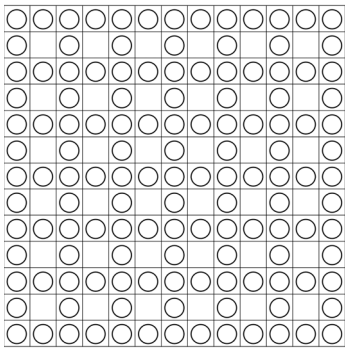


(b) Surface plot of the beamforming image

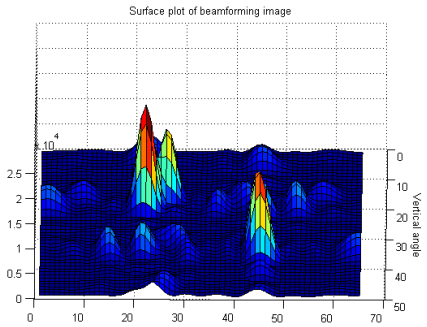


(c) Contour plot of the beamforming image

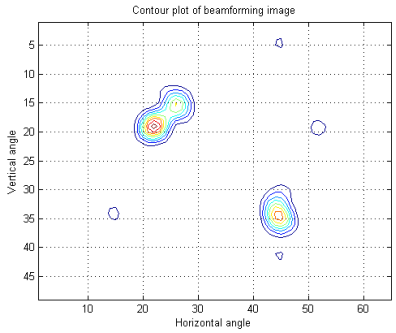
Figure 3.40: 9 by 9, hole size is 1 by 1



(a) The sensor array layout

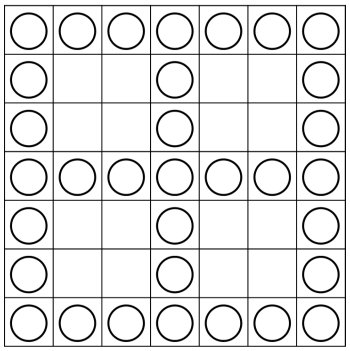


(b) Surface plot of the beamforming image

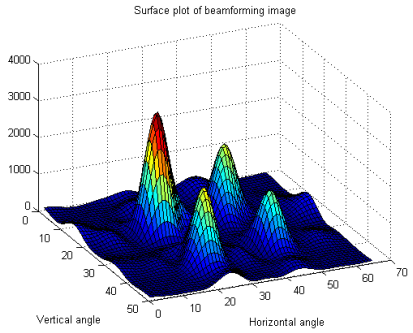


(c) Contour plot of the beamforming image

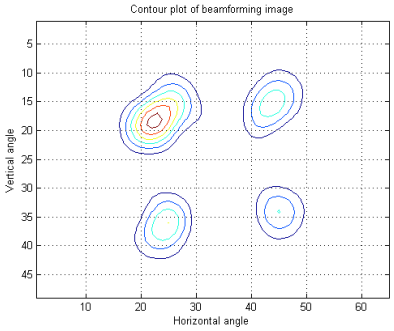
Figure 3.41: 13 by 13, hole size is 1 by 1



(a) The sensor array layout

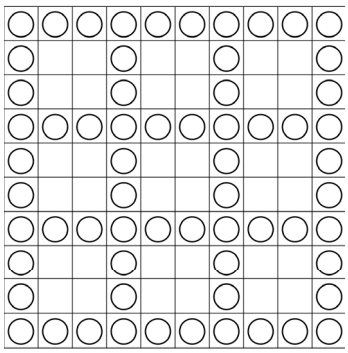


(b) Surface plot of the beamforming image

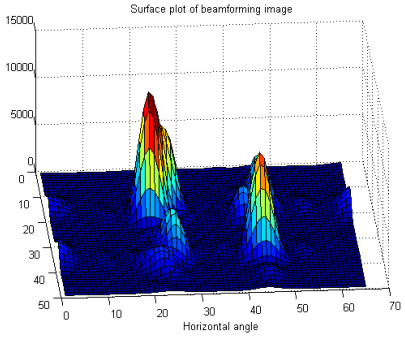


(c) Contour plot of the beamforming image

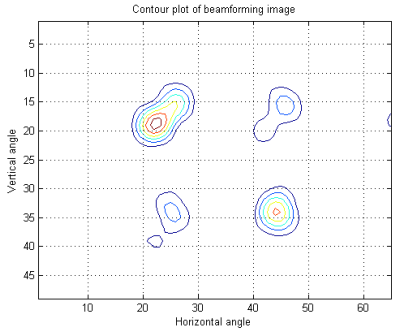
Figure 3.42: 7 by 7, hole size is 2 by 2



(a) The sensor array layout

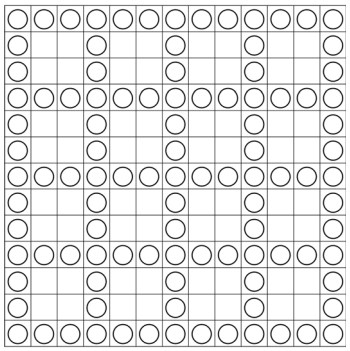


(b) Surface plot of the beamforming image

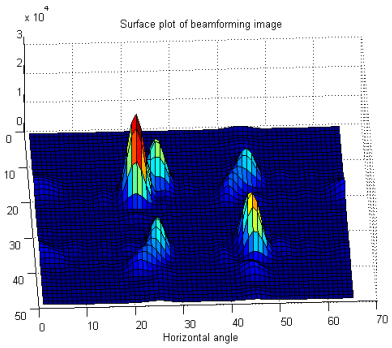


(c) Contour plot of the beamforming image

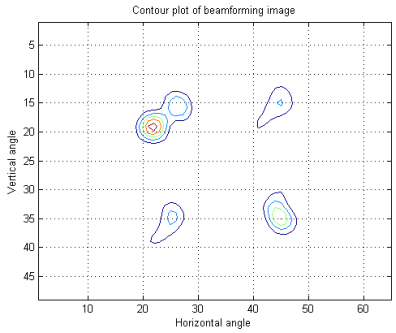
Figure 3.43: 10 by 10, hole size is 2 by 2



(a) The sensor array layout

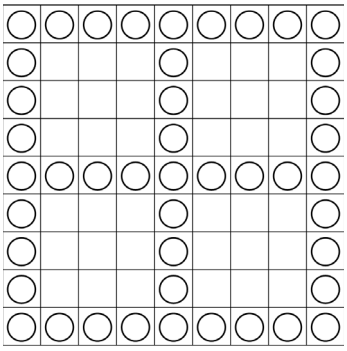


(b) Surface plot of the beamforming image

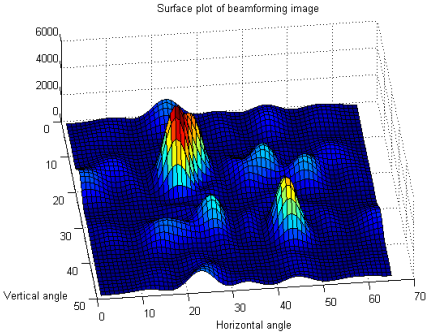


(c) Contour plot of the beamforming image

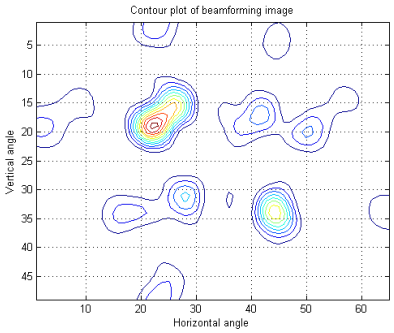
Figure 3.44: 13 by 13, hole size is 2 by 2



(a) The sensor array layout

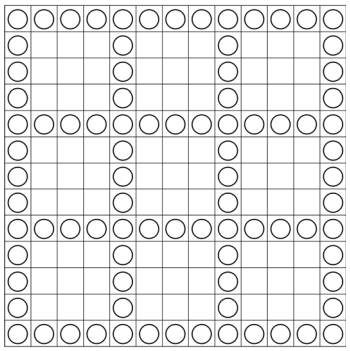


(b) Surface plot of the beamforming

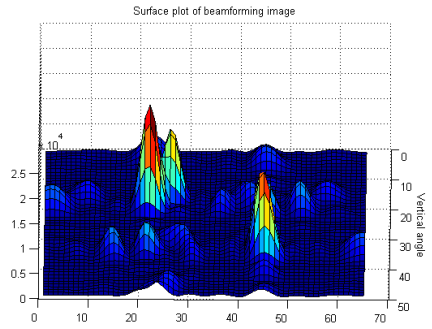


(c) Contour plot of the beamforming

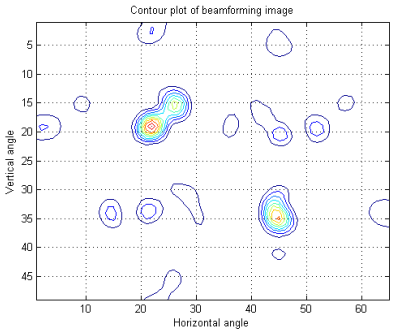
Figure 3.45: 9 by 9, hole size is 3 by 3



(a) The sensor array layout

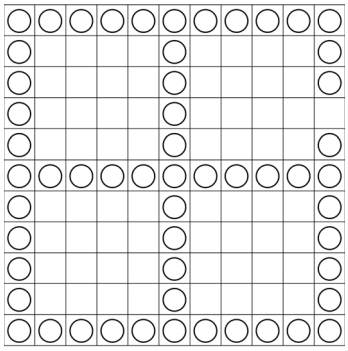


(b) Surface plot of the beamforming image

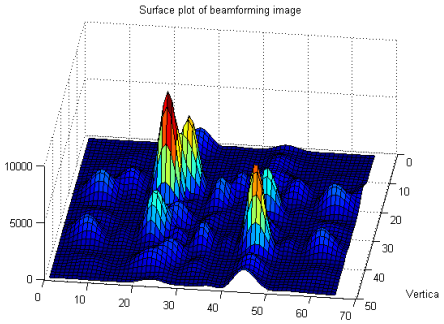


(c) Contour plot of the beamforming image

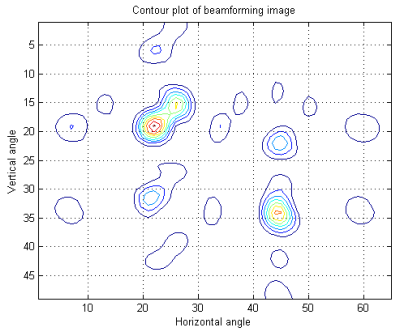
Figure 3.46: 13 by 13, hole size is 3 by 3



(a) The sensor array layout



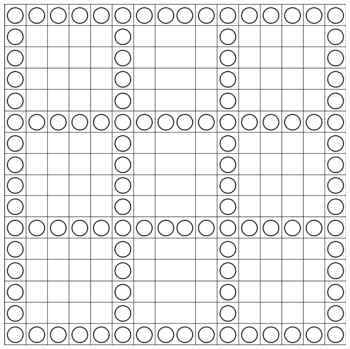
(b) Surface plot of the beamforming



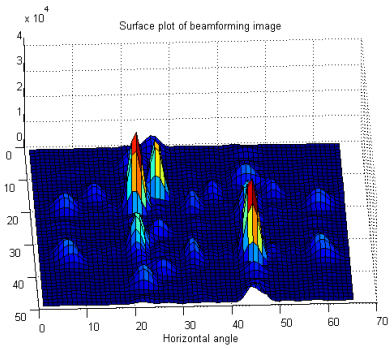
(c) Contour plot of the beamforming

Figure 3.47: 11 by 11, hole size is 4 by 4

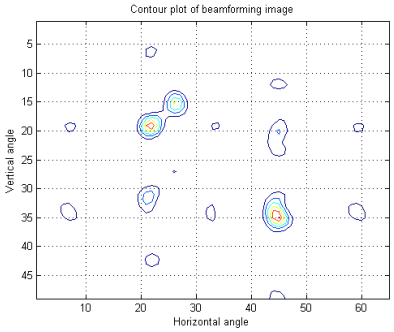
According to the simulation, the smaller the hole size is, the fewer ambiguities the image will have. However even when the hole size is 4 by 4, the target can



(a) The sensor array layout

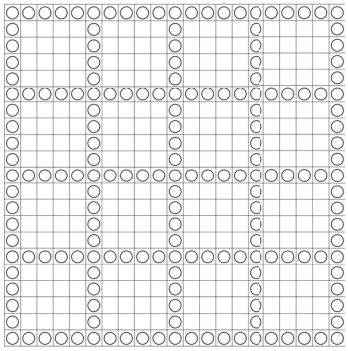


(b) Surface plot of the beamforming image

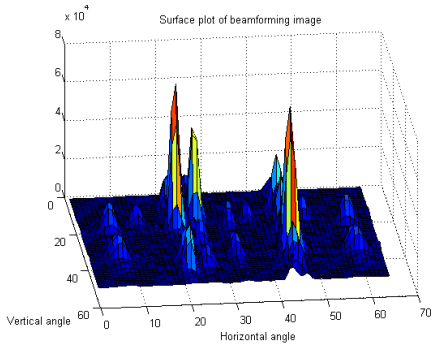


(c) Contour plot of the beamforming image

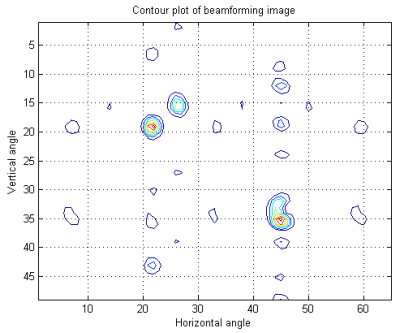
Figure 3.48: 16 by 16, hole size is 4 by 4



(a) The sensor array layout



(b) Surface plot of the beamforming image



(c) Contour plot of the beamforming image

Figure 3.49: 21 by 21, hole size is 4 by 4

be separated from the ambiguities. When the sensor number reaches 21 by 21, the two birds can be separated. If the structure of 21 by 21 sensors with 4 by 4 holes are used, compared with the full rectangular sensor array, more than half of the sensors are saved, and the performance is not affected much by the removed sensors.

3.9 Proposed algorithms

According to the analysis and simulation results list above, the design and algorithms most suitable for the applications are proposed as follows.

3.9.1 Sensor array structure

The sparse rectangular sensor arrays achieve the balance between the hardware resources, computation load and memory consumption. In some cases, where there are fewer than 3 targets, and the range differences between targets are clear, two line multiplication beamforming can be used as this method can achieve a high main lobe to side lobe ratio. The target with different range can be separated by pulsed mode then they will not be captured in the same beamforming image. In chapter 5, the method to reduce ambiguity by fusion of the optical image will be described.

Therefore the structure of the sensor array layout depends on the requirement of the application and the budget of the system. In a clear environment without many targets, two line multiplication can be chosen. Also, in some more complicated environments, the sparse rectangular array is the most suitable structure.

3.9.2 Procedure to construct the pulsed beamforming image

Window function can reduce the main lobe to side lobe ratio; in particular, the Dolph-Chebyshev window has the best performance. Frequency domain beamforming has high resolution especially when high resolution images are required. The pulsed mode can be used to detect the range of the target. To remove any false alarms, the matched filter is used to verify the signal, the matched amplitude should be higher than a specified value calculated by the time of flight. The procedure is shown in Figure 3.50.

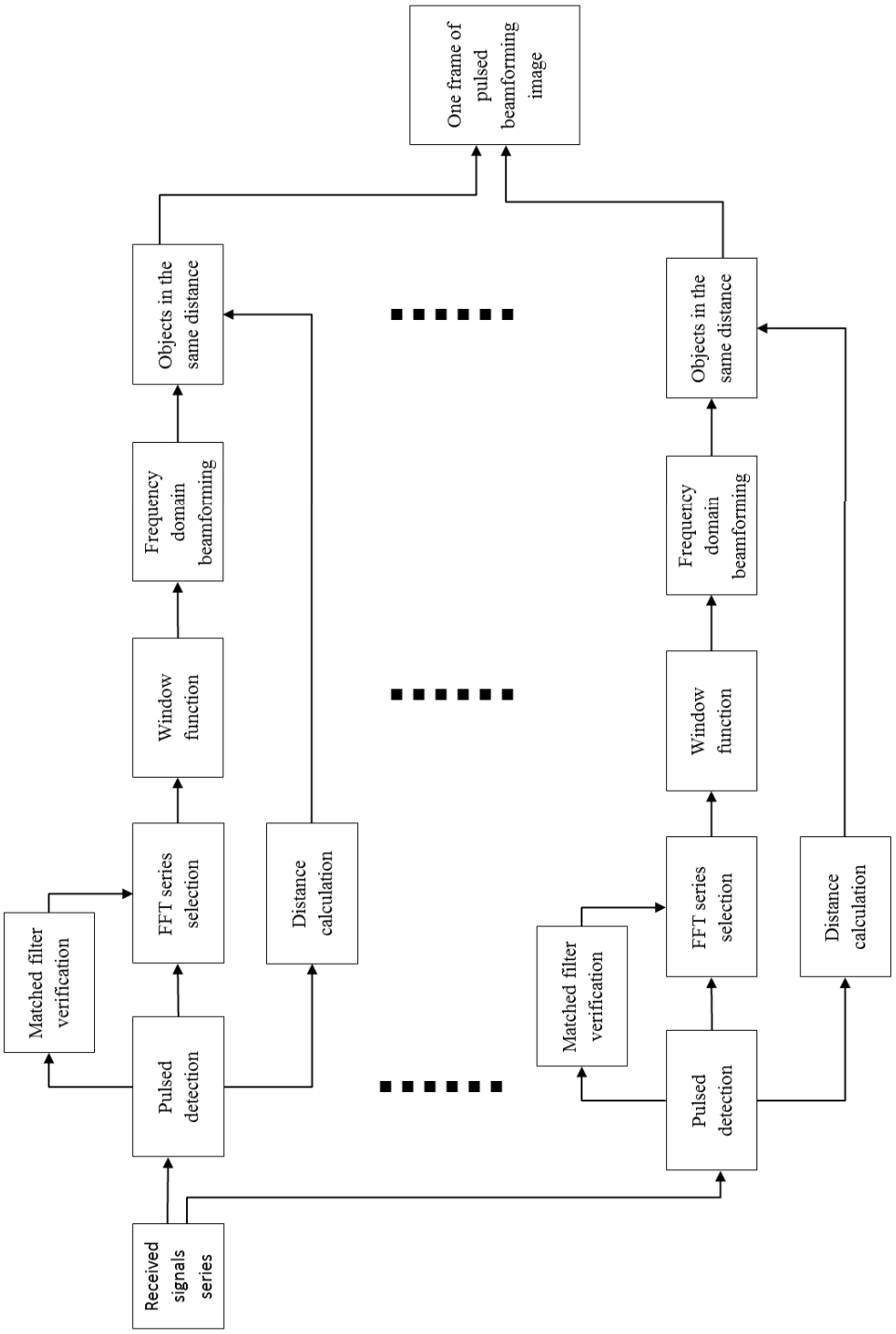
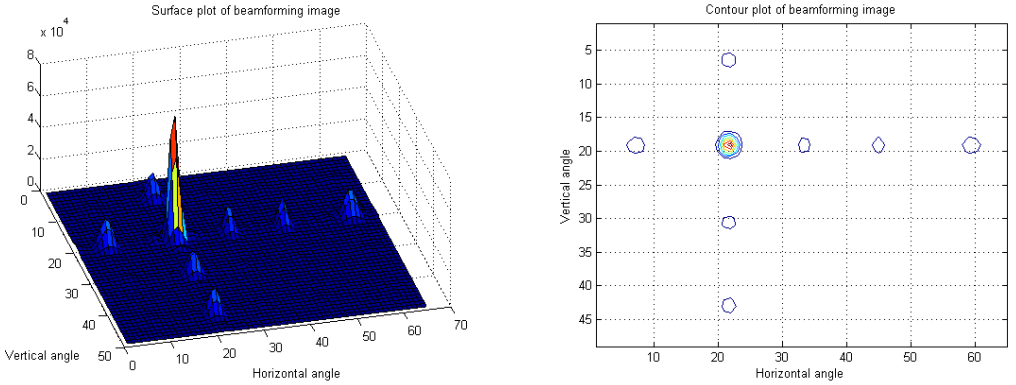


Figure 3.50: Proposed pulsed beamforming algorithm



(a) Surface plot of the beamforming image

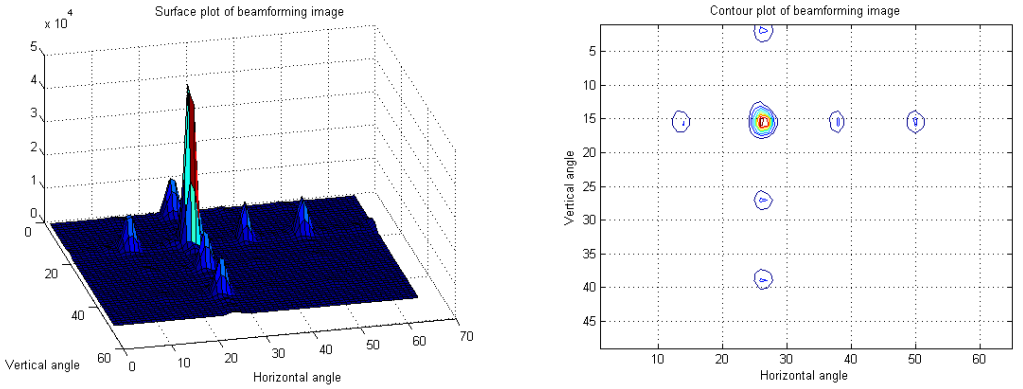
(b) Contour plot of the beamforming image

Figure 3.51: Target one pulsed beamforming image

3.9.3 Simulation results of the example frame

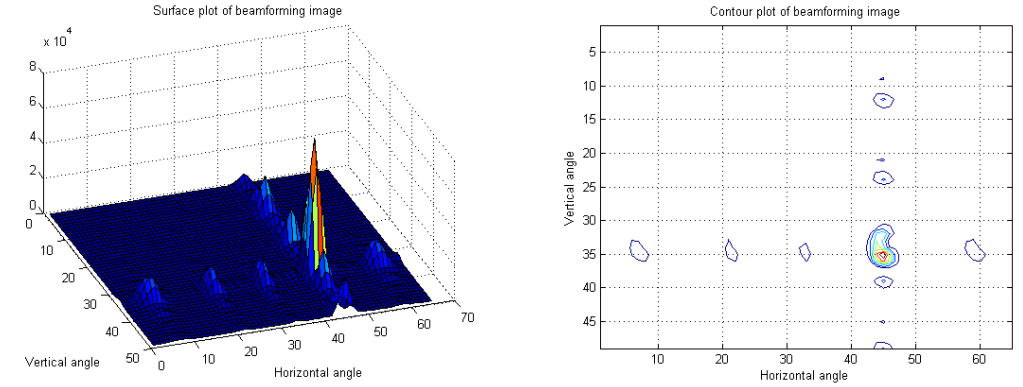
Below is the simulation results of the example image by the proposed algorithm. The three targets are at different distances, so they are in three different frames. Figure 3.51, Figure 3.52 and Figure 3.53 show the beamforming image of target one, target two and target three, separately.

When Figure 3.51, Figure 3.52 and Figure 3.53 are multiplied with the related distance, the example frame can be obtained, which is shown in Figure 3.54.



(a) Surface plot of the beamforming image **(b)** Contour plot of the beamforming image

Figure 3.52: Target two pulsed beamforming image



(a) Surface plot of the beamforming image **(b)** Contour plot of the beamforming image

Figure 3.53: Target three pulsed beamforming image

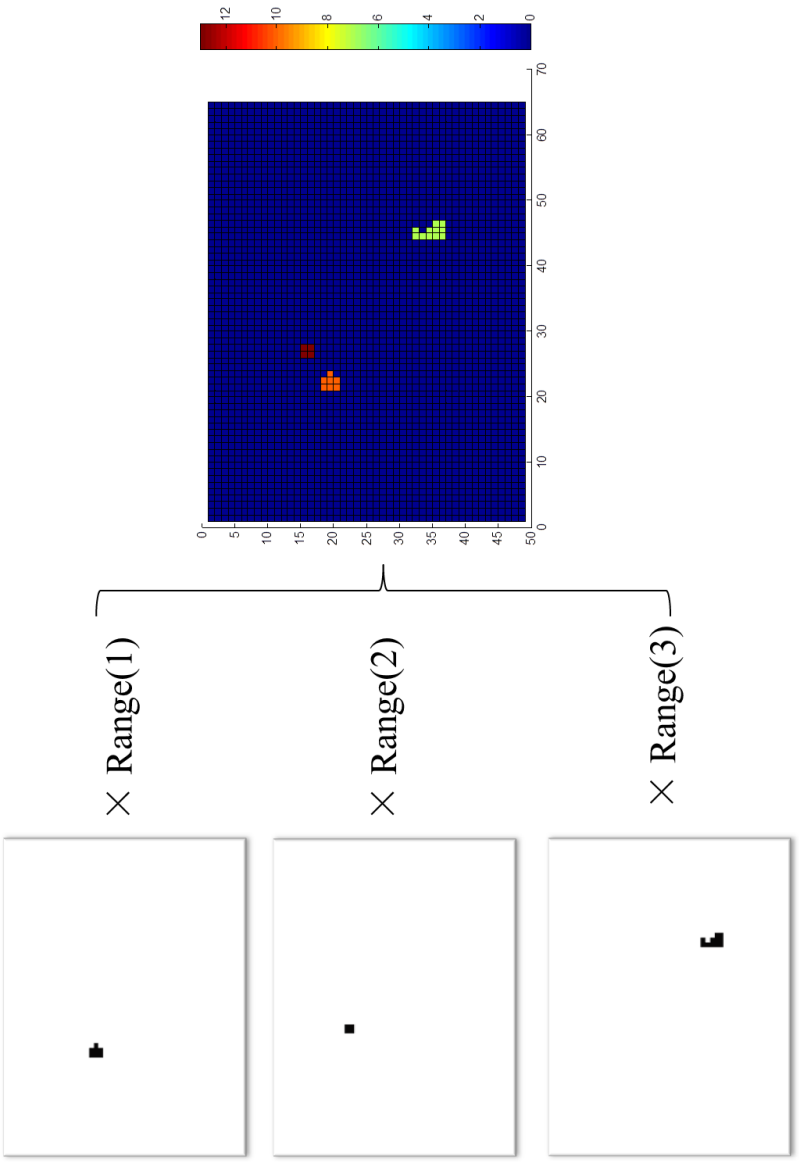


Figure 3.54: Proposed example pulsed beamforming image

4 Image construction of fifth layer with dynamic information using Doppler beamforming - Addresses Proposition 2

4.1 Introduction

In the last chapter the creation of a range layer by pulsed beamforming radar was discussed. That is considered as the fourth layer in addition to traditional RGB layers. In this chapter, the creation of a radial velocity layer by Doppler beamforming radar is further investigated. Pulsed beamforming radar can separate targets by distance while Doppler beamforming radar can separate targets by velocity. Therefore, Doppler beamforming radar has the ability to separate moving objects from stationary objects, which is very useful in computer vision.

Some aspects of Doppler beamforming are quite similar to pulsed beamforming. So the similar parts will not be discussed in this chapter.

For cost and structure simplicity consideration, there is no need to build a new radar for Doppler beamforming, so this radar shares the same hardware with pulsed beamforming radar with a different signal processing method. The pulsed radar needs to have a reduced pulse length to increase the resolution and extend the silent gap between the transmitted signals and received signals to increase the detection distance, while Doppler beamforming radar requires continuous waves to increase Doppler effect resolution. As the frame rate of both pulsed beamforming radar and Doppler beamforming radar is high, the radar can be switched between pulse mode and continuous mode in a time division way. To achieve high resolution both in pulse mode and continuous mode, the switch time needs to be balanced.

4.1.1 Benefit of Doppler processing

Background subtraction is an important step in optical image tracking. The quality of the subtraction of objects determines the difficulties and accuracy of the future processing. There are some difficulties in optical image background subtraction, such as the illumination quality, color similarity and disturbance. In most

cases, the objects to be tracked are moving objects. Many moving object tracking methods have been developed. The most popular methods include the frame differencing method and the optical flow method. Either of these methods require a considerable amount of computation and the quality is strongly affected by the application environment. In particular, when the object is moving towards the sensor, the differences between continuous images are insignificant. Doppler radar can only "see" moving objects which make the moving object subtracting much easier. Doppler radar can also easily distinguish two objects with different velocities. As Doppler radar is not based on light quality and color difference, Doppler radar tracking is a good complimentary tracking resource to optical imaging. Obviously, Doppler radars also have the benefit of providing the radial velocity which enhances the tracking accuracy.

4.1.2 Classification of Doppler radar

There are many types of Doppler radars: coherent pulsed Doppler radar, pulsed Doppler radar, continuous wave Doppler radar and frequency modulated Doppler radar.

Continuous wave radars send continuous waves towards the detection area and are easy to operate. There is no minimum or maximum range if the attenuation is acceptable. The scan rate can be very high. However continuous wave radar cannot determine the range of the object. Pulsed radar can solve this problem by sending pulsed waves. Pulsed radar has a low detection rate because it needs to wait for the returned signals. Doppler radars require filters working at different frequencies to separate objects with different velocities.

4.2 Doppler signal subtraction methods

The Doppler effect is that the observed frequency increases when the object approaches the receiver, while the frequency reduces when the object recedes from the receiver. The amount of frequency shift is proportional to the object's velocity towards or away from the receiver. The Doppler shift frequency f_d is given by:

$$f_d = f_c \cdot \frac{v}{c} \cdot \cos\phi \tag{4.1}$$

where f_c is the carrier frequency in hertz, c is the wave velocity in m/s, v is the velocity of the target in m/s and ϕ is the angle between the object moving direction and the radar beam.

4.2.1 Subtraction of the Doppler signals

There are many ways to subtract Doppler frequency from carrier frequency, such as Short Time Fourier Transform (STFT), low pass filtering and wavelet analysis.

4.2.1.1 Multiply the original signal and use a low pass filter

A traditional method which is frequently used in analogue radar processing is to multiply the receiver signals with the carrier signals. $X_T(t)$ is the transmitted signal and $X_R(t)$ is the received signal.

$$\begin{aligned} X_T(t) &= A_T \cos(2\pi \cdot f_c \cdot t) && (4.2) \\ X_R(t) &= A_R \cos(2\pi \cdot (f_c + f_d) \cdot t) && (4.3) \end{aligned}$$

where A_T is the amplitude of the transmitted signals, A_R is the amplitude of the received signals, f_c is the carrier frequency and f_d is the Doppler frequency.

So when multiplying the received signals with the carrier signals, the multiplication $X_m(t)$ can be written as

$$\begin{aligned} X_m(t) &= A_R \cos(2\pi \cdot (f + f_d) \cdot t) \cdot A_T \cos(2\pi \cdot f \cdot t) \\ &= \frac{1}{2} A_T A_R (\cos(2\pi \cdot (2 \cdot f + f_d) \cdot t) + \cos(2\pi \cdot f_d \cdot t)) && (4.4) \end{aligned}$$

where $X_m(t)$ is a signal mainly combined of two parts. If the carrier frequency f_c is much higher than f_d, $2 \cdot f_c + f_d$ is much higher than f_D, therefore f_D can be easily removed by a low pass filer. The remit signals are

$$X_{remit}(t) = A' \cos(2\pi \cdot f_d \cdot t + \varphi) \quad (4.5)$$

where A' and φ are the amplitude and phase of the remit signals. Usually, Analogue to Digital Convertors (ADCs) are used to sample the signals. Similar to the analogue processing method, the signal can be multiplied with a sampled carrier frequency. Then the sub-sampling is applied instead of the low pass filter.

$$y[n] = \sum_{k=0}^{K-1} x[nM-k] \cdot h[k] \tag{4.6}$$

where $h[k]$ is impulse response.

So the sampled Doppler signal is

$$X_{remit}[n] = A' \cos(2\pi \cdot f_d \cdot nT_s + \varphi) \tag{4.7}$$

Low pass filters can be either digital or analogue. The analogue low pass filter requires extra circuits while the digital low pass filter requires extra computation. If only the moving object needs to be detected, then it is only the Doppler frequency that needs to be processed; analogue low pass filters are suitable because the sample rates for the ADCs are much lower because the Doppler frequency is much lower than the carrier frequency.

4.2.1.2 Sub-sampling method

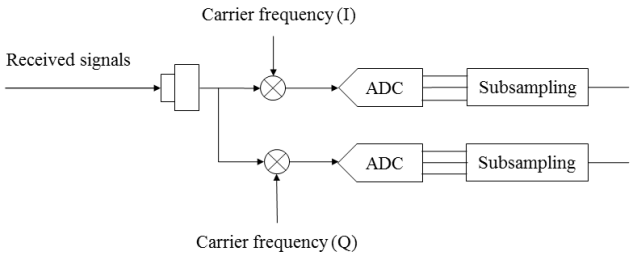
Low pass filters are not cost effective either in terms of computation load or hardware resource. One benefit of digital processing is that the high frequency signals can be filtered by sub-sampling. This is because of the Nyquist equation; only the signals lower than half the sample frequency can be recreated.

According to when the sub-sample happens in different process stages, two methods can be used as shown in Figure 4.1.

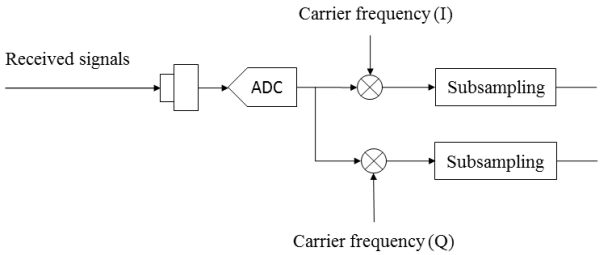
If the first sub-sampling method is used, at least twice of the carrier frequency is required, and if the second sub-sampling method is adopted, only a frequency higher than twice the highest Doppler frequency is required. How to choose the sample method depends on whether or not pulsed beamforming needs to be processed. If pulsed beamforming needs to be processed too, the second method is preferred, otherwise the first method is preferred.

4.2.1.3 Short-time Fourier transforming

The other method to subtract the Doppler signal is to transform the signals into frequency domain using the Fourier transform. The Doppler signal can be identified in the frequency spectrum, that is the difference between the received signal



(a) Sub-sampling method 1



(b) Sub-sampling method 2

Figure 4.1: Sub-sampling of Doppler signals

frequency and the carrier signal frequency. For real time target tracking, Short Time Fourier transform (STFT) is required, which is a time-frequency processing method. A non-zero window function is required to separate signals in a short period of time.

$$STFT\left[x(t)\right](\tau,\omega) \quad \equiv \quad X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)\,\omega(t-\tau)\,e^{-j\omega t}dt \tag{4.8}$$

where $\omega(t)$ is the window function. If the window function is too wide, very limited measurement of Doppler signals can be achieved; if the window function is too narrow, the frequency resolution of each measurement is low.

4.3 Linear array beamforming processing of Doppler signals

4.3.1 Frequency domain Doppler beamforming

The medium constraints and assumptions of Doppler beamforming are the same as in pulsed beamforming. In order to use Doppler beamforming to create a two dimensional image, it is better to start from the one dimensional linear array.

Here, it is assumed that N sensors are laid in a line which is shown in Figure 4.2. The distance between each sensor is equal, which is S_c. In non-Doppler beamforming process, to avoid ambiguities, $S_c \leqslant \frac{\lambda_c}{2}$. λ_c is the wavelength of the carrier. As the Doppler frequency f_d in most cases is lower than the carrier frequency f_c, $S_c \leqslant \frac{\lambda_c}{2} < \frac{\lambda_d}{2}$. If the sensor spacing is the same as in non-Doppler beamforming, there is no ambiguity in Doppler beamforming, either. If non-Doppler beamforming does not need to be processed, S_d can be up to $\frac{\lambda_d}{2}$, that is much bigger than S_c, and the beamwidth can be narrow. However, in this research, for cost consideration, pulsed beamforming and Doppler beamforming shared the same hardware with sensor spacing at $\frac{\lambda_c}{2}$.

The system structure of Doppler beamforming is similar to pulsed Doppler beamforming described in the last chapter. To avoid the complexity and save energy, only one transmitter was used.

According to the analysis of time domain beamforming and frequency domain beamforming in chapter 3, time domain beamforming has a limited beam number

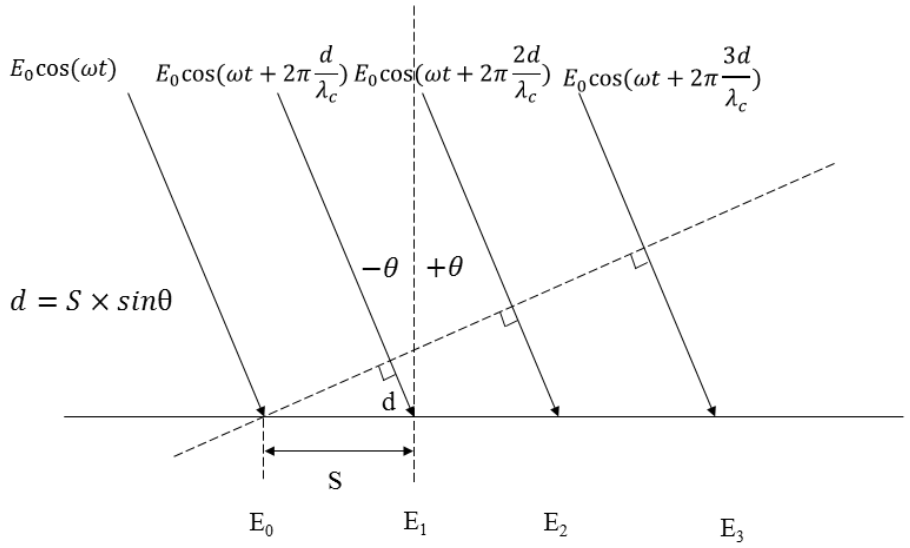


Figure 4.2: Beamforming calculation

unless the sample frequency increases to a very high level. In this chapter, only frequency domain beamforming is studied.

The received signal of qth sensor is

$$s_q(t) = E_q \cos\left(2\pi\left(f_c + f_d\right)\left(t - q \cdot 2\pi \cdot S \cdot \sin\theta / c\right)\right) \tag{4.9}$$

where $s_q(t)$ is the received signal, E_q is the maximum amplitude of received signal from the qth sensor after attenuation in air, S is the space between adjacent sensors and λ_c is the wavelength of the carrier. To simplify the equation, it is assumed that $\triangle t_q = -q \cdot 2\pi \cdot S \cdot \sin\theta / c$, therefore

$$s_q(t) = E_q \cos\left(2\pi\left(f_c + f_d\right)\left(t + \triangle t_q\right)\right) \tag{4.10}$$

According to the analysis of the Doppler signal subtraction, the method which multiplies the received signals with the transmitted signal and sub-samples the multiplied signals was used. The transmitted signal $x(t)$ is supposed to be

$$x(t) = E_x \cos(2\pi f_c t + \varphi_0)$$

So when $s_q(t)$ is multiplied by $x(t)$,

$$\begin{aligned} m_q(t) &= \frac{1}{2}E_q E_x \cdot (\cos(2\pi \cdot f_d \cdot t + 2\pi \cdot (f_c + f_d) \cdot \Delta t_q - \varphi_0) + \cdots \\ &\quad \cos(2\pi \cdot (2 \cdot f_c + f_d) \cdot t + 2\pi \cdot (f_c + f_d) \cdot \Delta t_q + \varphi_0)) \end{aligned} \tag{4.11}$$

After the low pass filter or sub sampling, the higher frequency can be removed and only the Doppler frequency is left. The sample frequency f_s should be in the range $2f_D \leqslant f_s < f_c$. The signal after the low pass filter or sub sampling is

$$\begin{aligned} d_q(t) &= \frac{1}{2}E_q E_x \cos(2\pi f_d (t + \Delta t_q) + 2\pi \cdot f_c \Delta t_q + \varphi_1) \\ &= \frac{1}{2}E_q E_x \cos\left(2\pi f_d (t + \frac{f_d + f_c}{f_d}\Delta t_q) + \varphi_1\right) \end{aligned} \tag{4.12}$$

It is defined that $E'_q = \frac{1}{2}E_q E_x$

$$d_q(t) = E'_q \cos\left(2\pi f_d (t + \frac{f_d + f_c}{f_d}\Delta t_q) + \varphi_1\right) \tag{4.13}$$

In pulsed beamforming, the equation becomes

$$s_q(t) = E_q \cos(2\pi f_c (t + \triangle t_q)) \tag{4.14}$$

Compare Equation 4.13 with Equation 4.14; the structures are very similar, so the relationship between Doppler beamforming and Non-Doppler beamforming is

$$f_d \leftrightarrow f_c \tag{4.15}$$

$$\frac{f_d + f_c}{f_d}\Delta t_q \leftrightarrow \triangle t_q \tag{4.16}$$

$d_q(t)$ need to be sub-sampled at frequency f_s and sample frequency $2f_D \leqslant f_s < f_c$.

$$\begin{aligned} d_q[n] &= E'_q \cos\left(2\pi f_d\left(nT_s - \frac{f_d+f_c}{f_d}q\cdot 2\pi\cdot S\cdot\sin\theta/c\right)\right) \\ &= E'_q\cos\left(2\pi f_d T_s\left(n - \frac{f_d+f_c}{f_d}q\cdot 2\pi\cdot S\cdot\sin\theta/c\cdot f_s\right)\right) \quad (4.17) \\ &= d_0\left[n - \frac{f_d+f_c}{f_d}q\cdot 2\pi\cdot S\cdot\sin\theta/c\cdot f_s\right] \quad (4.18) \end{aligned}$$

Similar to frequency domain pulsed beamforming, the frequency spectrum of the Doppler signal is

$$D_q[k] = \frac{1}{N}\sum\nolimits_{n=0}^{N-1} d_q[n]\cdot e^{-j2\pi nk/N} \quad (4.19)$$

where N is the sample length.

So the supposed time delay in direction φ of sensor q is $\triangle n_{\varphi,q} = -\frac{f_d+f_c}{f_d}q\cdot 2\pi\cdot S\cdot\sin\theta/c\cdot f_s$. So the delay and sum of the received signals $y[k,\varphi]$ is

$$y[k,\varphi] = \sum_{q=0}^{Q-1} D_q[k]\cdot e^{-j2\pi k\cdot\triangle n_{\varphi,q}/N} \quad (4.20)$$

If $\varphi = \theta$, the absolute value of $y[n,\varphi]$ achieves the maximum. The flowchart of frequency domain Doppler beamforming is shown in Figure 4.3.

4.3.2 Simulation of linear array Doppler beamforming

The simulation is shown in Figure 4.4, Figure 4.5 and Figure 4.6. The propagation velocity is 340 m/s, the carrier frequency is 40kHz, the sensor spacing is 0.00425 meters, the number of sensors is 9, the samples per cycle is 4, the FFT length is 1024 samples, the position of target is (2,4) and the radial velocity is 6 m/s.

Figure 4.4 shows the simulation when there is one moving object. The position of the moving object is (2,4), the radial velocity is 6 m/s.

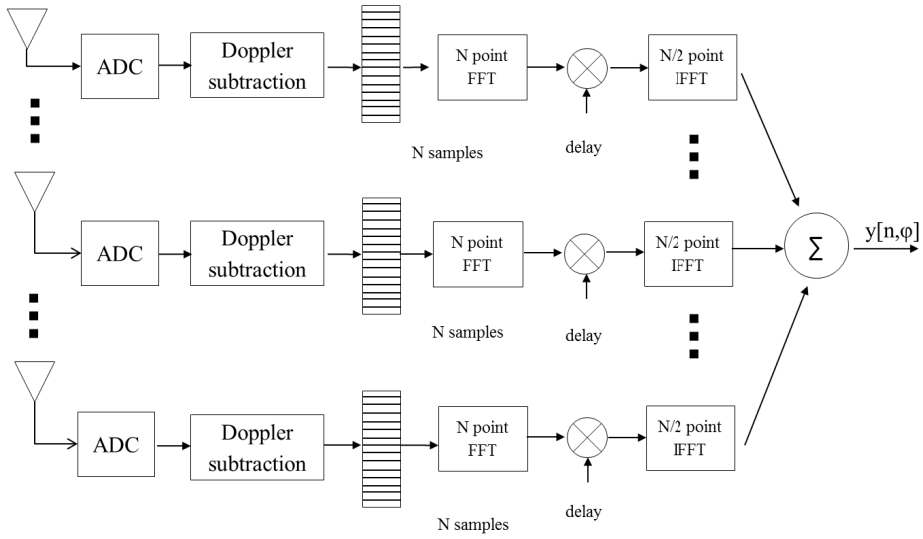


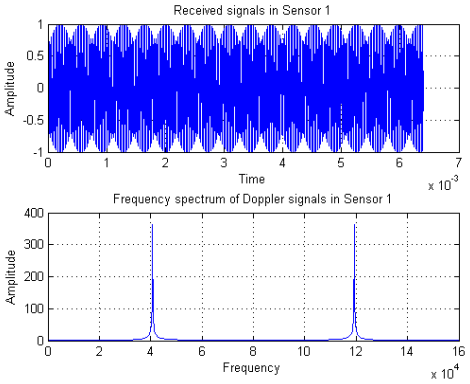
Figure 4.3: Doppler beamforming

Figure 4.5 shows the simulation of a stationary object and a moving object. The position of the moving object is (2,4), the radial velocity is 6 m/s. The position of the stationary object is (0,10). The result shows that only the moving object can be detected.

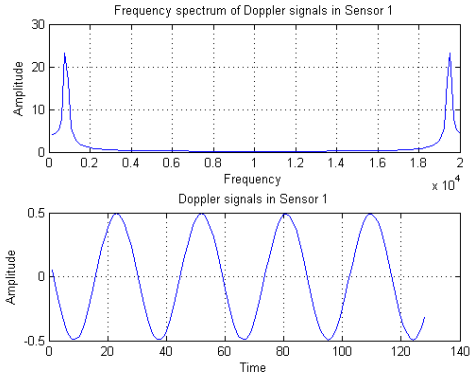
Figure 4.6 shows the simulation of two moving objects in position (2,4) and (0,10). The radial velocities are 6 m/s and 15 m/s, separately. The method used to separate these two targets will be discussed in section 4.4.

4.3.3 Window function

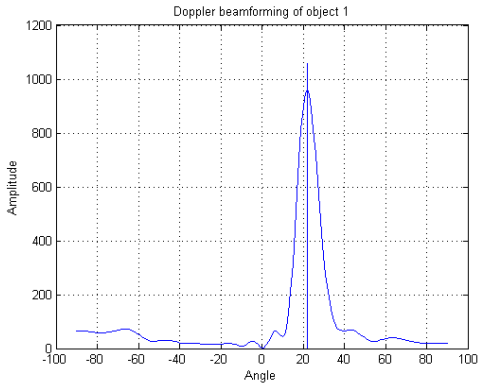
The window function for Doppler beamforming is also tested. Figure 4.7 shows Doppler beamforming with Gaussian windows, and Figure 4.8 shows Doppler beamforming with Dolph-Chebyshev windows.



(a) Received signals in Sensor 1

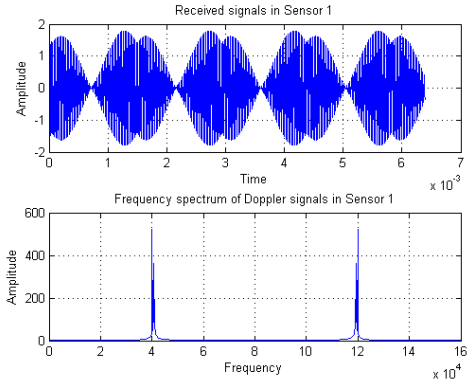


(b) Doppler signals in Sensor 1

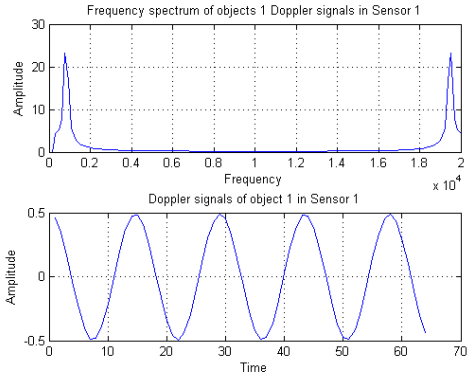


(c) Doppler beamforming of object 1

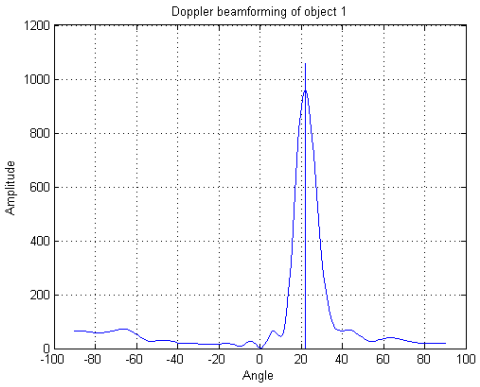
Figure 4.4: One moving object



(a) Received signals in Sensor 1

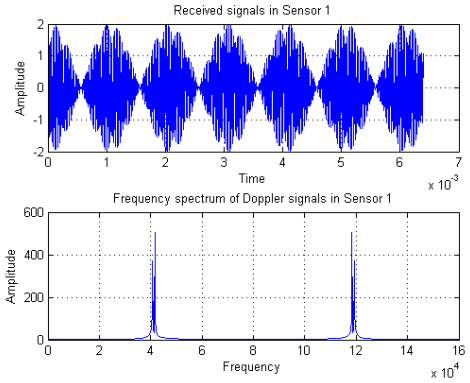


(b) Doppler signals of object 1 in Sensor 1

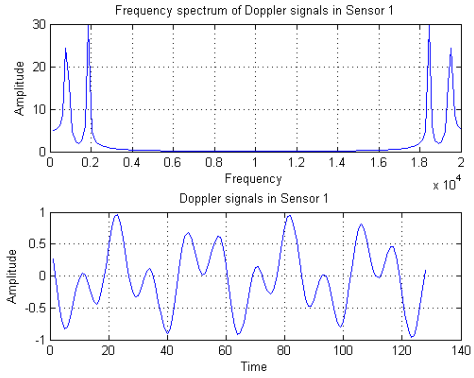


(c) Doppler beamforming of object 1

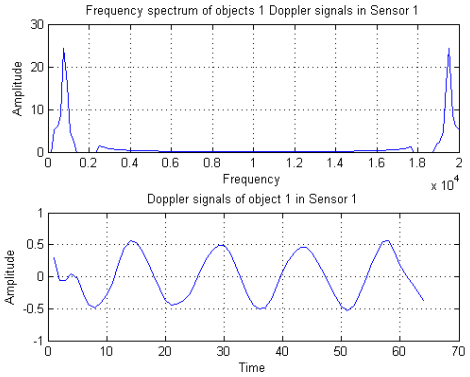
Figure 4.5: One stationary object and one moving object



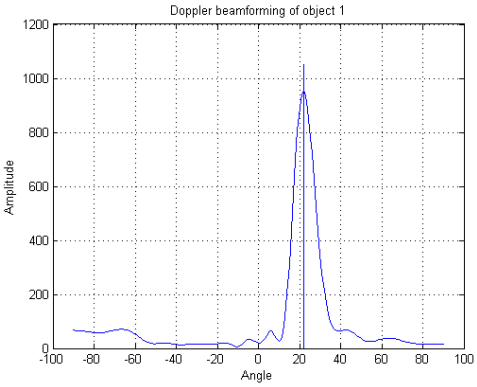
(a) Received signals in Sensor 1



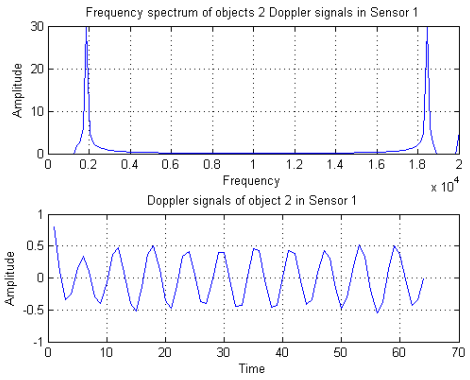
(b) Doppler signals in Sensor 1



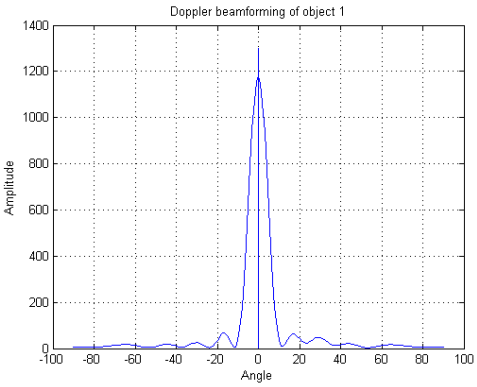
(c) Doppler signals of object 1 in Sensor 1



(d) Doppler beamforming of object 1

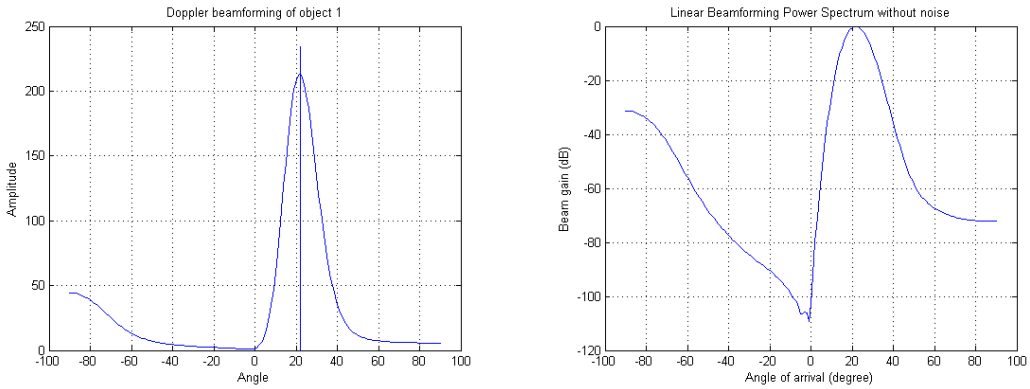


(e) Doppler signals of objects 2 in Sensor 1



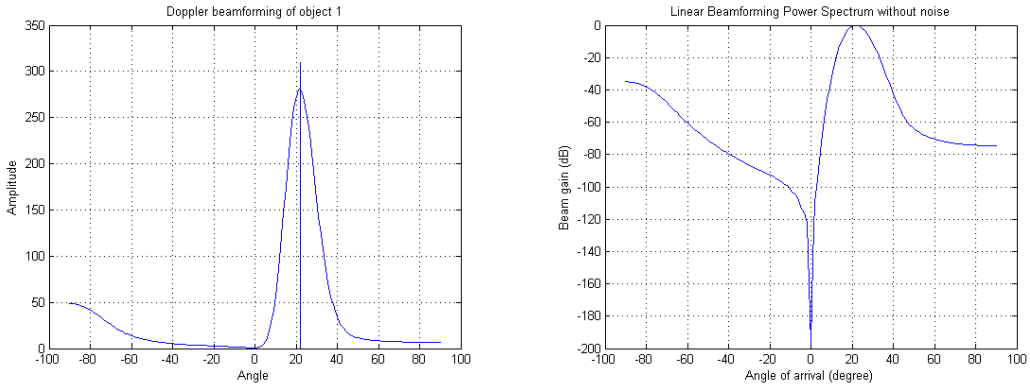
(f) Doppler beamforming of object 2

Figure 4.6: Two moving objects



(a) Gaussian window **(b)** Gaussian window (dB)

Figure 4.7: Linear array Doppler beamforming with Gaussian windows



(a) Dolph-Chebyshev window **(b)** Dolph-Chebyshev window (dB)

Figure 4.8: Linear array Doppler beamforming with Dolph-Chebyshev windows

4.3.4 Relationship between the sensor space and resolution

The hardware in this thesis was designed for both pulsed beamforming and Doppler beamforming and the sensor space was fixed after the radar was built. When the sensor spacing is less than $\lambda_c/2$, there is no ambiguity for pulsed beamforming. For Doppler beamforming, according to Equation 4.1, the Doppler frequency is lower than the carrier frequency if the radial velocity of the target is less than the velocity of sound. In this condition, if the sensor spacing is less than $\lambda_c/2$, there is no ambiguity in Doppler beamforming either. However, if the sensor spacing does not change, the relationship between phase difference and angle of arrival in Doppler beamforming changes. In detail, according to subsection 4.3.1, the time difference $\triangle t$ is

$$\triangle t = S \sin(-\theta)/c \tag{4.21}$$

$$\triangle t = \frac{\phi}{2\pi} \cdot \frac{\lambda_d}{c} \tag{4.22}$$

where S is the sensor spacing, θ is the Angle of Arrival (AoA) of the returned signal containing the Doppler signal, c is the velocity of sound, ϕ is the phase difference of returned signals between adjacent sensors, and λ_c and f_c are the wavelength and frequency of the carrier. S cannot be changed according to the Doppler signal frequency, so $S = \frac{\lambda_c}{2}$. If Doppler frequency is f_d, then

$$\theta = -\arcsin\left(\frac{\phi}{2\pi} \cdot \frac{S}{\lambda_c} \cdot \frac{f_c}{f_d}\right) \tag{4.23}$$

Compare with Equation 3.9

$$\sin\theta_d = \sin\theta \cdot \frac{f_c}{f_d} \tag{4.24}$$

Therefore, the AoA needs to be transformed by the ratio $\frac{f_c}{f_d}$ in Doppler beamforming.

4.4 Separating of multiple targets with different velocities

4.4.1 FFT filtering

When there are multiple targets, in the frequency spectrum, there are multiple peaks. section 4.3 demonstrates the relationship between the sensor spacing, the phase shift and the AoA.

To get the correct angle θ_d which is related to $\frac{f_c}{f_d}$, the Doppler frequency needs to be known. If there is more than one Doppler frequency, they must be separated first and processed separately.

In frequency domain beamforming, all the sampled signals in the time series need to be converted into the frequency domain. Therefore, the FFT filtering is the most suitable filter as it does not increase the computation load very much.

In detail, FFT filtering is also called zeroing bins in the frequency domain. The filtering is to set the value of the unwanted frequency to zero and to leave the wanted frequency. It is the same as multiplying by a rectangular window in the frequency domain. Multiplying by a window in the frequency domain is the same as circular convolution by the transform of that window in the time domain. The transform of a rectangular window is the Sinc function $(\sin(\omega t)/\omega t)$. The Sinc function has many large ripples that extend the full width of the time domain aperture. If a time-domain filter that can output all those ripples (ringing) is a "bad idea", then so is zeroing bins. These ripples will be largest for any spectral content that is "between bins" or non-periodic in the FFT aperture width. So if the original FFT input data is a window on any data that is somewhat non-periodic in that window, then those particular artifacts will be produced by zeroing bins.

Another way to look at it is that each FFT result bin represents a certain frequency of sine wave in the time domain. Thus zeroing a bin will produce the same result as subtracting that sine wave or equivalently adding a sine wave of an exact FFT bin center frequency but with the opposite phase. If the frequency of some content in the time domain is not purely periodic in the FFT width, then trying to cancel it by adding the inverse of an exactly periodic sine wave will produce, not silence, but something that looks more like a "beat" note. Conversely, if the original time domain signal is just a few pure unmodulated sinusoids that are exactly periodic in the FFT aperture width, then zeroing FFT bins will remove the designated ones without artifacts.

4.4.2 FIR Filter bank

FFT filtering is not an optimal method to isolate the different frequencies from the original signals. Other than FFT filters there are many types of filters. The best

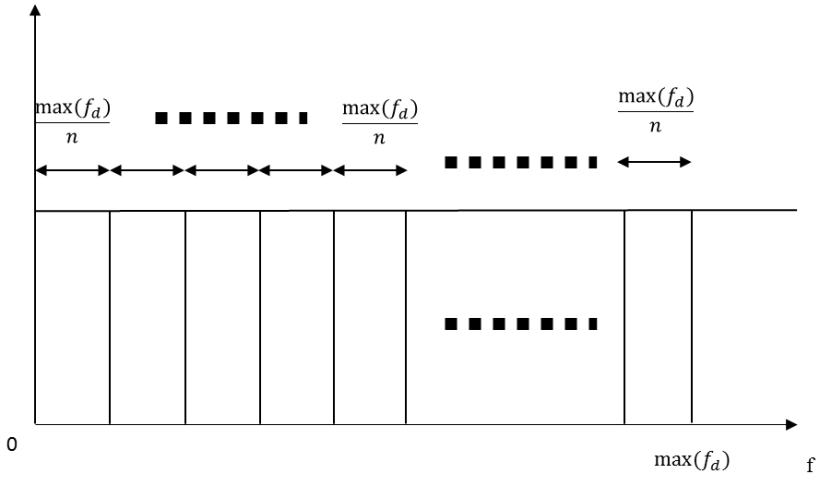


Figure 4.9: Filter bank with equal bandwidth

way is to use the FIR filter bank to separate them. The FIR bank is a number of FIR filters working at different frequencies. In most case, the bandwidths of every FIR filter are equal. The structure is shown in Figure 4.9.

According to Equation 4.23, the AoA highly relies on the ratio of the carrier frequency to the Doppler frequency. In low frequency Doppler signals, the tiny change in the frequency will induce a great shift in the AoA. One way to solve this is to set a threshold for the Doppler signals and the frequency lower than the threshold should be considered as from stationary objects. Also, the bandwidth in the low frequency should be narrow, while the bandwidth in the high frequency should be wide. The disadvantage of this method is that the distinguishing of the velocity of high speed objects is lower. The high speed targets with similar velocity are shown in the same beamforming image. The structure is shown in Figure 4.10.

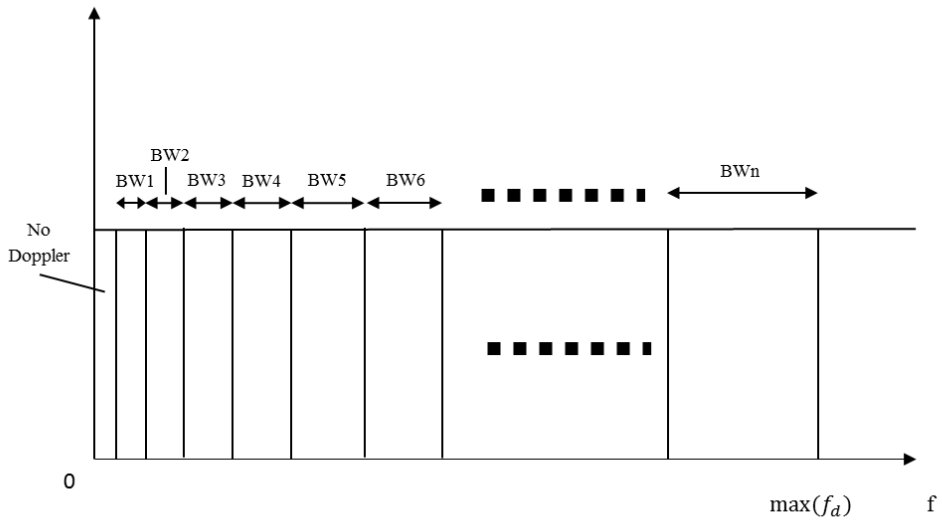


Figure 4.10: Filter bank with unequal bandwidth

4.5 Two dimensional beamforming on Doppler signals

4.5.1 Full rectangular array beamforming

An image layer is assumed to be created with m by n pixels. In each pixel, the value represents the detected scaled Doppler velocity of the object in that area. To find the two dimensional position of the Doppler signal, two dimensional Doppler beamforming needs to be applied.

$$\begin{bmatrix} q_{0,0} & q_{0,1} & \cdots & q_{0,n} & \cdots & q_{0,N} \\ q_{1,0} & q_{1,1} & \cdots & q_{1,n} & \cdots & q_{1,N} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ q_{m,0} & q_{m,1} & \cdots & q_{m,n} & \cdots & q_{m,N} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ q_{M,0} & q_{M,1} & \cdots & q_{M,n} & \cdots & q_{M,N} \end{bmatrix} \tag{4.25}$$

It is assumed that the Doppler frequency is f_d. The phase delay of sensor $q_{m,n}$ compared with the first sensor in the first line comes from two parts, that is the

phase delay between $q_{0,0}$ and $q_{0,n}$, and the phase delay between $q_{0,n}$ and $q_{m,n}$. It is assumed that the AoA in the horizontal direction is α and the AoA in the vertical direction is β. Both the horizontal and vertical sensor spacings are half a wavelength, that is $S_h = S_v = \lambda_c/2$. So the total phase delay of $q_{m,n}$ is

$$\Delta t_{m,n} = -n \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\alpha \cdot \frac{f_s}{f_c} - m \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\beta \cdot \frac{f_s}{f_c} \quad (4.26)$$

Therefore, the Doppler signal subtracted from sensor $q_{m,n}$ is

$$\begin{aligned} s_{n,m}(t) &= E_{n,m} \cos\left(2\pi f_d \left(t - n \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\alpha \cdot \frac{f_s}{f_c}\right.\right. \\ &\quad \left.\left. -m \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\beta \cdot \frac{f_s}{f_c}\right)\right) \end{aligned} \quad (4.27)$$

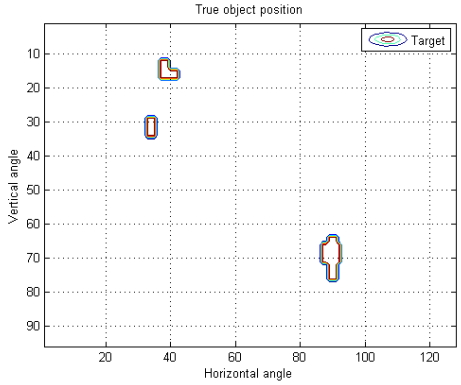
Convert the signal to a discrete signal,

$$\begin{aligned} s_{n,m}[q] &= E_{n,m} \cos\left(2\pi f_d T_s \left(q - n \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin(\alpha) \cdot \frac{f_s}{f_c}\right.\right. \\ &\quad \left.\left. -m \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\beta \cdot \frac{f_s}{f_c}\right)\right) \\ &= s_{0,0}\left[q - n \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\alpha \cdot \frac{f_s}{f_c}\right. \\ &\quad \left. -m \cdot \frac{f_d + f_c}{f_d} q \cdot \pi \cdot \sin\beta \cdot \frac{f_s}{f_c}\right] \end{aligned} \quad (4.28)$$

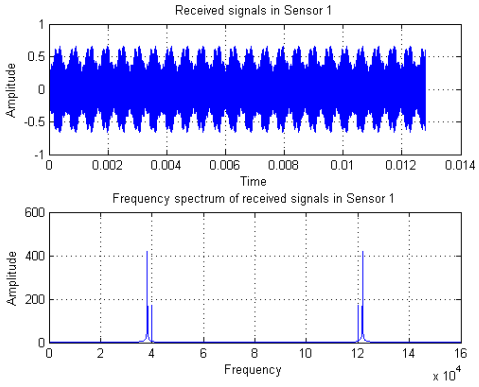
The rest of the procedures are similar to two dimensional pulsed frequency domain beamforming. The supposed time delay in the horizontal direction with the AoA at ϕ and in the vertical direction with the AoA at φ of sensor $q_{n,m}$ are added and $y[q, \phi, \varphi]$ is calculated. When $y[q, \phi, \varphi]$ achieves the maximum, $\phi = \alpha$ and $\varphi = \beta$.

Simulation is shown in Figure 4.11 and Figure 4.12. The picture is the same as the example in chapter 3. The propagation velocity is 340 m/s, the carrier frequency is 40kHz, the sensor spacing is 0.00425 meters, the number of sensors is 9, and the sample number per cycle is 4. If the FFT series length is 256 or 512, the object is not clear in the Doppler beamforming image. When the series length increases to 1024, the target is very clear in the Doppler beamforming image.

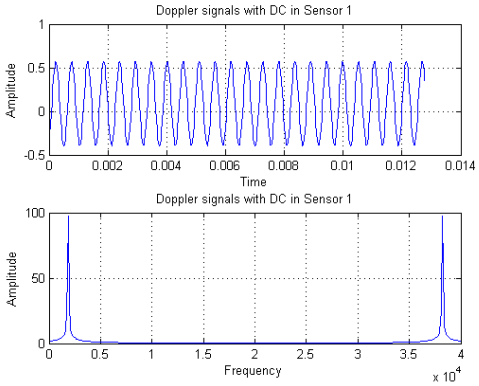
Figure 4.11 shows the simulation with one moving object.



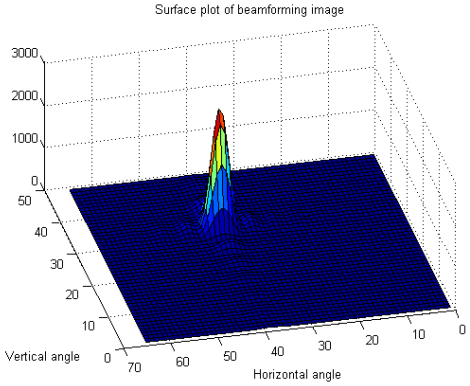
(a) True Object position



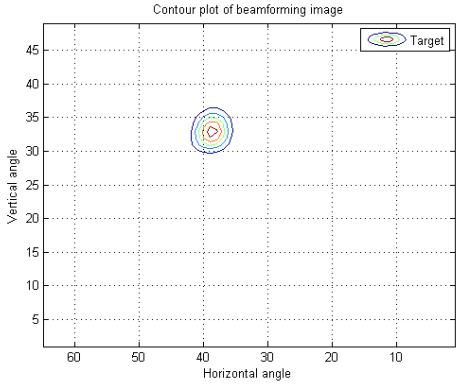
(b) Received signals in Sensor 1



(c) Doppler signals in Sensor 1



(d) Surface plot of beamforming image



(e) Contour plot of the beamforming image

Figure 4.11: Full rectangular sensor array Doppler beamforming with one moving object

Figure 4.12 shows the simulation with two moving objects. These two objects are moving with different velocities, therefore they need to be separated by FFT filtering.

4.5.2 Two line multiplication beamforming

Two line multiplication Doppler beamforming was also investigated. In most environments, there are only a few moving objects. The chance that there are moving objects with the same velocity is very rare, so the possibility that there is ambiguity in the Doppler beamforming image is very low. The principle of the two line multiplication beamforming is to create the beamforming spectrum in the horizontal and the vertical directions separately, and then multiply the results. The simulation is shown in Figure 4.14. There are two moving objects in the environment with different velocities. Both the beamforming in the horizontal and the vertical directions can detect two objects. The FFT filtering is used to separate the two objects and only the object with the same velocity in the two directions needs to be multiplied.

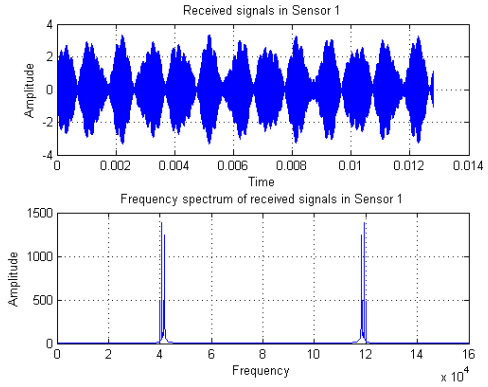
According to the results, the main lobe to side lobe ratio is very high and the targets can be clearly subtracted.

4.5.3 Crossed array beamforming

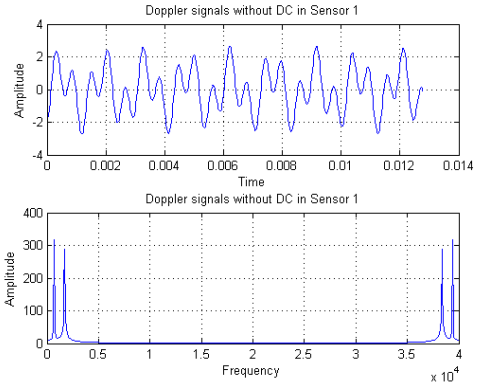
Here crossed array Doppler beamforming was also tested. In Figure 4.16 and Figure 4.18, two targets with different velocities are detected. As discussed in section 4.4 there are two ways to separate the moving objects with different velocities. As in the crossed array beamforming, there is more noise than the full rectangular array beamforming and the two line multiplication beamforming, the performance of the FFT filter and the FIR filter bank is compared.

Figure 4.16 shows the two moving objects separation using the FFT filters and Figure 4.18 shows the two moving objects separation using the FIR Filter bank.

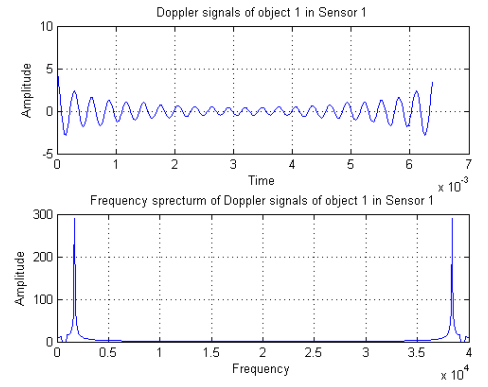
The results show that there is no large difference between the performance of the FFT filtering and the FIR Filter bank. The performance is acceptable.



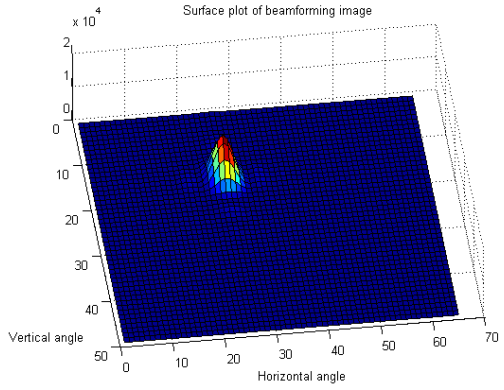
(a) Received signals in Sensor 1



(b) Doppler signals in Sensor 1

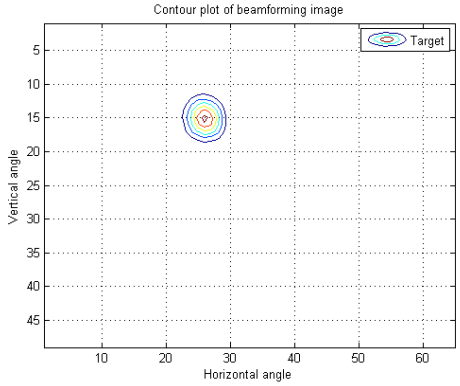


(c) Doppler signals of object 1 in Sensor 1

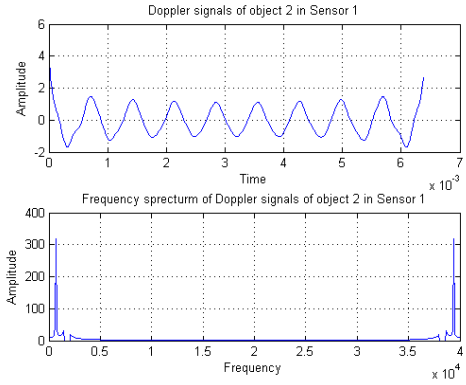


(d) Surface plot of the beamforming image of object 1

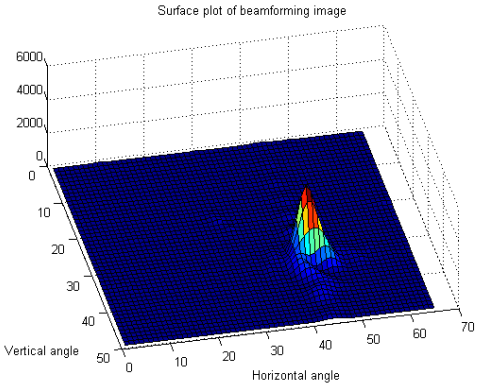
Figure 4.12: Full rectangular sensor array Doppler beamforming with two moving objects (1)



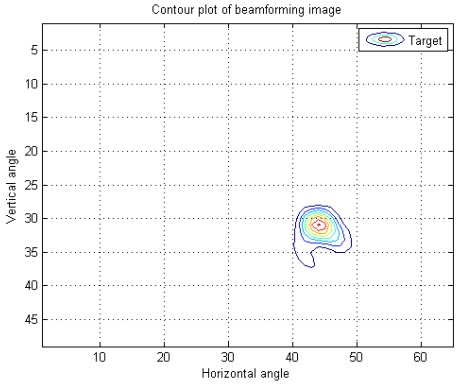
(a) Contour plot of the beamforming image of object 1



(b) Doppler signals of object 2 in Sensor 1

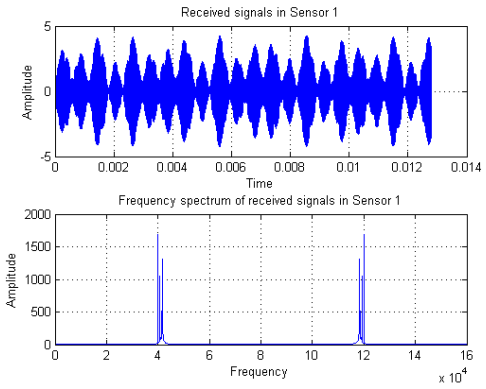


(c) Surface plot of the beamforming image of object 2

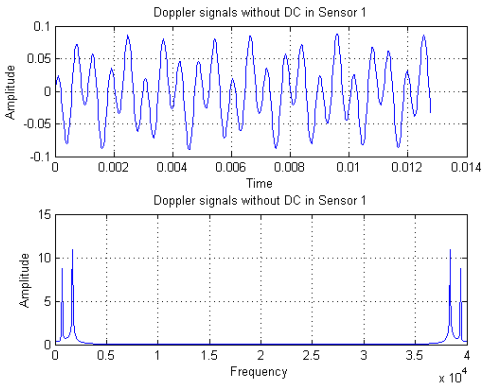


(d) Contour plot of the beamforming image of object 2

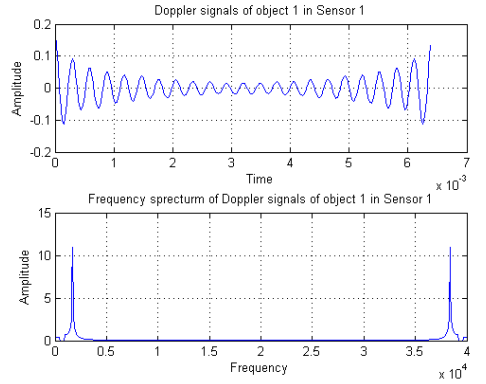
Figure 4.13: Full rectangular sensor array Doppler beamforming with two moving objects (2)



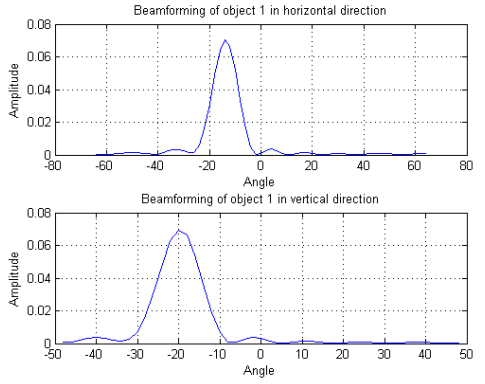
(a) Received signals in Sensor 1



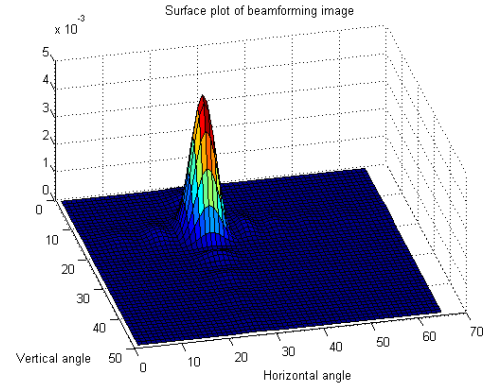
(b) Doppler signals in Sensor 1



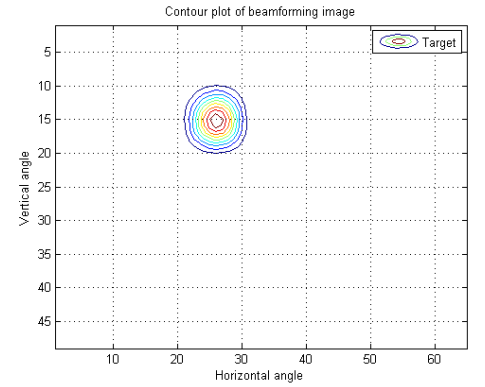
(c) Doppler signals of object 1 in Sensor 1



(d) Beamforming of object 1 in horizontal direction

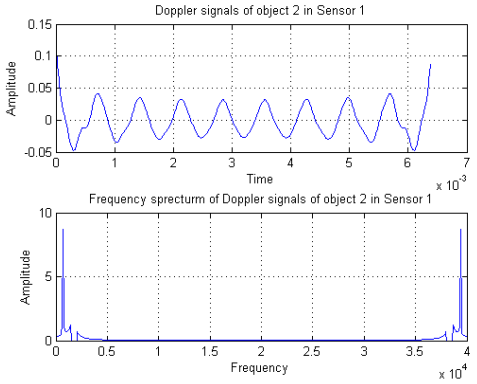


(e) Surface plot of the beamforming image of object 1

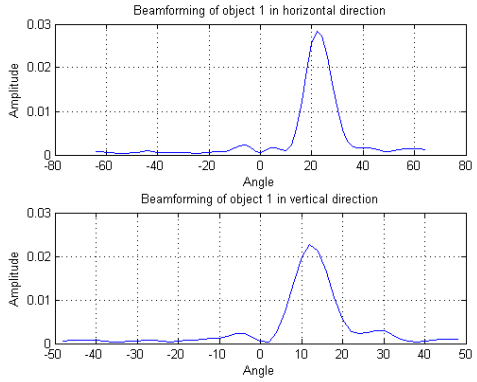


(f) Contour plot of the beamforming image of object 1

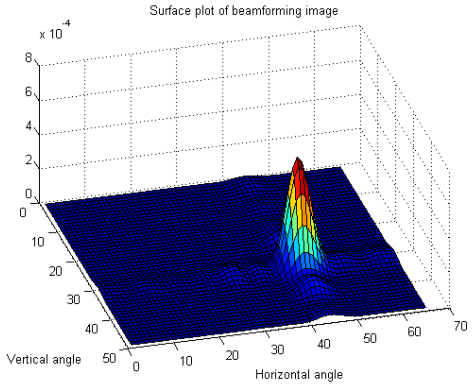
Figure 4.14: Two line multiplication with two moving objects (1)



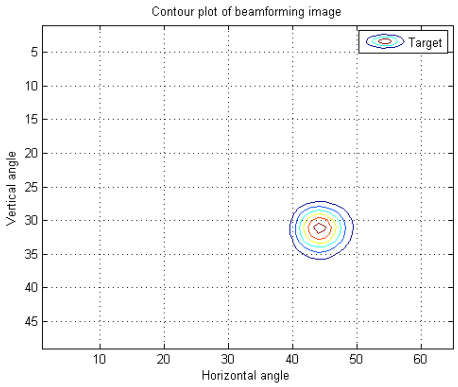
(a) Doppler signals of object 2 in Sensor 1



(b) Beamforming of object 2 in horizontal direction

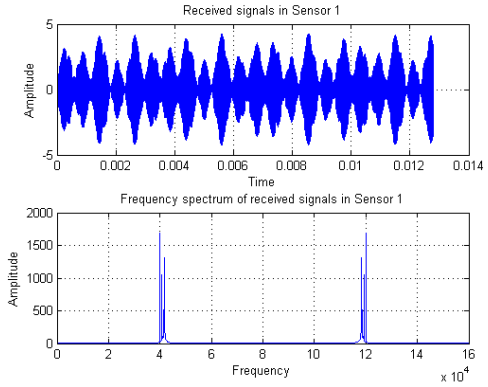


(c) Surface plot of the beamforming image of object 2

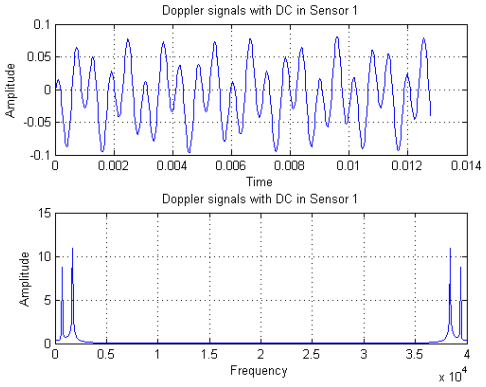


(d) Contour plot of the beamforming image of object 2

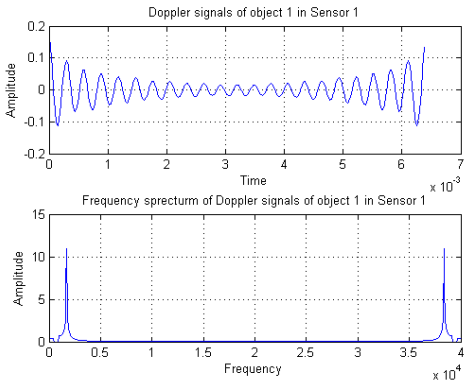
Figure 4.15: Two line multiplication with two moving objects (2)



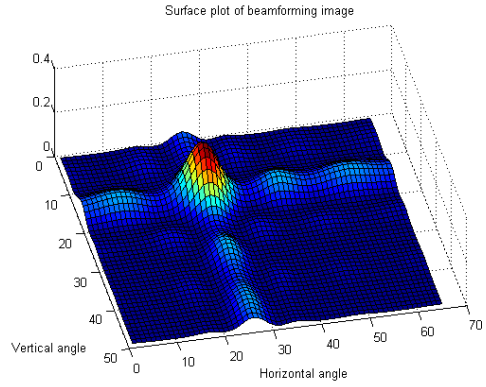
(a) Received signals in Sensor 1



(b) Doppler signals in Sensor 1

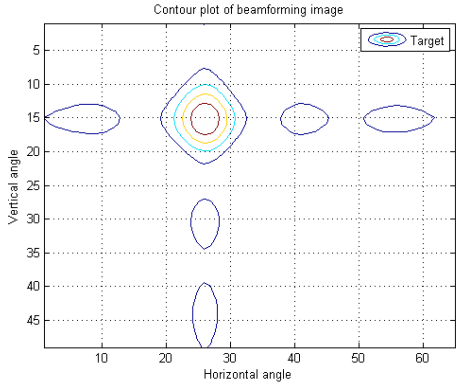


(c) Doppler signals of object 1 in Sensor 1

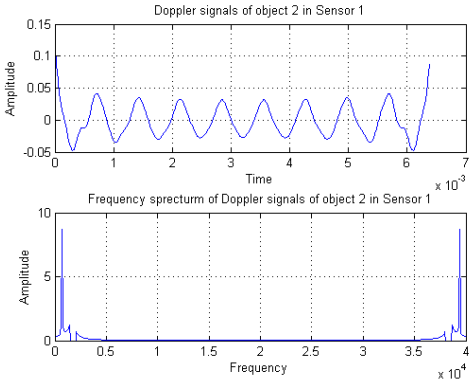


(d) Surface plot of the beamforming image of object 1

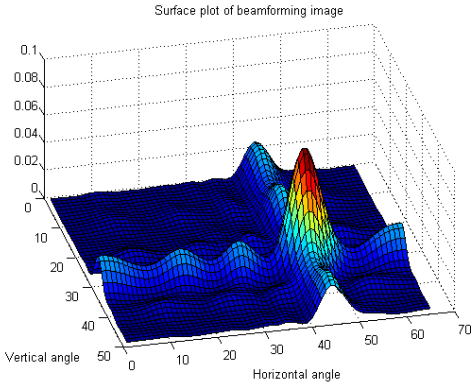
Figure 4.16: Crossed array Doppler beamforming with two moving objects using the FFT filtering (1)



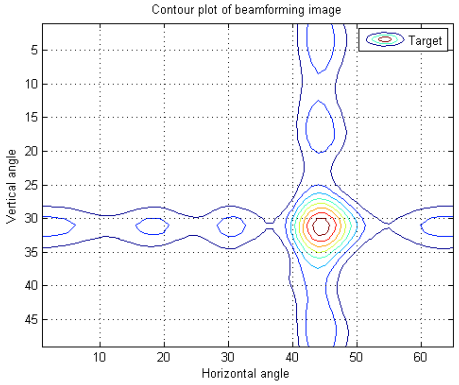
(a) Contour plot of the beamforming image of object 1



(b) Doppler signals of object 2 in Sensor 1

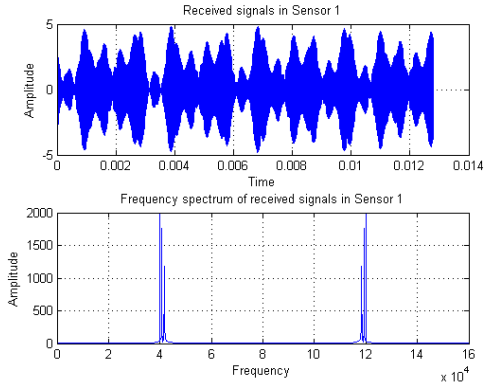


(c) Surface plot of the beamforming image of object 2

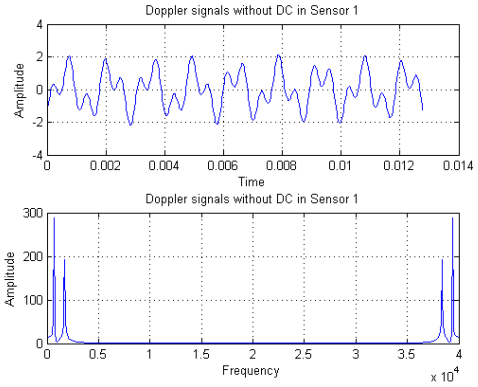


(d) Contour plot of the beamforming image of object 2

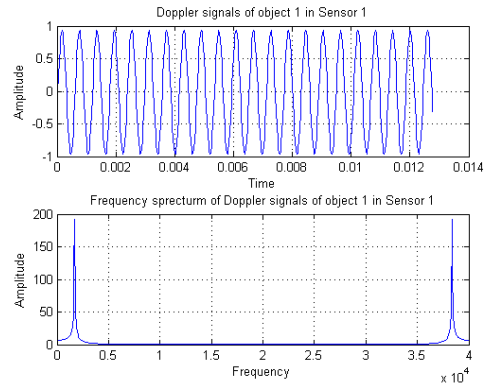
Figure 4.17: Crossed array Doppler beamforming with two moving objects using the FFT filtering (2)



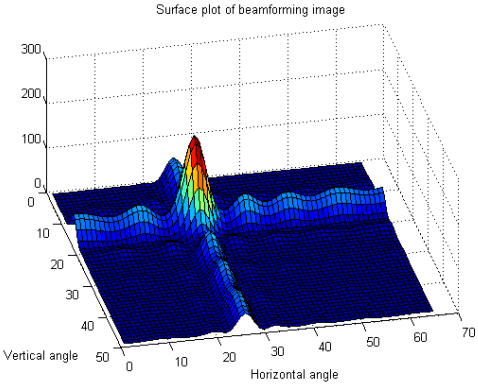
(a) Received signals in Sensor 1



(b) Doppler signals in Sensor 1

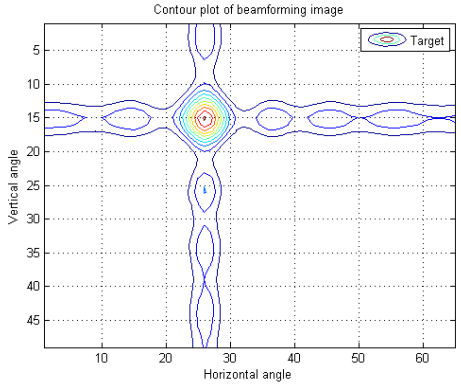


(c) Doppler signals of object 1 in Sensor 1

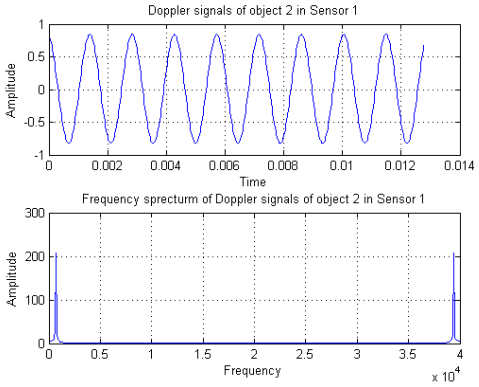


(d) Surface plot of the beamforming image

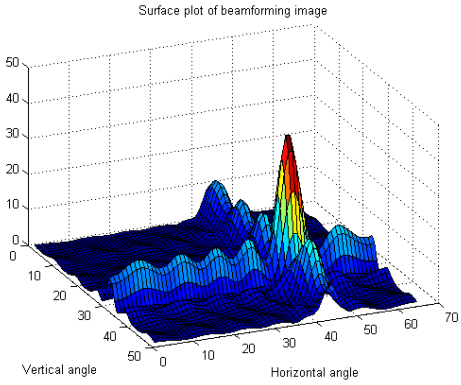
Figure 4.18: Crossed array Doppler beamforming with two moving objects using the FIR filter banks (1)



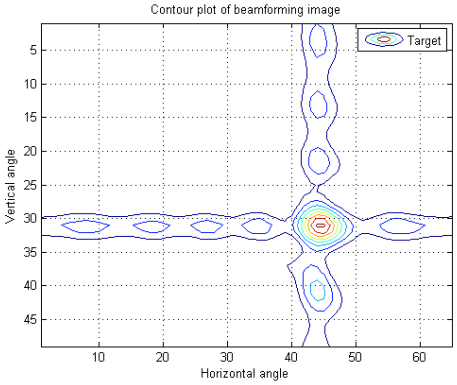
(a) Contour plot of the beamforming image



(b) Doppler signals of object 2 in Sensor 1



(c) Surface plot of the beamforming image



(d) Contour plot of the beamforming image

Figure 4.19: Crossed array Doppler beamforming with two moving objects using the FIR filter banks (2)

4.5.4 Remove zero frequency components

For the Doppler signal detection, the zero frequency (ZF) components left after the Doppler signal subtraction do not affect the results, but in Doppler beamforming, the ZF component difference between different sensors can also generate a Doppler beamforming picture. The ZF components are not only related to the moving object, but also related to all the objects in front of the radar. That means not only the moving object, but also stationary targets can be detected.

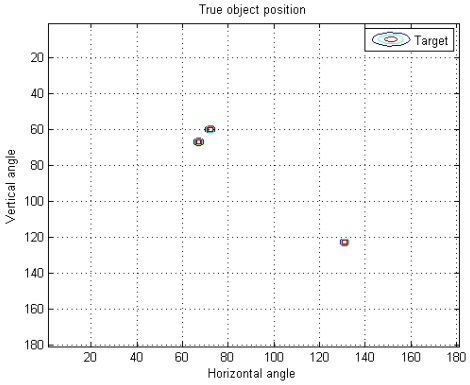
To solve this problem, high pass filters can be used to remove the ZF components; the pass band should be set below the lowest frequency of Doppler signals which need to be detected. For digital signal processing, the high pass filter is not cost effective; one simple alternative solution is to calculate the mean of the sample series and subtract the mean from each sample, then the ZF components can be removed. Thus in Doppler beamforming, only moving objects can be sensed.

In Figure 4.20 and Figure 4.22, the performance of Doppler beamforming with ZF and without ZF is compared. The results show the ZF components produce much more noise. So in Doppler beamforming, the ZF components must be filtered in advance.

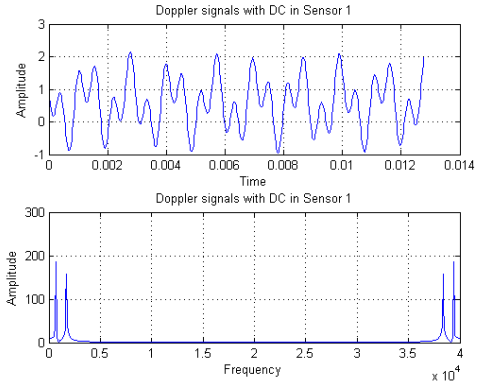
4.5.5 Window functions to reduce side lobe

Window function can sacrifice beam width for an increase in the main lobe to side lobe ratio. Similar to the window function in the linear sensor array, the one dimensional window function can be extended to a two dimensional window function to reduce the side lobes. Gaussian windows and Dolph-Chebyshev windows are mainly used. Figure 4.24 and Figure 4.25 show crossed array Doppler beamforming with Gaussian windows and Dolph-Chebyshev windows, separately. The FFT series length is 2048 and the size of the crossed sensor array is 13 by 13. In the Dolph-Chebyshev window, the ratio of the main lobe to side lobe is 40.

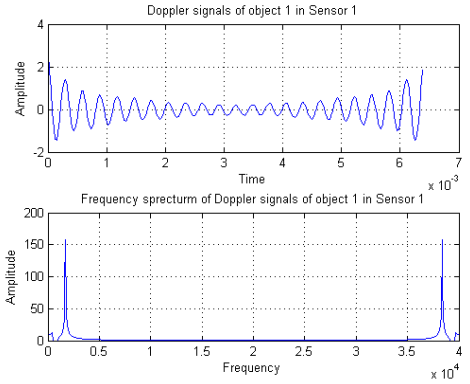
To make the radar image reflect the optical image, the AoA of the radar image is set from -60^o to 60^o in the horizontal direction and -45^o to 45^o in the vertical direction.



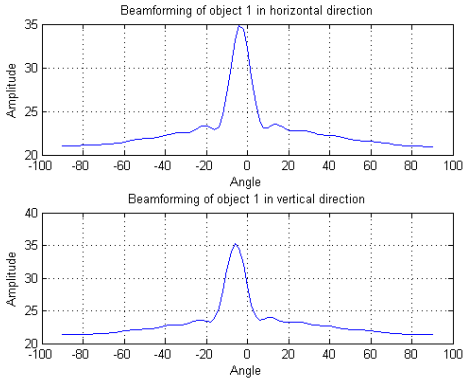
(a) True object position



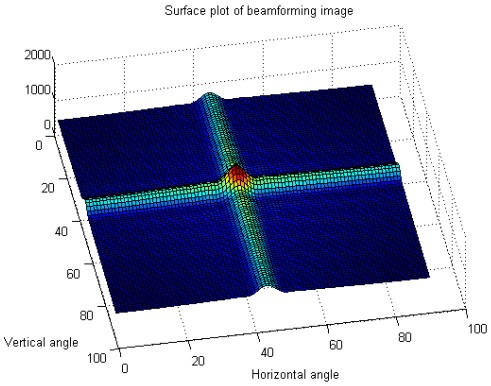
(b) Doppler signals with ZF in Sensor 1



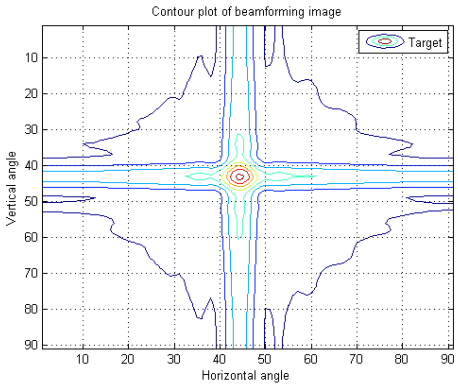
(c) Doppler signals of object 1 in Sensor 1



(d) Beamforming of object 1

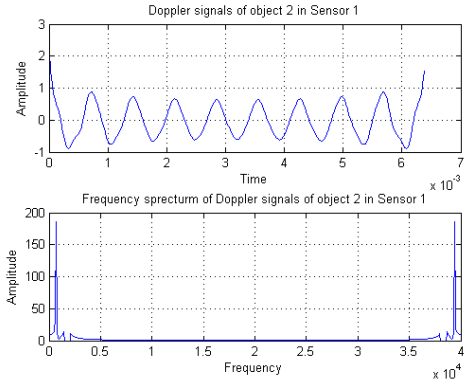


(e) Surface plot of the beamforming image of object 1

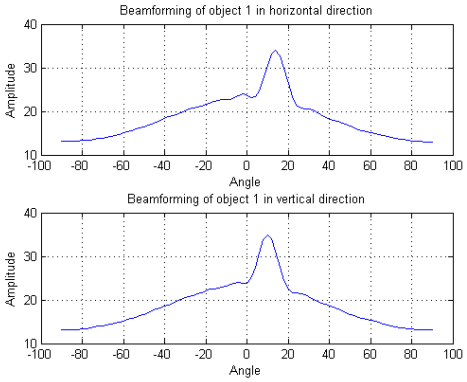


(f) Contour plot of the beamforming image of object 1

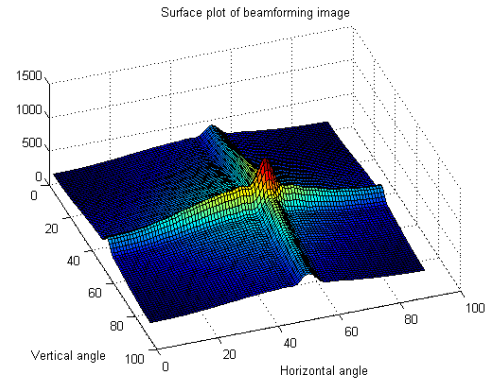
Figure 4.20: Doppler beamforming with ZF components (1)



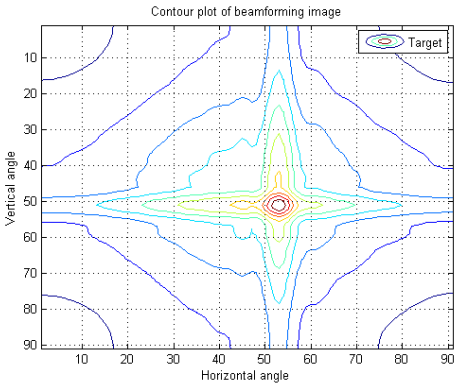
(a) Doppler signals of object 2 in Sensor 1



(b) Beamforming of object 2

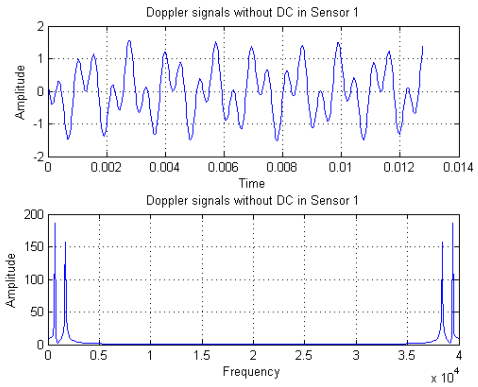


(c) Surface plot of the beamforming image of object 2

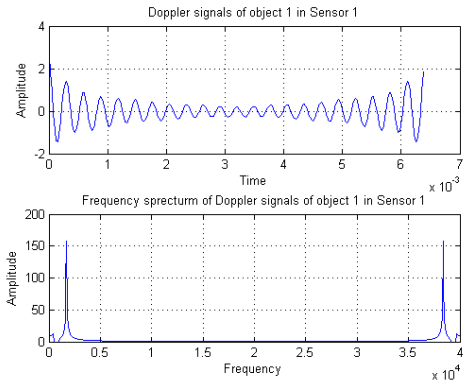


(d) Contour plot of the beamforming image of object 2

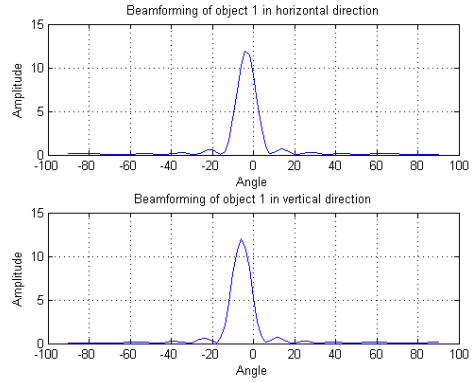
Figure 4.21: Doppler beamforming with ZF components (2)



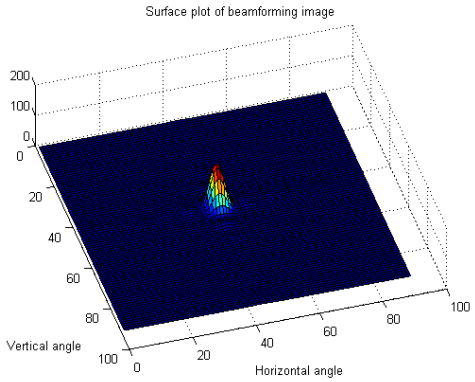
(a) Doppler signals without ZF in Sensor 1



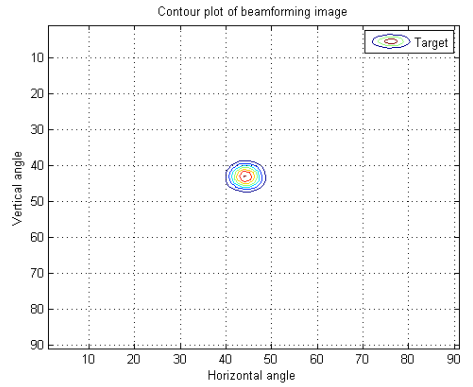
(b) Doppler signals of object 1 in Sensor 1



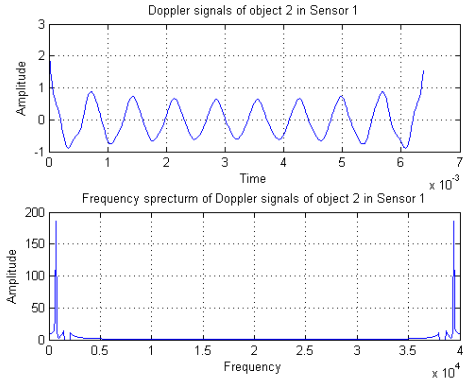
(c) Beamforming of object 1



(d) Surface plot of the beamforming image of object 1

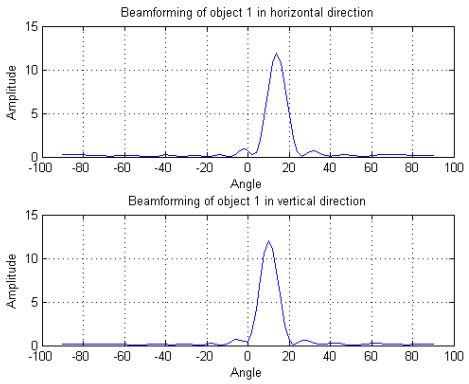


(e) Contour plot of the beamforming image of object 1

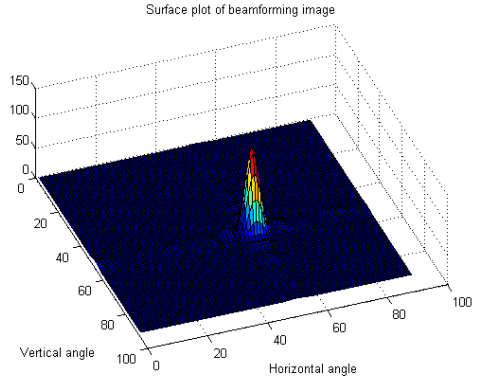


(f) Doppler signals of object 2 in Sensor 1

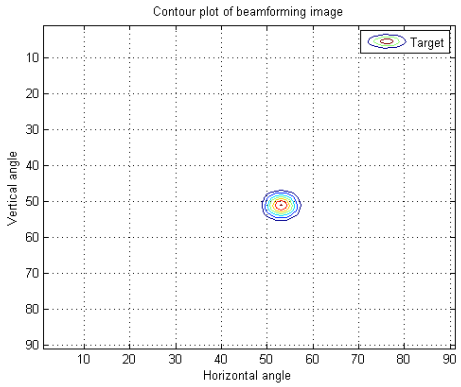
Figure 4.22: Doppler beamforming without ZF components (1)



(a) Beamforming of object 2

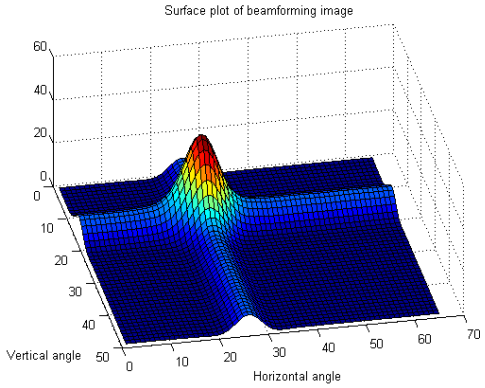


(b) Surface plot of the beamforming image of object 2

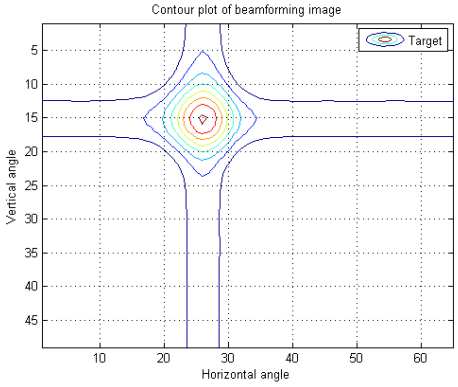


(c) Contour plot of the beamforming image of object 2

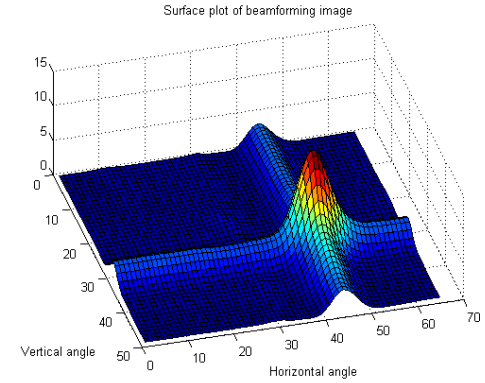
Figure 4.23: Doppler beamforming without ZF components (2)



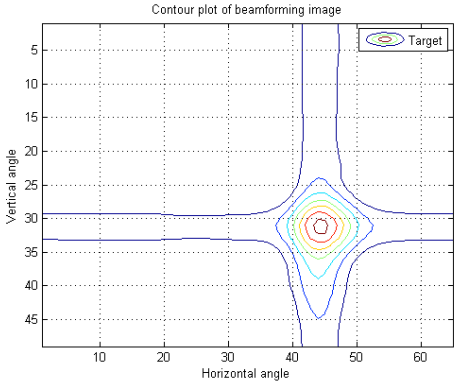
(a) Surface plot of the beamforming image of object 1



(b) Contour plot of the beamforming image of object 1

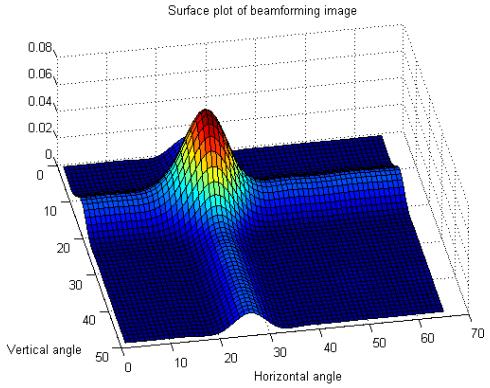


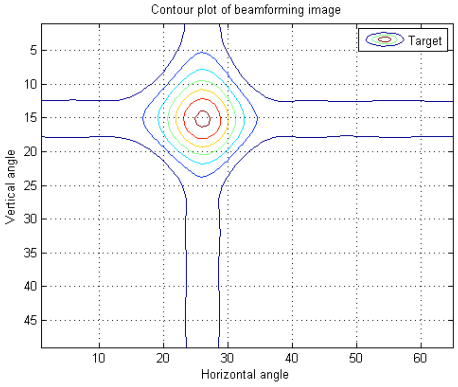
(c) Surface plot of the beamforming image of object 2



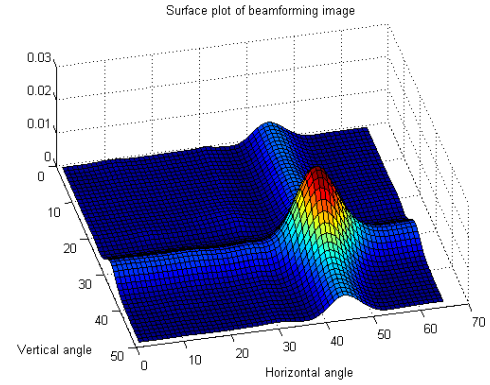
(d) Contour plot of the beamforming image of object 2

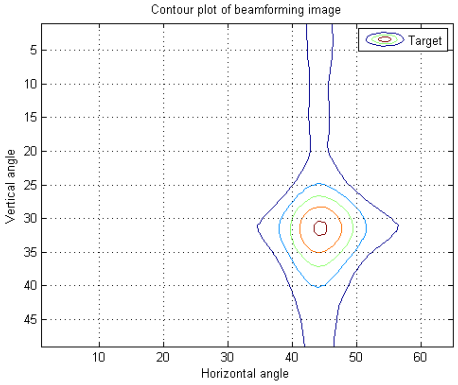
Figure 4.24: Crossed array Doppler beamforming Simulation with Gaussian windows





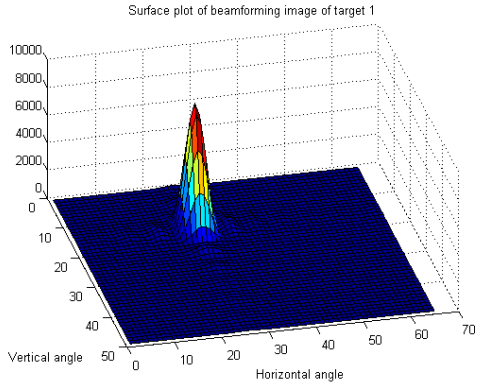
(a) Surface plot of the beamforming image of object 1 **(b)** Contour plot of the beamforming image of object 1



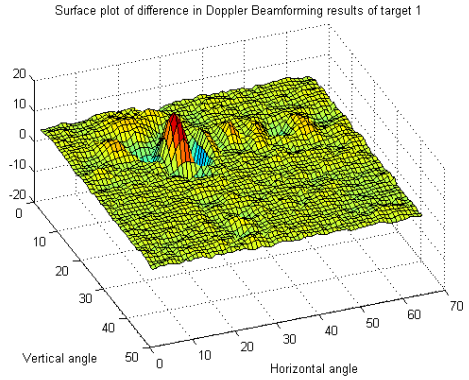


(c) Surface plot of the beamforming image of object 2 **(d)** Contour plot of the beamforming image of object 2

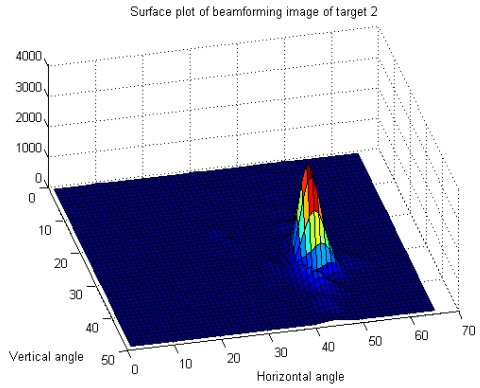
Figure 4.25: Crossed array Doppler beamforming simulation with Dolph-Chebyshev windows



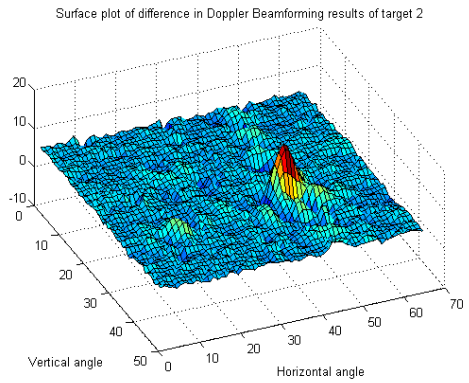
(a) Surface plot of the beamforming image of object 1



(b) Surface plot of the difference



(c) Surface plot of the beamforming image of object 2

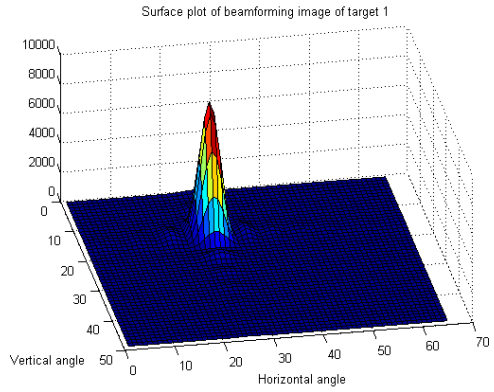


(d) Surface plot of the difference

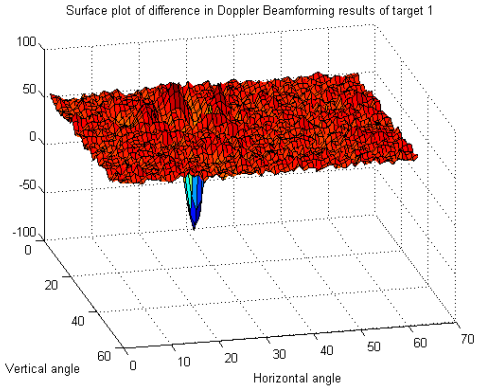
Figure 4.26: Noise analysis SNR=10dB

4.5.6 Performance when added noise

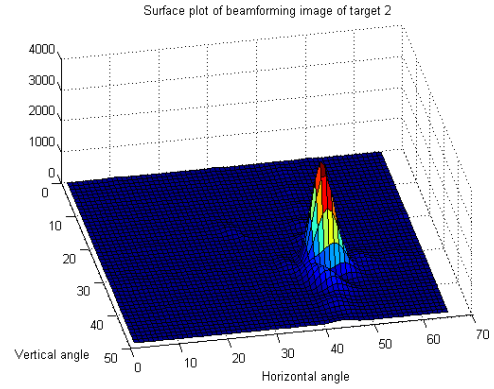
To test whether the system was robust or not, white Gaussian noise was added to the received signal of each channel. Then the beamforming images with and without noise are calculated to find out the root mean square errors (RMS). The results are shown from Figure 4.26 to Figure 4.29.



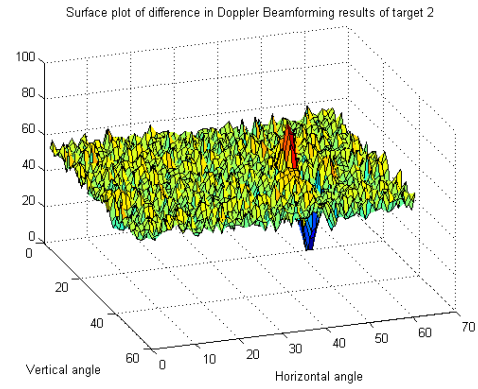
(a) Surface plot of the beamforming image of object 1



(b) Surface plot of the difference

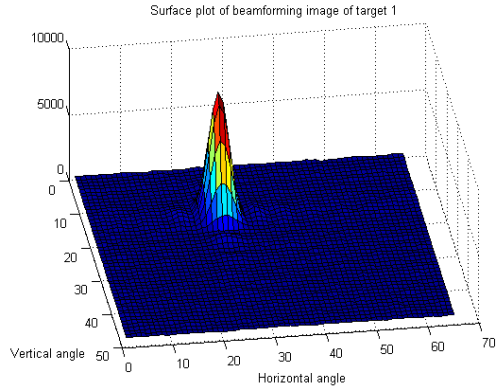


(c) Surface plot of the beamforming image of object 2

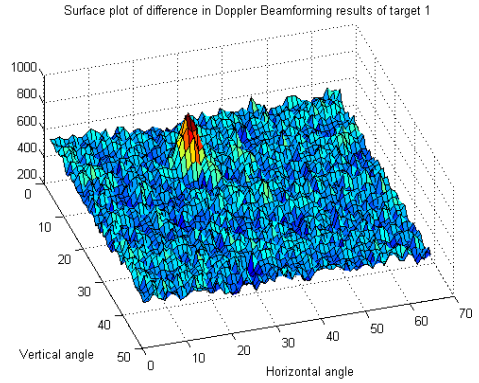


(d) Surface plot of the difference

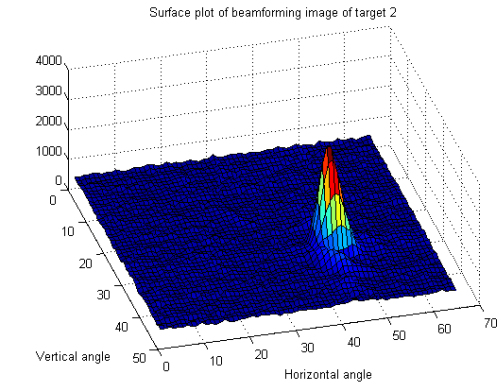
Figure 4.27: Noise analysis SNR=0dB



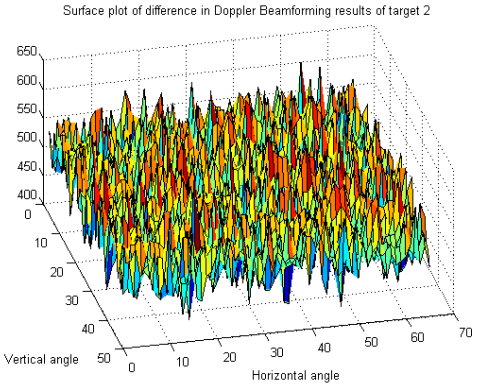
(a) Surface plot of the beamforming image of object 1



(b) Surface plot of the difference

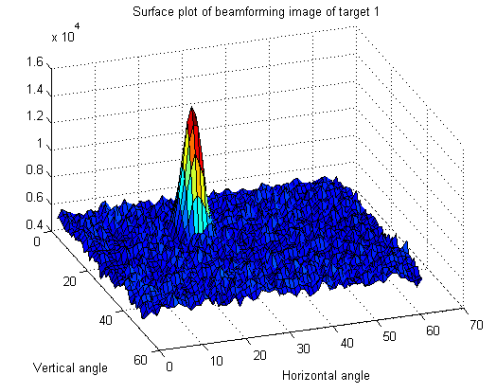


(c) Surface plot of the beamforming image of object 2

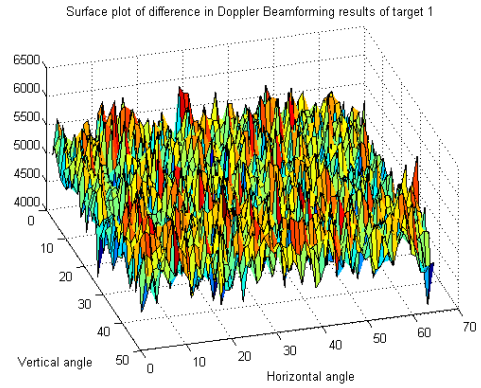


(d) Surface plot of the difference

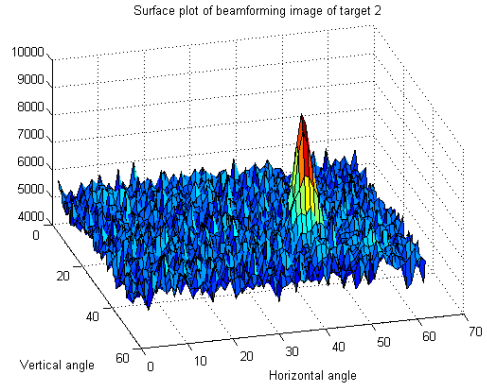
Figure 4.28: Noise analysis SNR=-10dB



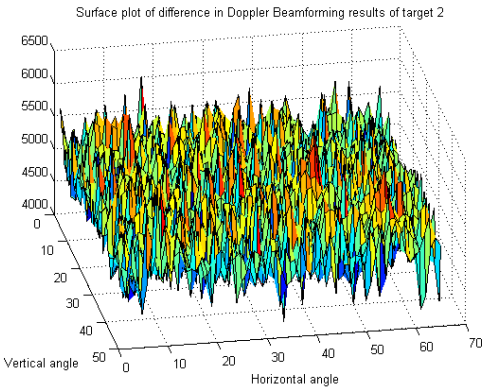
(a) Surface plot of the beamforming image of object 1



(b) Surface plot of the difference



(c) Surface plot of the beamforming image of object 2



(d) Surface plot of the difference

Figure 4.29: Noise analysis SNR=-20dB

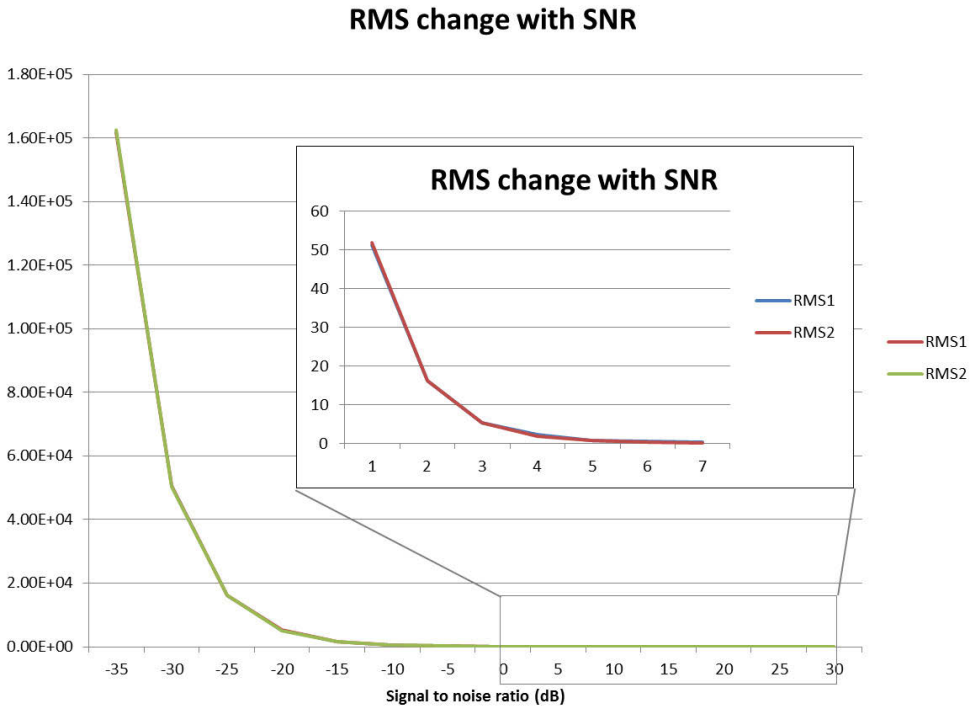


Figure 4.30: RMS errors changing with SNR

Here the beamforming results are quite robust even when the signal to noise (SNR) is lower than -10dB. When the SNR is -20dB, the target cannot be detected. The RMS errors changing with the SNR is also tested which is shown in Figure 4.30. When the SNR is higher than -10dB, the RMS errors are very small.

4.6 Mode switch between pulsed beamforming radar and Doppler beamforming radar

As pulsed beamforming radar needs to send pulsed waves while Doppler beamforming radar needs to send continuous waves, how can the appliance obtain both the range layer and the velocity layer at the same time? One solution is that the controller switches the radar between two modes: the Pulsed mode and the

Doppler mode. As in each frame the time is very short compared with the camera frame rate, so the two modes can be quickly switched. Though this method may sacrifice some time resolution, the frame rate is high enough for fusion with the optical image. The Doppler mode period can also be limited to filter long distance objects.

4.7 Example of fifth image layer

4.7.1 Simulation setting

The example image used in this chapter is similar to the image used in chapter 3. One person and two birds are supposed to be captured by the designed radar, which is shown in Figure 4.31.

The resolution of the image is 640 by 480. The Cartesian coordinate system can be used to measure the position of the objects.

The middle of the image is zero, which is the same as in chapter 3. The distance of the human is supposed to be 7 meters, and the distance of the left bird and the right bird is 10 meters and 13 meters, respectively. So the centroid position of the

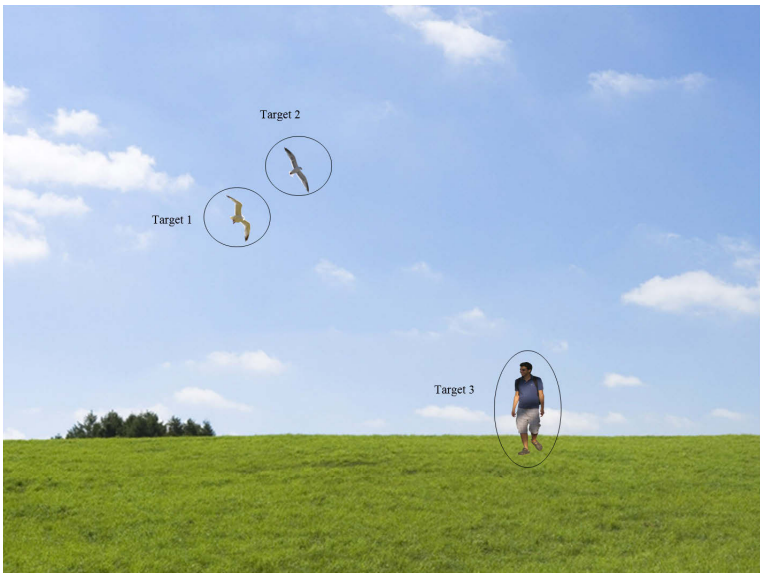
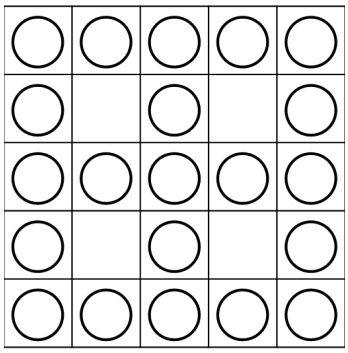


Figure 4.31: The supposed optical image

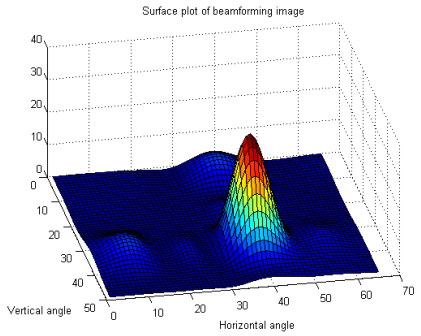
person is (3.26, 2.71, 5.92), the centroid position of the left bird is (-4.32, -2.02, 9.00), and the centroid position of the right bird is (-3.19, -4.58, 11.91). This time the radial velocity of the left bird is 0 m/s, the right bird is -10 m/s and the radial velocity of the human is 3 m/s.

4.7.2 Simulation on sparse rectangular array

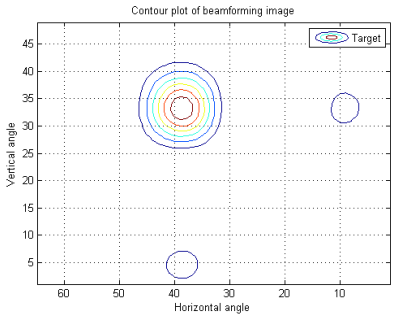
Here some types of sparse rectangular sensor arrays were used to test the resolution according to the sensor number and the attenuation ratio between the target and the ambiguities. As the situation when multiple targets have a similar velocity is rare, in the simulation, the velocities of the two targets were different. For computation cost considerations, the simulation was based on the FFT filtering. The results are shown from Figure 4.32 to Figure 4.42.



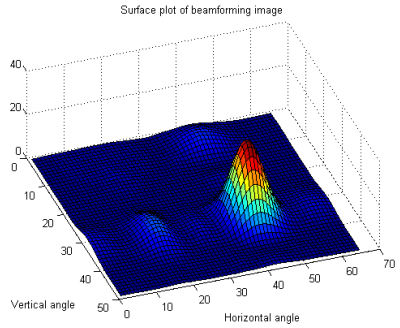
(a) The sensor array layout



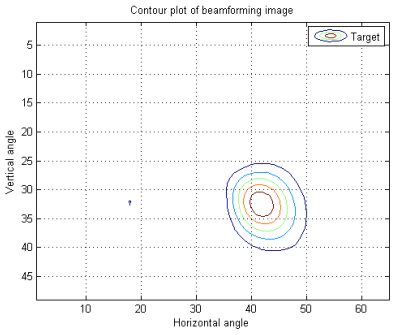
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

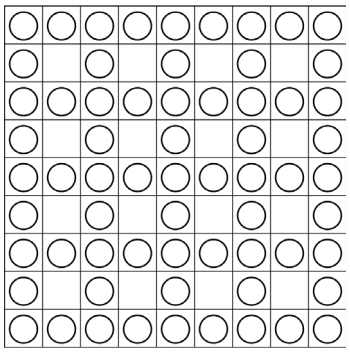


(d) Surface plot of the beamforming image of object 2

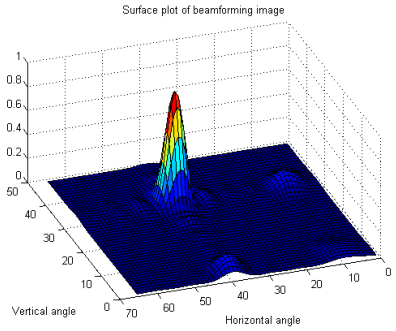


(e) Contour plot of the beamforming image of object 2

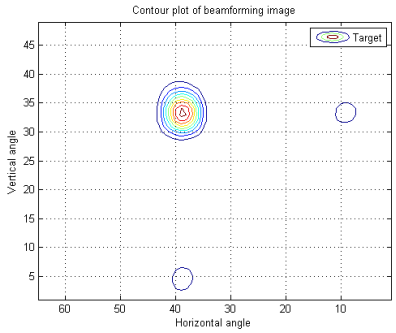
Figure 4.32: 5 by 5, hole size is 1 by 1



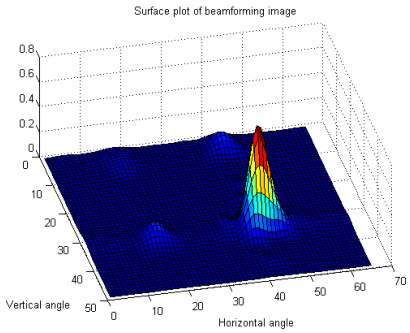
(a) The sensor array layout



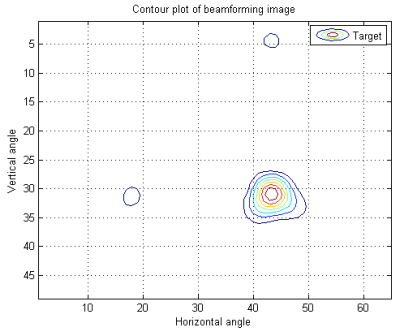
(b) The surface plot of the beamforming of object 1



(c) Contour plot of the beamforming image of object 1

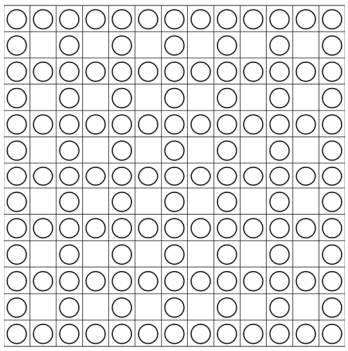


(d) Surface plot of the beamforming image of object 2

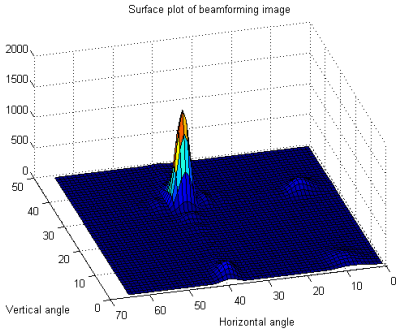


(e) Contour plot of the beamforming image of object 2

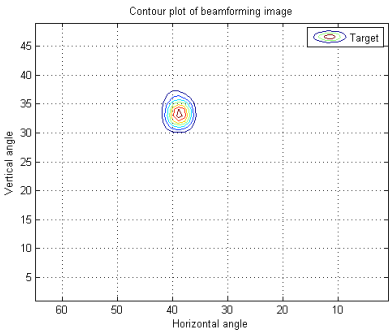
Figure 4.33: 9 by 9, hole size is 1 by 1



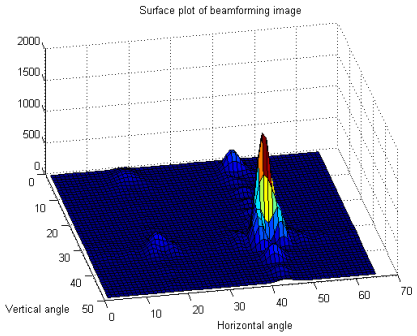
(a) The sensor array layout



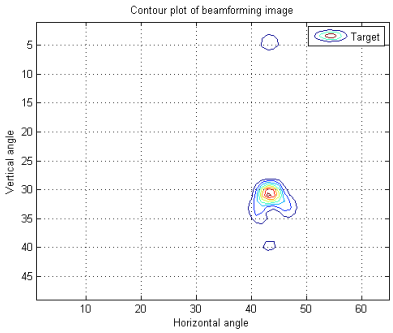
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

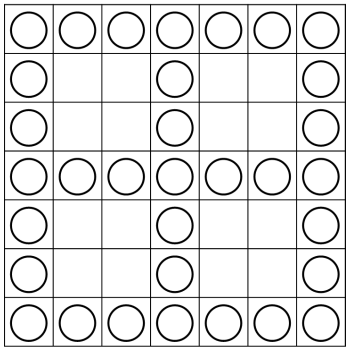


(d) Surface plot of the beamforming image of object 2

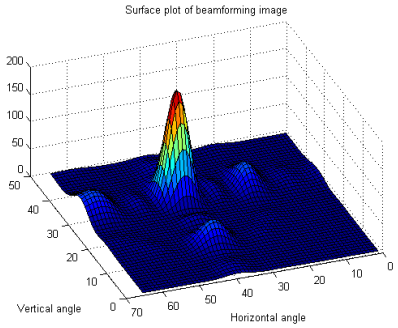


(e) Contour plot of the beamforming image of object 2

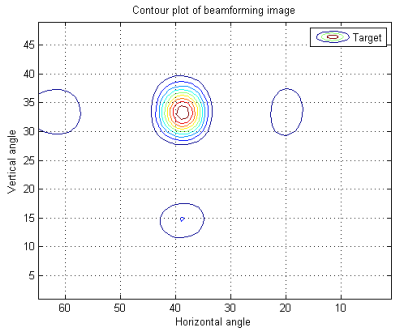
Figure 4.34: 13 by 13, hole size is 1 by 1



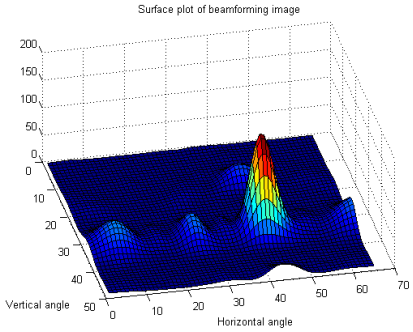
(a) The sensor array layout



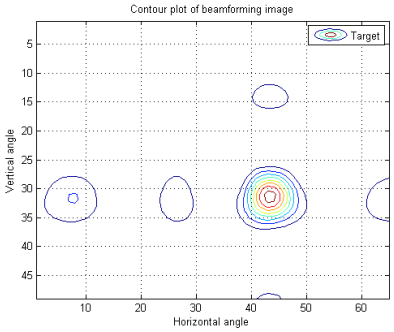
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

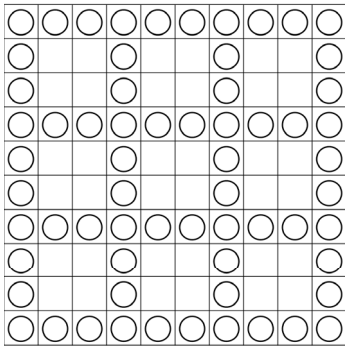


(d) Surface plot of the beamforming image of target 2

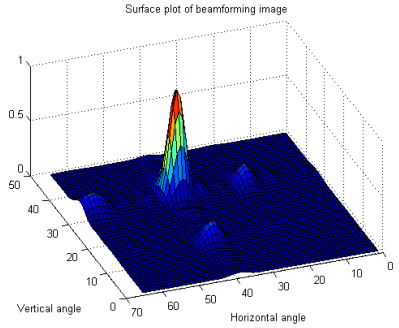


(e) Contour plot of the beamforming image of object 2

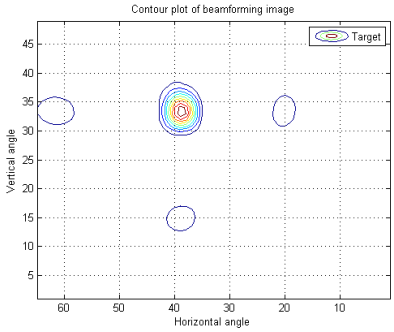
Figure 4.35: 7 by 7, hole size is 2 by 2



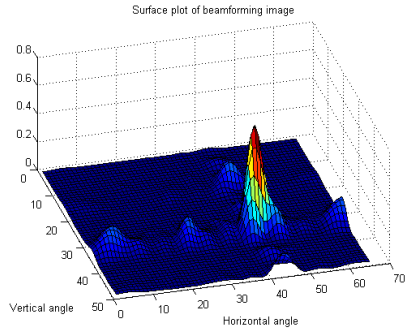
(a) The sensor array layout



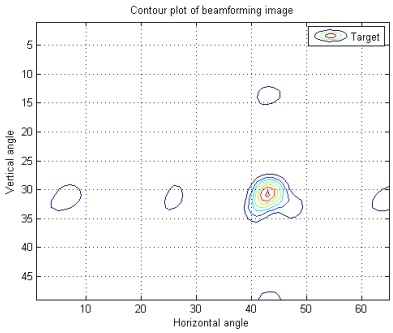
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming of object 1

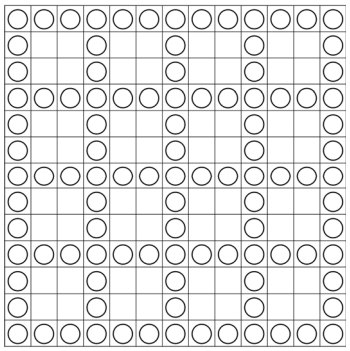


(d) Surface plot of the beamforming image of object 2

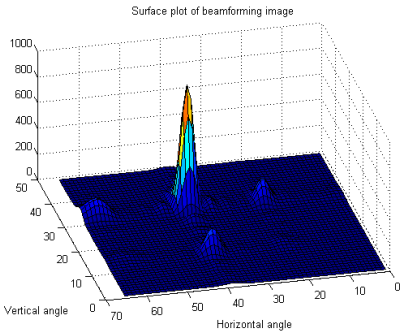


(e) Contour plot of the beamforming image of object 2

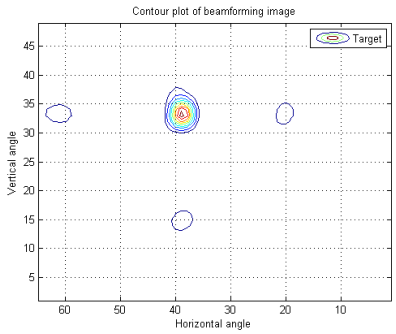
Figure 4.36: 10 by 10, hole size is 2 by 2



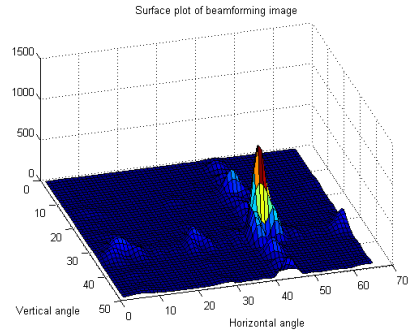
(a) The sensor array layout



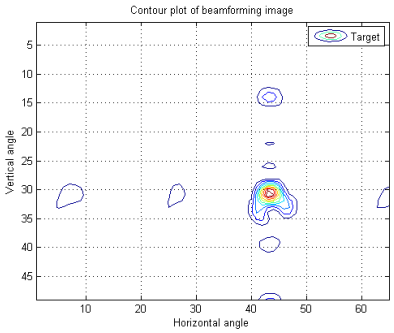
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

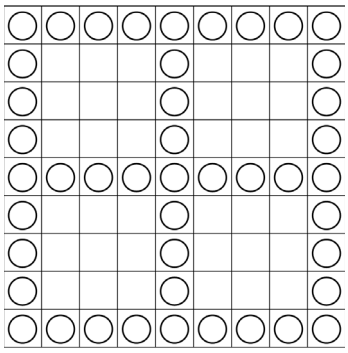


(d) Surface plot of the beamforming image of object 2

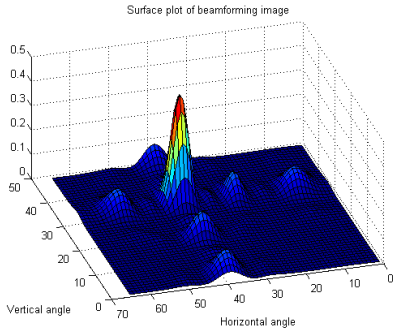


(e) Contour plot of the beamforming image of object 2

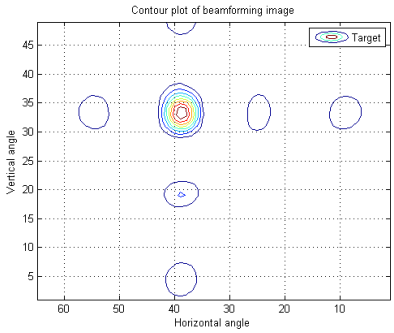
Figure 4.37: 13 by 13, hole size is 2 by 2



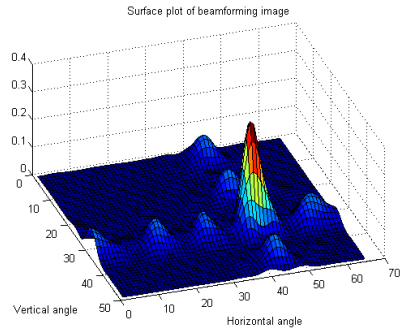
(a) The sensor array layout



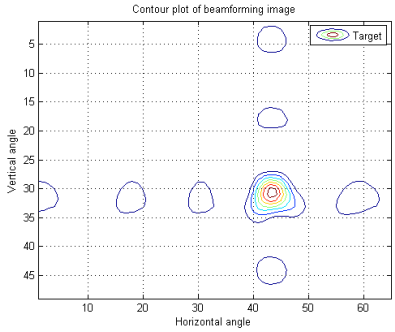
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image

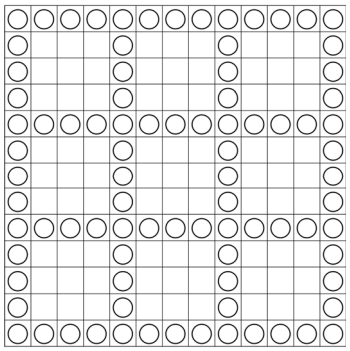


(d) Surface plot of the beamforming image of object 2

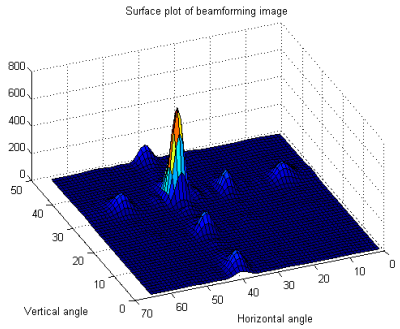


(e) Contour plot of the beamforming image of object 2

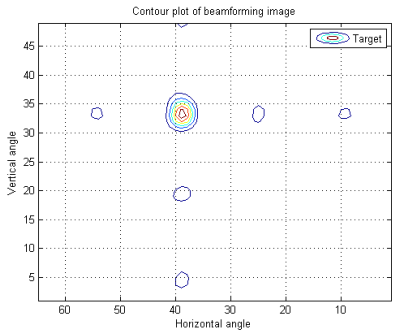
Figure 4.38: 9 by 9, hole size is 3 by 3



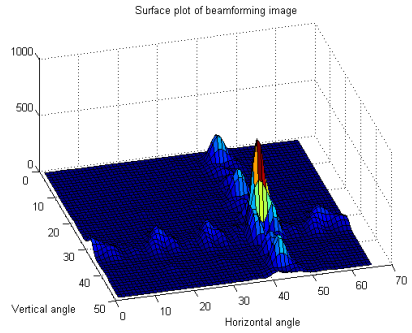
(a) The sensor array layout



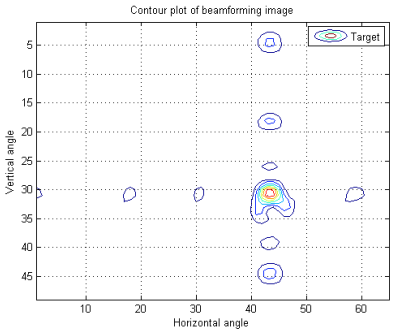
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

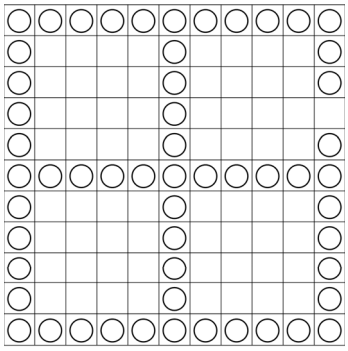


(d) Surface plot of the beamforming image of object 2

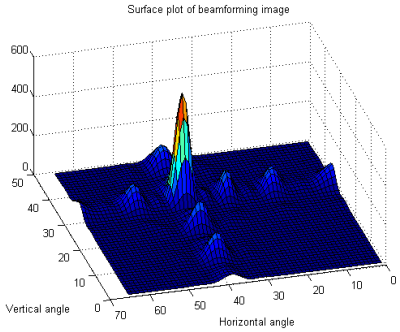


(e) Contour plot of the beamforming image of object 2

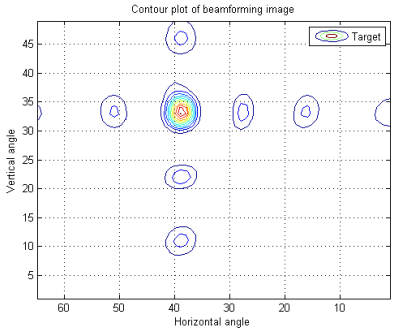
Figure 4.39: 13 by 13, hole size is 3 by 3



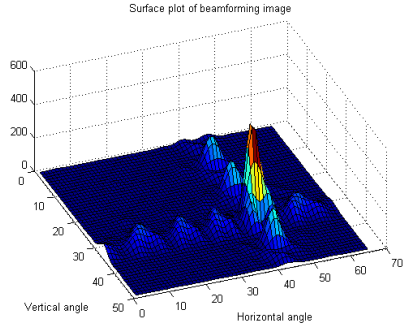
(a) The sensor array layout



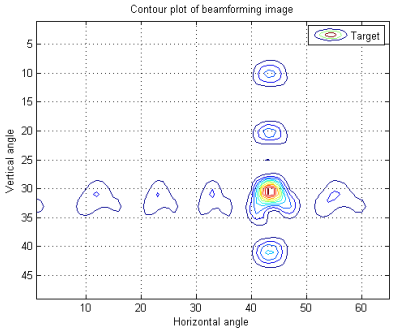
(b) Surface plot of the beamforming image 1



(c) Contour plot of the beamforming image of object 1

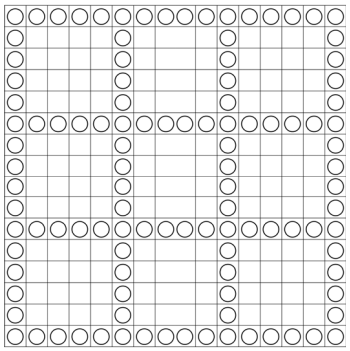


(d) Surface plot of the beamforming image of object 2

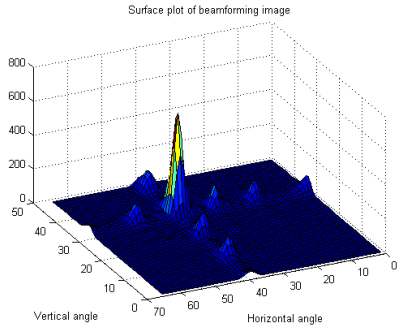


(e) Contour plot of the beamforming image of object 2

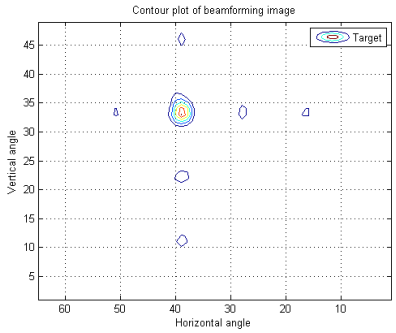
Figure 4.40: 11 by 11, hole size is 4 by 4



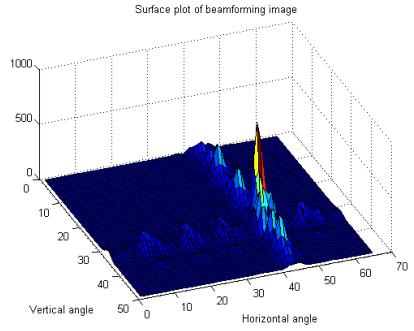
(a) The sensor array layout



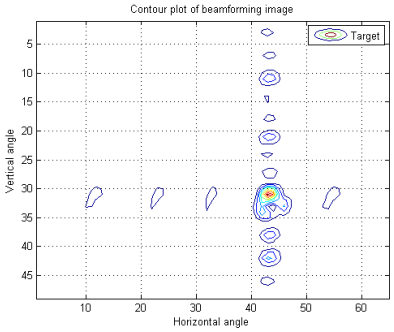
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1

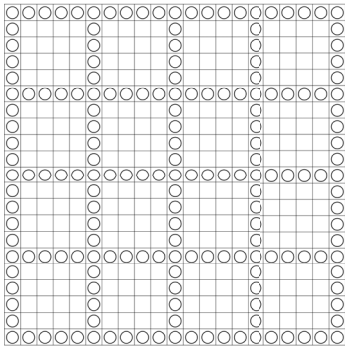


(d) Surface plot of the beamforming image of object 2

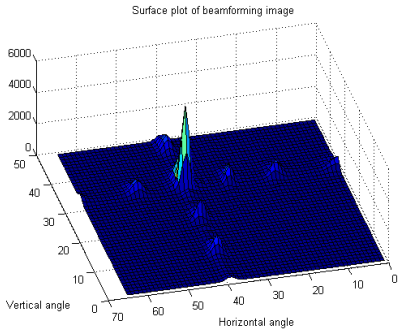


(e) Contour plot of the beamforming image of object 2

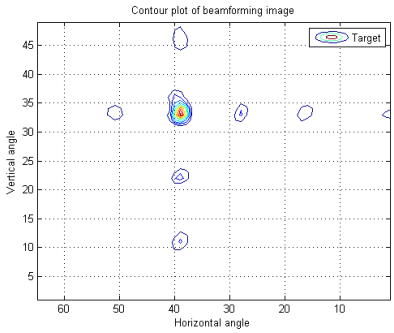
Figure 4.41: 16 by 16, hole size is 4 by 4



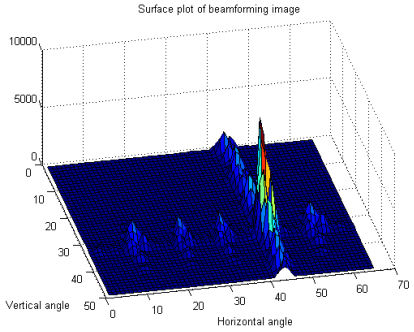
(a) The sensor array layout



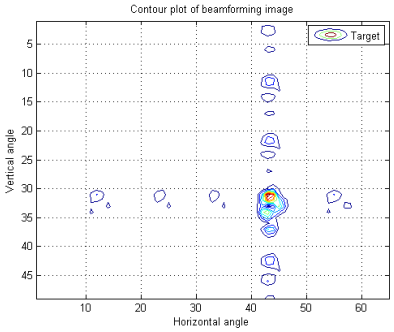
(b) Surface plot of the beamforming image of object 1



(c) Contour plot of the beamforming image of object 1



(d) Surface plot of the beamforming image of object 2



(e) Contour plot of the beamforming image of object 2

Figure 4.42: 21 by 21, hole size is 4 by 4

4.8 Proposed algorithms

According to the analysis on the Doppler beamforming theory and the simulation results in this chapter, the design and the algorithm are concluded as follows.

4.8.1 Sensor array structure

The sparse rectangular sensor arrays achieve the balance between the hardware resource, computation load and memory consumption. In some cases, if there are fewer than 3 targets, and the range difference between targets is clear, two line multiplication beamforming can be used as this method can have high a main lobe to side lobe ratio with a low computation cost. The target with a different range can be separated by the pulsed mode and they will not be shown in the same beamforming image. In chapter 5, the method to reduce ambiguity by fusion with the optical image is described.

4.8.2 Procedure to construct the pulsed beamforming image

Window function can reduce the main lobe to side lobe ratio, and the Dolph-Chebyshev window has a good performance. Frequency domain beamforming has a high resolution. Zero frequency (ZF) components need to be filtered to improve the signal to noise ratio. Also, FFT filtering is required to separate targets with different velocities. Though FIR filter banks have better performance, for the cost, computation load and memory consumption consideration, the FFT filtering is preferred. By the comparison of the performance between FFT filter and FIR filter banks, the results show the performance of FFT filtering is acceptable.

The FFT series length in Doppler beamforming is much longer than pulsed beamforming. In pulsed beamforming, to increase the distance resolution the pulse length needs to be reduced. In Doppler beamforming, for some low moving objects, if the pulse is too short, the full length of the Doppler signal in the series may be shorter than a wavelength. In this situation, the Doppler frequency may not be accurate and the position of the target in the image may be shifted considerably. In simulation, the benefits of increasing the FFT length is shown. For some gait analysis applications, the Doppler frequency always changes with time. In this case, the FFT length needs to be short to adapt to the change of the Doppler frequency. The window functions are also applied to increase the main lobe to side lobe ratio.

The final algorithm is shown in Figure 4.43.

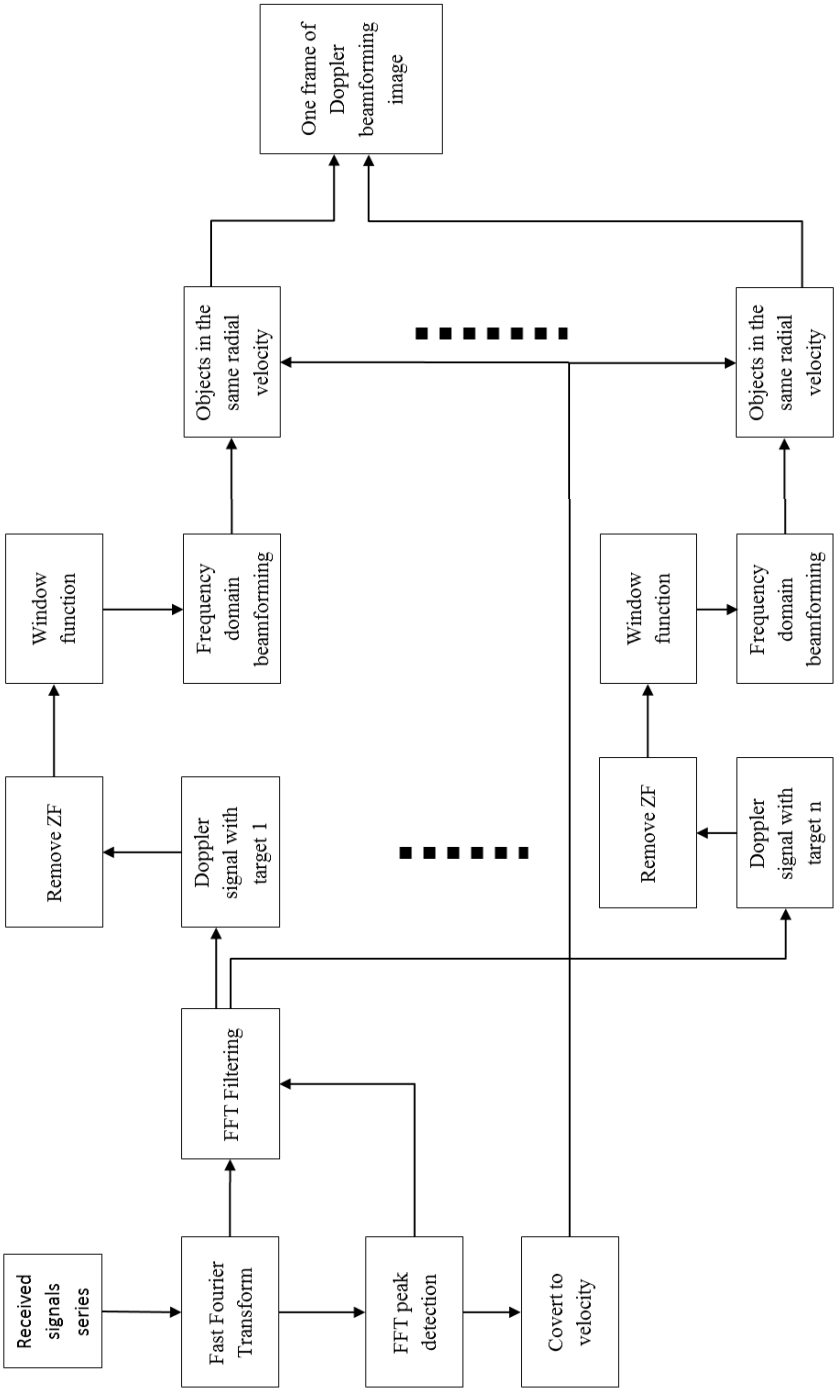
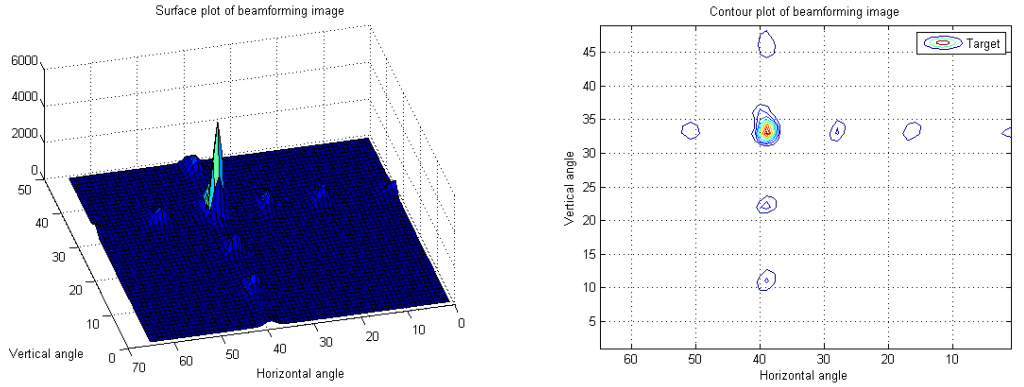


Figure 4.43: Final Doppler beamforming algorithm



(a) Surface plot of the beamforming image

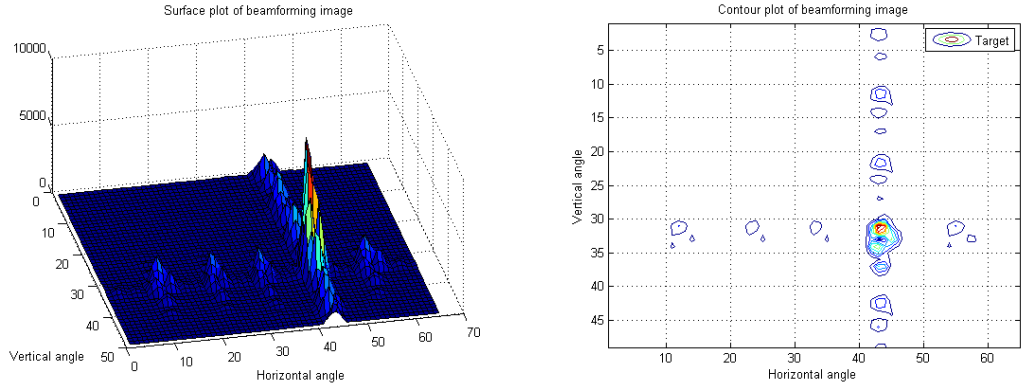
(b) Contour plot of the beamforming image

Figure 4.44: The target one (right bird) Doppler beamforming image

4.8.3 Simulation of the example frame

Below is the simulation results of the example image by the proposed algorithm. There are three targets in the environment, while two of them are moving objects with different velocities. According to the analysis above, the moving objects can be detected in two different frames. Figure 4.44 and Figure 4.45 show the beamforming images of target one and target two, separately.

When Figure 4.44 and Figure 4.45 are multiplied with the related velocities, the example frames can be obtained, which are shown in Figure 4.46.



(a) Surface plot of the beamforming image

(b) Contour plot of the beamforming image

Figure 4.45: The target two (human) Doppler beamforming image

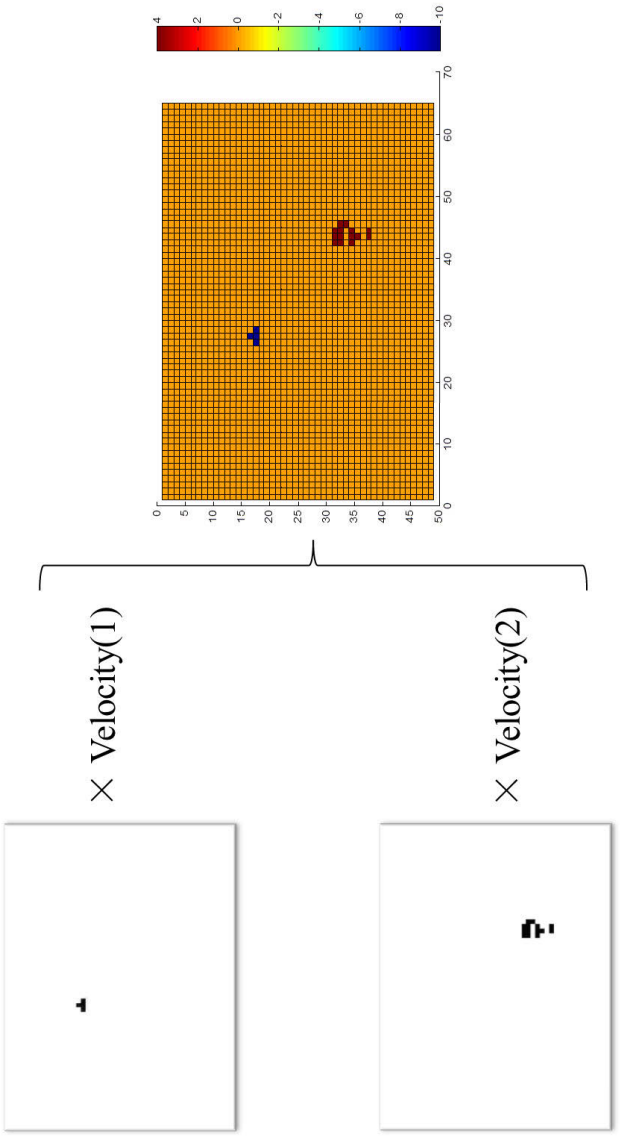


Figure 4.46: The proposed example Doppler beamforming image

5 Fusion of pulsed Doppler beamforming radar imaging with optical imaging to enhance target tracking - Addresses Proposition 3

5.1 Introduction

In the last two chapters, the creation of the fourth layer which represents the object's location and the fifth layer which represents the object's radial velocity were described. Compared with the light intensity information in traditional RGB layers, these two layers show the presence and kinematic state of the objects. This information is more useful for target tracking.

Much research effort has been put into finding the best way to increase the ability to track objects. Actually, due to the complexity of the application environment and diversity of sensor types, in each application, the algorithm should be re-designed. As it is hard to create a general algorithm, in this chapter, several scenarios are studied to use the extra two image layers to enhance object tracking.

The common image fusion techniques can be divided into three categories: pixel level fusion, feature level fusion, and decision level fusion. In most cases, information from different sensors is fused in the same layer. There are also some fusion methods in multiple levels. For example, processing the image from one kind of sensor, subtracting the features and making a decision can speed and enhance the processing of the raw data of other sensors. Furthermore, in dynamic systems, sensors need not only track the current state of objects but also need to estimate the future state of objects. Kalman filtering and its variations are the most popular methods to estimate the states. Sensors can independently update their estimation of the objects and fuse the results or make an estimation based on other sensor's information.

According to the stage when the fusion occurs in processing, the fusion methods can be divided into three categories: (1) Radar images and optical images are processed in parallel and fusion occurs in the same procedure of processing. (2) Radar images are processed first and the results are used to speed or enhance the optical image processing. (3) Optical images are processed first and the results are used to speed or enhance the Radar image processing.

5.2 Radar and camera calibration and image fusion

Before fusion the camera and radar should be calibrated and the captured optical image and the radar image should be carefully registered.

5.2.1 Calibration of optical image and radar

How the position of real objects is projected into an optical image is determined by the relative position of objects and cameras, the camera focal length and the size of films, which is CCD or CMOS in digital cameras. The relationship can be described as follows.

$$\theta = \arctan\left(\frac{s}{2}/D\right) \tag{5.1}$$

where, s is the diameter of the lens and D is the camera focal length.

So the angle of view is

$$Aov = 2\arctan\left(\frac{s}{D/2}\right) \tag{5.2}$$

In film cameras, the size of a standard 135 film is 35mm width by 24mm height. In digital cameras, there are many types of sensor size. The sensor size of full-frame digital SLR cameras is 36mm by 24mm film frame. Also half-frame digital SLR cameras use sensors with a size of 18mm by 24mm. In portable digital cameras, web-cams and phone cameras, the sensor sizes are usually much smaller. For example, the popular diagonal length is 1/1.8 inch or 1/2.4 inch. To better measure the object size, the focal lengths are usually multiplied with a ratio to get the equivalent focal length to 135 film. For example, a half-frame SLR would have a lens with 18mm focal length. The ratio between full-frame SLR and half frame SLR is 1.5. So the equivalent focal length is 18mm*1.5=27mm.

In a monitor system, the angle of view is wide to prevent the object running out of the vision area. This project assumes the angle of view is 120°. So the equivalent focal length is 10 mm. For a digital image, it is easy to adjust the focal length. However for a radar system, the angle of view is fixed. The radar image can be mathematically fitted to the optical image using a technique which is similar to digital zoom verses optical zoom.

5.2.2 Image registration

In most image fusion algorithms, images are supposed to be perfectly registered. Unregistered images can bring errors to the fused image. The accuracy of the fusion is very sensitive to the alignment of source image [12]. The angles of view of the radar and the camera are usually slightly different. Even though the coordinate system transform matrix is carefully chosen, the locations of objects in different layers have slight shifts [66].

Image registration can be divided into two categories. In the first case, the difference between two images is usually the spatial difference. In the second case, the difference between two images is not only the spatial difference, but also other changes, such as the difference in illumination and the data type difference as they are captured by different types of sensors. However, only the spatial difference can be removed by registration and the result is not an exact match between two images.

In this radar system, the resolution of the radar system is not intended to increase to the optical image level. The ratio of pixel number of the optical image to the radar image may be around 10-50 to 1. So the registration method is that one pixel in the radar image is equivalent to a region in the optical image. The key idea to register the radar and optical image is to find some control points which can be shown in both the radar image and the optical image. For example, some objects can emit both light and Doppler signals, which can easily be detected by Doppler radars and cameras.

5.3 Scenario 1: Fusion of video camera frames with Doppler beamforming radar to enhance moving object detection

5.3.1 Introduction

In robotic systems and computer vision, real-time moving object detection is an important step. Both video cameras and radars have the ability to detect moving objects. However there are some difficulties which are shown in Table 5.1.

Apparently, some advantages and disadvantages of these two types of sensors in detecting moving object are complementary. Much research has been done on how to improve the accuracy of moving object detection in one type of sensor. It is supposed that the fusion of these two methods can enhance the performance. The system should be at a low cost and robust. The improvement is achieved

Table 5.1: Difficulties in moving object detection

Sensor type	optical image	radar image
shortfall	Noise in images Complex object motion Computation load in real-time processing Cannot represent the three dimensional information Scene illumination changes Occlusions and complex object shapes Shadows	Noise in signals Complex object motion Computation load in real-time processing Sound sources interference for acoustic radar and radio wave emitters interference for radar Low resolutions

by providing complementary information to another type of sensor to reduce the uncertainty of detection.

In this scenario current methods are first investigated and some fusion methods for background detection are studied. Then the methods are demonstrated and evaluated by simulation in Matlab.

5.3.2 Background

In images, the object to be detected is called the foreground and the rest is called the background. There are two methods to determine the foreground and the background: (1) The optical flow method subtracts the background by monitoring the pattern of apparent motion, surface and edges in a visual scene caused by the relative motion between the camera and the objects. The problem with the optical flow method is that light changes will greatly affect the optical flow. The computation load of optical flow is high. Various configurations of optical flow sensors exist, which combine the image sensor with a processor programmed to run an optical flow algorithm. (2) The basic idea of the second method is subtracting the current frame with a reference frame. The reference frame should have no moving object and be regularly updated to adapt to the varying luminance conditions and geometry settings. Methods in this category require a relatively low computation load and low memory consumption. However, the illumination of the environment could change, and slight position changes of the camera could bring much noise to the subtracted image. Frame differencing methods only compare the difference between several continuous frames. This method does not need the prior knowledge of the background or scene, so it is much more robust to light changes. Many improved methods have been studied, such as running Gaussian average and kernel density estimation. These methods have limitations:

- In most case, only the edge of the moving object can be detected.
- When the objects are radially approaching or leaving the camera, the differences in continuous frames are tiny, which make the object hard to be detected.
- If the environment contains complicated turbulent objects such as tree leaves or rain, this method cannot remove the interference.
- The shadows of the objects can also be detected as moving objects.
- The frame difference in slow moving objects are tiny, and for a fast moving object, due to the low frame rate, one object can be detected as two objects.
- In an environment with low illumination, or the moving object has similar light intensity or similar color to the background, the performance is not stable.
- The threshold for converting the binary image needs to change when illumination changes.

From the radar point of view, a radar with Doppler beamforming processing can only sense moving objects. One single frame can tell where the moving objects are. The resolution is much lower than optical images, so it is difficult to detect the edge. Doppler beamforming does not rely on light but is affected by sound resources for acoustic radars or by electromagnetic waves for electromagnetic radars. Another benefit of Doppler beamforming is that it can separate objects with different velocities. In crossed sensor array radar, there are ambiguities when multiple objects are moving in the same radial velocity. If a full rectangular array is used, there are no ambiguities, but the computation load increases sharply when the resolution is improved.

Based on the characteristics of moving object detection in video cameras and Doppler beamforming radars, this scenario aims to fuse these two moving object detection methods to enhance the moving object subtraction qualities, such as solving the turbulence interference problem and shadow problem in background subtraction in optical images.

5.3.3 Frame differencing

As optical flow requires some specified hardware and needs to analyze the light changes of each point continuously, for cost consideration, this method is not discussed in this scenario. Also though a stable background can detect both stationary and moving objects, in this case, only moving objects need to be detected. Therefore the fusion of image processing using frame differencing and Doppler beamforming to enhance the moving objects detection is discussed in this section.

The basic idea of frame differencing is to compare the absolute difference in each corresponding pixel of successive frames and to divide the image into unchanged and changed regions.

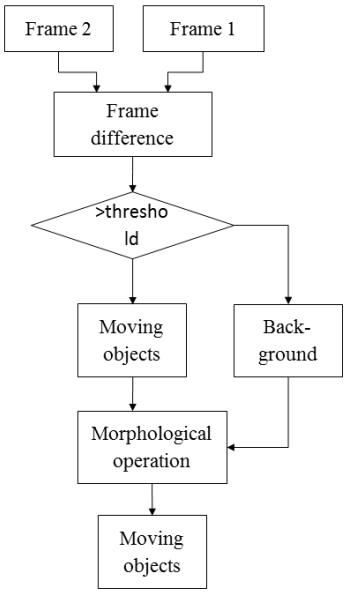


Figure 5.1: Frame differencing procedure

In particular, frame differencing takes two consecutive frames as inputs and produces a new image by subtracting the second image pixel values from the first image pixel values. For color images, the images need to be converted to grey scale values first. The general operation is as follows:

$$I_{diff}[i,j] = I_1[i,j] - I_2[i,j] \tag{5.3}$$

where $I_{diff}[i,j]$ is the grey scale value in pixel (i,j). $I_{diff}[i,j]$ need to be converted into a binary image by using binary threshold and the resultant binary image is processed by morphological operation. Figure 5.1 shows the procedure of frame differencing.

Static background subtraction is sensitive to illumination changing, camera motion and unwanted objects such as tree leaves. Another problem is that the shadow of the moving object is also taken as a moving object. Frame differencing can deal with part of the illumination changing and the camera motion problem.

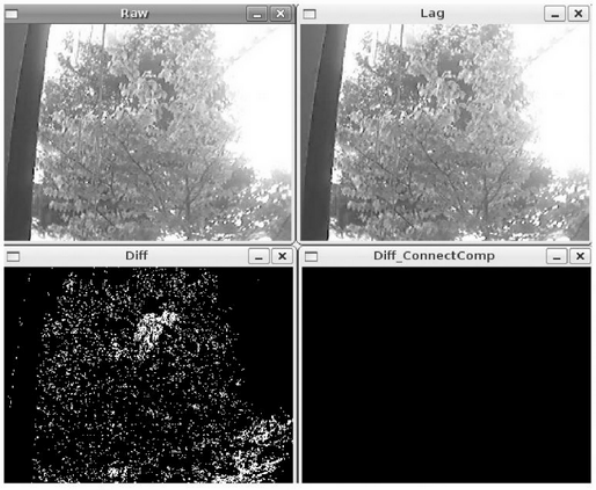


Figure 5.2: The small moving leaves can be removed by the connected component method [3]

5.3.3.1 Connected components

Connected components labeling is a method to remove the redundancy and improve the image quality based on the assumption that all the noise areas are small. For example, for a tree with leaves in the background, connected components labeling can solve part of the problem. The solution can be seen in [3] and the simulation is shown in Figure 5.2.

Connected components labeling marks whether the neighborhood is connected or not. For most of the time, an object is a rigid object with a size much bigger than the interference. The small areas will be filtered and only the large rigid objects are left.

The drawback is that in some applications, connected components can not solve the problem. In the case of the ball tracking application, the size of the target is small. Also, for a long distance or slow moving objects, as only the edge of the target can be detected, they are not marked as a region, therefore the moving object may be removed as noise.

5.3.4 Doppler beamforming

Doppler beamforming processing is shown in chapter 4. For each receiver, the carrier needs to be first subtracted, then the signals need to be transferred into the

frequency domain to distinguish objects with different radial velocities. The beamforming algorithm in chapter 4 is applied after that. The beamforming image with different Doppler frequency is summed to create the final Doppler beamforming image.

5.3.4.1 Ability of Doppler beamforming in subtracting small unwanted moving objects

In Doppler beamforming, the velocities of multiple small unwanted moving objects in the background are usually slow. For a ball tracking system, the velocity of the moving object is much higher than that of the unwanted moving objects in the background, therefore, objects can be easily subtracted by velocities. As tree leaves do not have stable velocities, Doppler beamforming can filter the regions with discontinuous velocities.

In image frame differencing, only the edge of moving objects can be detected. The edge information is not sufficient to subtract the moving object especially for objects with low velocities. By fusion of Doppler beamforming images, the subtraction of targets can be improved. Further processing such as template matching can be used to confirm the objects.

5.3.5 Fusion of optical frame differencing with Doppler beamforming images

In this section, the pixel level fusion of frame differencing images with Doppler beamforming images is discussed. As the feature subtraction of Doppler beamforming is difficult and the meanings of pixel values in the optical image and the Doppler beamforming image are different, there is no correlation between these two types of images. The sensors need to be calibrated carefully in advance.

The pixel values in the optical image represent the light intensity of objects, while the pixel values in a Doppler beamforming radar image represent the possibilities of moving objects. However, the optical frame differencing image can also be considered as representing the possibilities of the moving objects. Based on this assumption, though the outputs from these two types of sensors are different, the values have similar meanings.

The shortfall of the optical frame differencing image is that it is not reliable because of the multiple small moving objects in the background, illumination changes and shadow, while Doppler beamforming radar does not rely on illumination. Connected components labeling can solve part of the ambiguity problem but the effects are limited in most applications. Therefore, the possibilities to remove ambiguity in frame differencing is studied.

5.3.5.1 Principal component analysis

Principal component analysis (PCA) [67] is a method to project higher dimensional data into lower dimensional space, which is widely used in feature extraction. PCA finds a linear projection and projects all the data into a new coordinate system. In this new coordinate system, the first coordinate presents the most significant changes (the first principal component) and the second coordinate presents the second most significant changes (the second principal component). Lower layer data are thought to present the original data better and higher layer data are thought to present noise. By ignoring the high layer data and keeping the lower layer data, the dimensions can be reduced.

Suppose the original data are $(y_1, y_2, ..., y_n)$, where n is the number of samples, y_i is a vector whose length is M. Suppose the dimension of eigenvector needs to be reduced from M to m, and the dimension of y_i need to be reduced into f_i; these steps can be followed.

First calculate the average value μ and covariance matrix C of original sample set:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{5.4}$$

$$C = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)^T \tag{5.5}$$

If the rank of matrix C is L, L eigenvalues $\lambda_1, \lambda_2, ..., \lambda_L$ and eigenvector $e_1, e_2, ..., e_L$ of matrix C by singular value decomposition (SVD) can be calculated, respectively.

Generally speaking, if the eigenvalue is larger, the change in original data is larger. Sort the eigenvalues by size, keep the first m eigenvalues and their eigenvectors. The original data y_i can be converted to f_i , by the equation below:

$$f_i = [e_1, e_2, ..., e_m]^T \quad y_i \tag{5.6}$$

So the dimension is reduced from M to m. The weight of them in original data is:

$$\rho = \sum_{i=1}^{m} \lambda_i / \sum_{j=1}^{L} \lambda_j \tag{5.7}$$

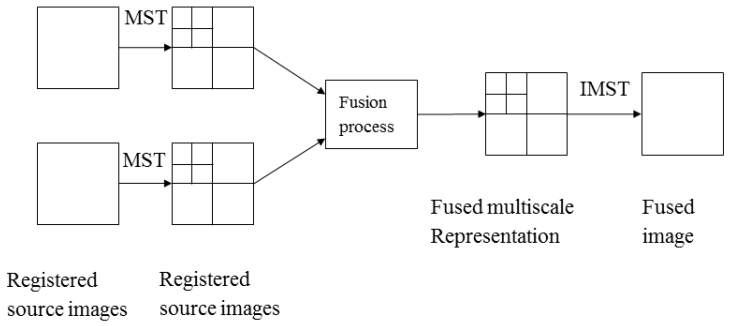


Figure 5.3: Multi-scale image fusion procedure

5.3.5.2 Multi-scale Decomposition-based fusion methods

Pixel level fusion methods can be divided into multi-scale decomposition based fusion methods and non-multi-scale decomposition based method. For images with different resolutions, a pixel in a low resolution image usually reflects a region of pixels in a high resolution image. One way to fuse these two kinds of images while keeping the details of the high resolution image is fusing them by a multi-scale decomposition-based fusion method.

Figure 5.3 illustrates the block diagram of an image fusion scheme based on multi-scale analysis.

Generally, multi-scale decomposition-based fusion methods include the pyramid transform (PT), the discrete wavelet transform (DWT) and the discrete wavelet frame.

5.3.5.3 Non-multi-scale decomposition based methods

Basically, non-multi-scale decomposition based methods include weighted averaging, nonlinear method, estimation theory based method and color composite fusion.

The basic idea is the additive and multiplicative fusion; these methods are usually used to enhance the contrast. For example

$$I_f = A\left(w_1 I_a + w_2 I_b\right) + B \tag{5.8}$$

$$I_f = A \cdot I_a \cdot I_b + B \tag{5.9}$$

where A and B are scaling factors and w_1 and w_2 are weighting parameters. I_f , I_a and I_b refer to pixel values of the final fused image and the input image a and b, respectively.

These methods can improve the SNR. The optimal weighting coefficients can be determined by the PCA and the weights for each input frame are obtained from the eigenvector corresponding to the largest eigenvalue.

5.3.6 Simulation and implementation

The videos are captured by a webcam and the Doppler beamforming images are created by a simulation according to the measurements. Then Matlab is used to process both the optical image and Doppler beamforming data and apply different fusion strategies into these images.

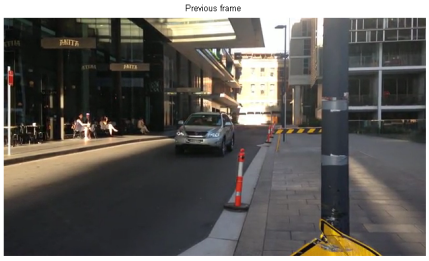
In case 1, there is a single moving object without walking people, which is shown Figure 5.4 and Figure 5.5.

Here the car is subtracted well, the fusion improves the moving object subtraction qualities. Both the multiplication method and weighted average method work well.

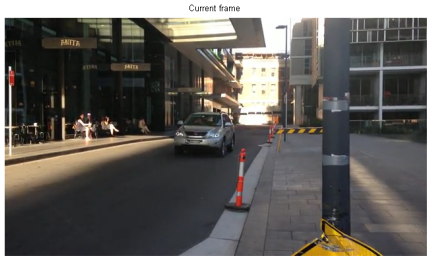
In case 2, there is a single moving object with tree leaves and walking people, which is shown in Figure 5.6 and Figure 5.7.

Here the tree effect can be mostly removed and walking people is not well subtracted due to the low velocity. The shadow of cars can be removed by multiplication. In weighted average, the shadow is reduced but still exists.

In case 3, there are multiple moving objects, which is shown in Figure 5.8 and Figure 5.9.

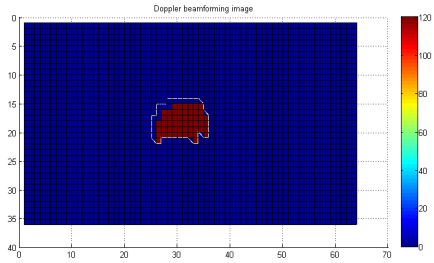


(a) Previous frame

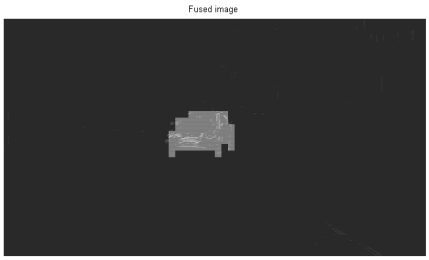


(b) Current frame

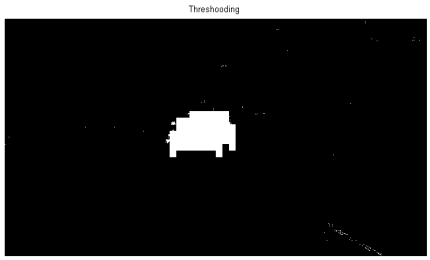
(c) Optical background subtraction



(d) The Doppler beamforming image



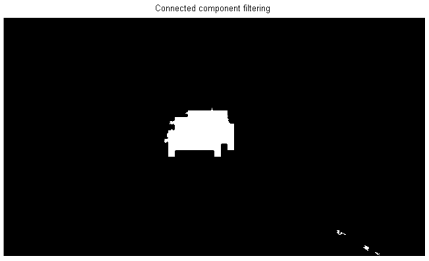
(e) Fused image



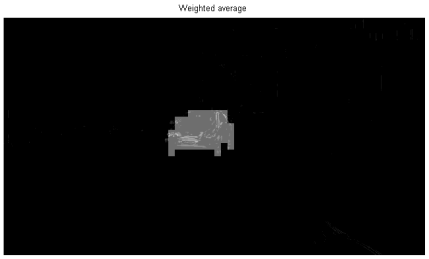
(f) Thresholding

Figure 5.4: Single moving object (1)

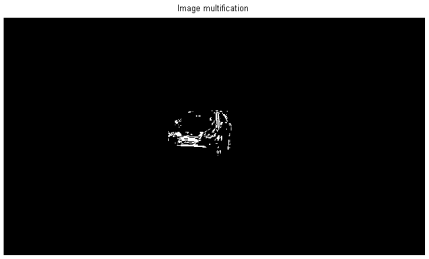
(a) Quality enhancement



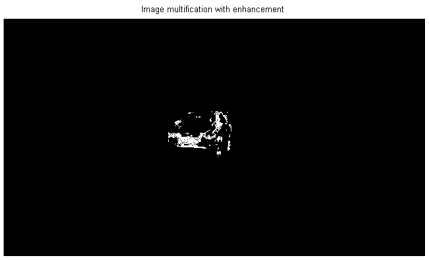
(b) Connected component filtering



(c) Weighted average



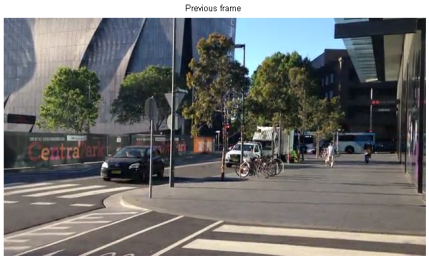
(d) Image multiplication



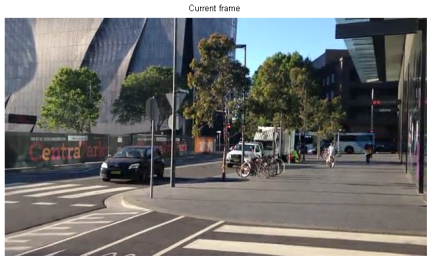
(e) Image multiplication with enhancement

(f) Subtracted object(s)

Figure 5.5: Single moving object (2)

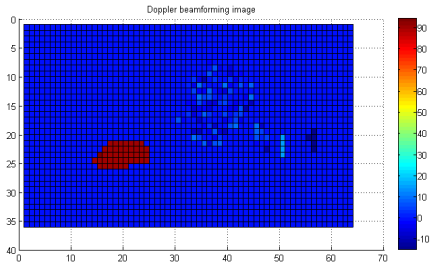


(a) Previous frame



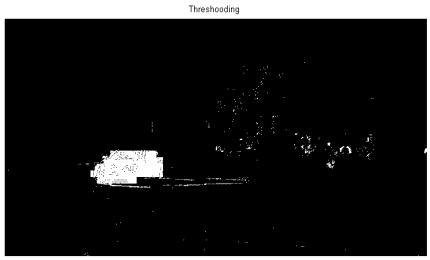
(b) Current frame

(c) Optical background subtraction



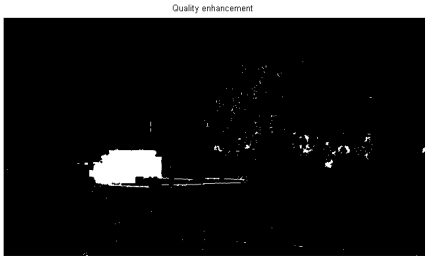
(d) The Doppler beamforming image

(e) Fused image

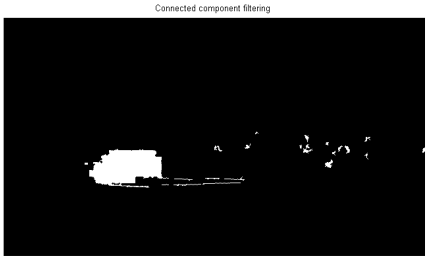


(f) Thresholding

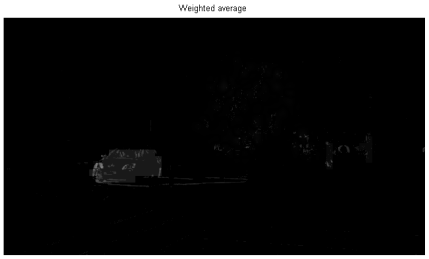
Figure 5.6: Single moving object with tree leaves and walking people (1)



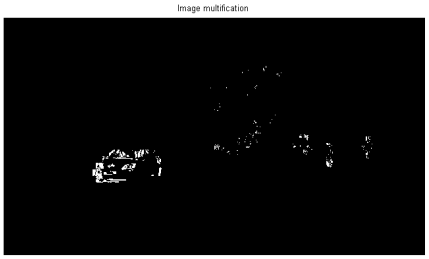
(a) Quality enhancement



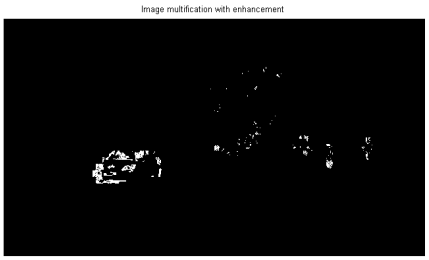
(b) Connected component filtering



(c) Weighted average



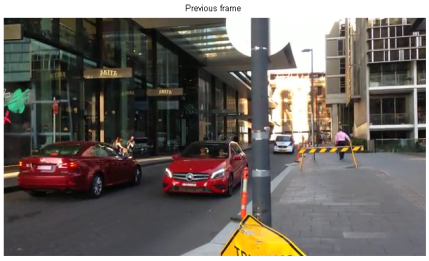
(d) Image multiplication



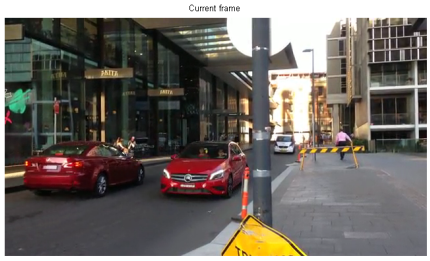
(e) Image multiplication with enhancement

(f) Subtracted object(s)

Figure 5.7: Single moving object with tree leaves and walking people (2)

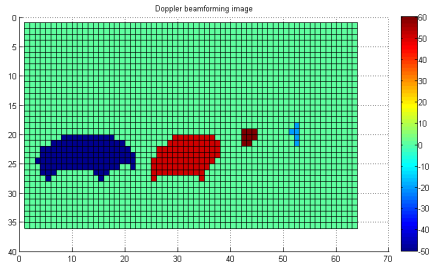


(a) Previous frame



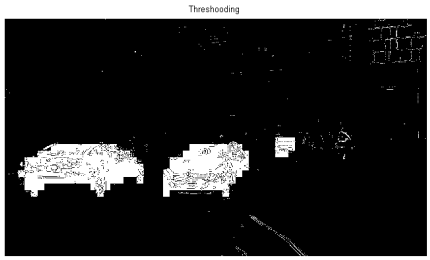
(b) Current frame

(c) Optical background subtraction



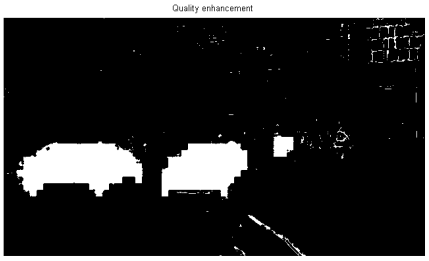
(d) The Doppler beamforming image

(e) Fused image

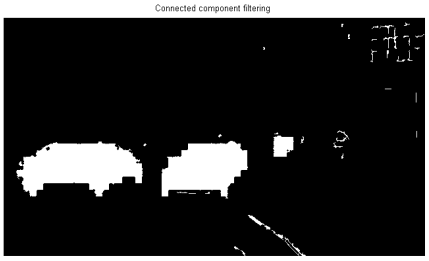


(f) Thresholding

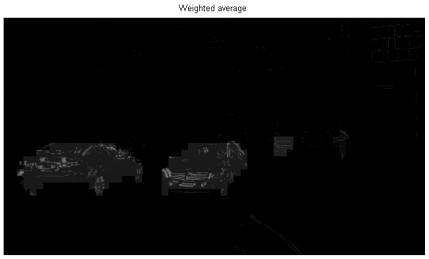
Figure 5.8: Multiple moving objects (1)



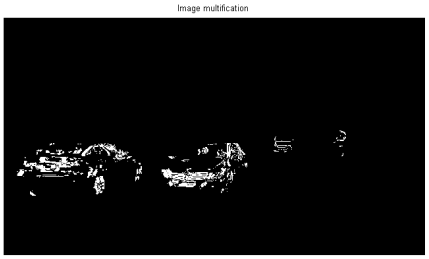
(a) Quality enhancement



(b) Connected component filtering



(c) Weighted average



(d) Image multiplication

(e) Image multiplication with enhancement

(f) Subtracted object(s)

Figure 5.9: Multiple moving objects (2)

Here three cars are clearly subtracted. In particular the car at a far distance is not clear in the optical frame differencing image, but is clear in the Doppler beamforming radar image as all the cars have similar radial velocities. The people with low velocities can still not be clearly subtracted.

5.3.7 Conclusions

Both the optical frame differencing image and the Doppler beamforming image represent the possibilities of whether there are moving objects or not. The optical frame differencing image can be affected by illumination changing, multiple small moving objects and insufficient pixel labeling in radial moving objects, while the Doppler beamforming radar image has a low resolution and can not subtract the contour of objects. In this scenario, the fusion methods on how to remove unwanted effects and improve the subtraction quality of moving objects are studied. The simulation shows that fusion can remove or reduce the ambiguities such as tree leaves, shadows and can also improve the subtraction of far distance objects with high velocities.

5.4 Scenario 2: Fusion of crossed array Doppler beamforming radar with optical image to enhance golf ball tracking

5.4.1 Introduction

Doppler beamforming radars have the ability to detect moving objects. As discussed in chapter 4, the way to increase the resolution is to increase the sensor number. The cost of increasing the sensor number in one dimension is not high. The ratio is N_2/N_1, where N_2 is the number of sensors in the high resolution radar, while N_1 is the number of sensors in the low resolution radar. As processors such as FPGAs and DSPs have the ability to process in parallel, both the computation load and hardware resource will increase according to the ratio N_2/N_1. When it comes to the two dimensional beamforming, for example a M_1 by N_1 full rectangular sensor array, if the sensor number is increased to M_2 by N_2, the ratio of computation load and hardware resource is $(M_2 \times N_2) \,/\, (M_1 \times N_1)$. According to the beamforming radar resolution computation in chapter 3, both M_2 and N_2 should be more than ten to achieve a relatively high resolution, which means the

total sensor number should be more than 100. In a low cost radar system, the cost of this number of sensors is not acceptable.

The alternative way to detect objects in two dimensions is to use two perpendicular linear arrays with M_1 horizontal sensors and N_1 vertical sensors. The total sensor number is M_1+N_1. In a crossed sensor array, the total sensor number is M_1+N_1-1 as one sensor can be shared by the horizontal and the vertical directions. When the horizontal and vertical sensor numbers are increased to M_2 and N_2, separately, the ratio is $(M_2 + N_2 - 1) / (M_1 + N_1 - 1)$. If both M_2 and N_2 are around 10, the total sensor number is around 20, which is acceptable in a low cost system. When the sensor number is increased, the increase of sensor number is the same as in the one dimension case.

The way to extend one dimensional beamforming to two dimensional crossed array is that horizontal and vertical angles are calculated separately. If the two dimensional beamforming attenuation value is multiplied in each angle, the two dimensional radar image can be created and the value in pixel $P(m_1, n_1)$ represents the possibilities of a target. This method works well if only a single object needs to be detected. However there are ambiguities if there are two or more than two objects. This is because if there are two objects in the radar detection area, both horizontal and vertical sensor arrays can sense two objects, and 4 objects will be shown in the image. Two objects in the diagonal are real objects, and the other two objects are false alarms. In a pulsed beamforming radar, the ambiguity can be removed if the range of objects can be clearly distinguished. This is because the objects with different ranges are processed in a different beamforming time series. In each time series, there is only one object in the image. In Doppler beamforming radar, the situation is similar. If the radial velocities of the two objects are different, these objects can be distinguished. However, there are some situations in which some objects are located at similar ranges and moving at similar velocities. Many methods have been tried and have not solved this problem. The method using the fusion of an optical image to help the crossed array radar to solve the ambiguity problem is discussed here.

5.4.2 Design constraints

The design should be based on low cost crossed sensor array radar.

The system should be able to work in real time.

The computation cost of both radar signal processing and optical imaging should be low.

5.4.3 Computation load and memory consumptions for sensors

According to the investigation of the relationship between the sensor number and the resolution in chapter 4, this scenario supposes that there are 20 by 20 sensors,

Table 5.2: FPGA resources

Altera Cyclone IV FPGA (EP4CE22F17C6N)	Quantity
Logic elements (LEs)	22,320
Embedded memory (Kbits)	594Kbits
18 x 18 multipliers Embedded	66
General-purpose PLLs	4
Maximum FPGA I/O pins	153
50 MHz clock oscillator	1
12-bit Analog/Digital converter (NS ADC128S022)	8 channel
SDRAM	32 MB
On-board USB blaster programming interface	1
USB mini-AB port	1
I2C EEPROM	2Kb

and the beamforming resolution in angle θ is 2.4º. The distance d from the target to the radar is 20 meters. The beamforming resolution is $d{\cdot}\theta{\cdot}\pi/180 = 0.873\,(meters)$. The beamforming resolution is different from the radar detection resolution. The radar detection resolution means if the target is smaller than the detection resolution, the target cannot be detected. The beamforming resolution is that if the target is smaller than the resolution, it can still be detected, but if the distance between two objects is smaller than the beamforming resolution, they cannot be separated in the beamforming spectrum. To separate objects with small size, the sensor number in both the horizontal and the vertical direction needs to be increased.

Take the acoustic radar as an example to calculate the computation load and hardware resource requirements of a single sensor. If the carrier frequency is 40 kHz and the sample frequency is 160 kHz, and the output of the ADC is 10bits, for one sensor, the data rate is $10bit * 160000/s = 1.6Mbit/s$. In the simulation, for FFT processing, 512 or 1024 samples should be captured each time, that is 5.12 kbit raw data. To store the FFT data, there is another 5.12 kbit data required. Therefore at least 10 kbit data space is needed for each sensor. For example, if 100 sensors are used, the number is 1Mbit. In a Nios II FPGA board, there is only 512k RAM. Other programs also consume the FPGA resources. The number of sensors greatly affects computational load and memory requirements. Table 5.2 shows the resources of the Nios II FPGA board.

5.4.4 Background

If N objects are to be detected, in a crossed sensor array radar, the radar image can show up to N^2 targets. This kind of radar is suitable for continuously tracking some specific objects which are located at some specific position or moving at some specific velocity. For example, in some ball games, the velocity of the ball is different from that of the players and other slowly moving objects. Here a case study on how to use the image to reduce ambiguity when two golf balls are flying was researched.

To create the image, first, Doppler beamforming was processed both in the horizontal and the vertical directions. The power strength in each angle was multiplied to create the Doppler beamforming image. Then the radars and the cameras were calibrated, and captured images were registered. The regions of ambiguity were marked according to the Doppler beamforming image. In each region, the image processing method such as template matching and frame differencing were used to determine the possibility whether or not there is an object.

5.4.5 Restriction of application

Crossed sensor array Doppler beamforming has ambiguities because the image is a multiplication of two line spectra. When the target number increases, the ambiguous object number increases sharply. If there are N targets, the maximum ambiguous targets is $N^2 - N$. The benefit of Doppler beamforming is that it can segment targets with different velocities, and in most applications, the number of fast moving objects to be detected at a similar velocity is no more than 3, so the ambiguous target number is fewer than 6. This is not a big computation load for the image processing.

According to the above reason, the application is limited to the number of fast moving objects with the same velocity being fewer than 3. That means the maximum region of interest (RoI) is 9. This prevents the image computation load becoming too high and ensures real time processing.

5.4.6 Doppler beamforming

As in crossed sensor arrays, the increase of the sensor number does not sharply increase the computational load and the hardware resource requirements. The radar is supposed to have 20 sensors both in the horizontal and vertical directions, therefore the resolution is 2.4^o. It is supposed that the angle of view of the camera is 120^o in the horizontal direction and 90^o in the vertical direction, and the optical image resolution is 1920 by 1080. The radar image has 50 by 38 pixels.

The maximum golf ball velocity is 100m/s. The ball which needs to be tracked usually moves at a velocity higher than 10m/s. Therefore, if the carrier frequency is

40kHz, according to the Doppler effect equation, the Doppler shift should be more than 1200 Hz. After the subtraction of the carrier, in the Doppler beamforming image, only the Doppler frequencies higher than a threshold are processed. Here only some fast flying birds or golf balls can be detected.

5.4.7 Template matching

There are two ways of template matching: the Sum of Absolute Difference (SAD) and the Normalized Cross Correlation (NCC)

5.4.7.1 Sum of absolute difference

If there are two images $I_{RoI}(x, y)$ and I_T with size M by N and P by Q, separately, the sum of absolute difference can be expressed as follows:

$$I_f(x,y) = \sum_{i=0}^{P-1}\sum_{j=0}^{Q-1} |I_{RoI}(x+i, y+i) - I_T(i,j)| \tag{5.10}$$

where, $x = 0, 1, 2, \cdots, M-1$ and $y = 0, 1, 2, \cdots, N-1$

When the SAD achieves the minimum, these two images are matched well.

5.4.7.2 Normalized cross correlation

If there are two images $I_{RoI}(x, y)$ and I_T with size M by N and P by Q, separately, the normalized cross correlation can be expressed as follows:

$$I_f(x,y) = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \frac{I_{RoI}(x+i, y+j)\, I_T(i,j)}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} I_{RoI}^2(i,j)}\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} I_T^2(i,j)}} \tag{5.11}$$

5.4.8 Frame differencing

In most cases of tracking systems such as in the ball tracking systems, the fast moving targets in the radial direction also have movements in other directions. The detail of frame differencing is shown in section 5.3. The benefit of frame differencing is that the computation cost is low, but the memory consumption is relatively high. When the size of the Region of Interest (RoI) is reduced, both the computation load and memory consumption is reduced. Even a portable micro controller can execute the algorithm in real time.

5.4.9 Decision making strategy

Ambiguity exists in the crossed array beamforming radar. As Doppler beamforming radar has the ability to separate the objects with different velocities, the maximum target number is supposed to be fewer than 3. So the question is simplified into that if there are 2 targets or 3 targets with the same velocities, how can the ambiguities be removed? If more than 4 targets are in the application, other algorithms are studied.

When there are 2 or 3 objects, the number of objects in the radar image should be 1 by 2 (no ambiguities), 1 by 3 (no ambiguities), 2 by 2, 2 by 3 and 3 by 3. Specifically, (1) If the object number is 2 by 2, 2 or 3 objects are the true targets. (2) If the object number is 2 by 3, then 3 objects are true targets. (3) If the object number is 3 by 3, 3 objects are true targets, and in each row and column, there must be one moving object.

If a template matching method is used, in each picture, the ratio of the correlation value of the highest region to the correlation value of other regions is calculated. According to the ratio, the possibilities of whether or not there are targets can be judged. One solution is setting a threshold for the ratio. Another solution is to compare the value difference in each line and each column and make a decision tree.

5.4.10 Simulation and implementation

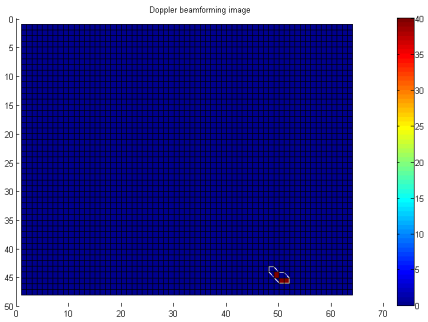
In this system, only the highest velocity needs to be processed. In both the horizontal and vertical Doppler beamforming, the highest Doppler frequency needs to be subtracted. In most cases, the highest frequency in both the horizontal and vertical should be the same or with only a slight difference. This is because the ball is moving approximately perpendicular to the radar sensor array surface plane. If the highest frequency does not match, there may be some unwanted moving objects. The second and third highest Doppler frequencies should be compared and matched.

Before the player hits the ball, the ball is not moving. So the highest Doppler frequency comes from the end of the club.

At the moment the player hits the ball, there are two fast moving objects with similar velocities. One is the end of the club and the other is the moving golf ball. In this situation, there are ambiguities in the radar image, which is shown in Figure 5.10 and Figure 5.11.

(a) The moment hitting the ball

(b) Frame differencing

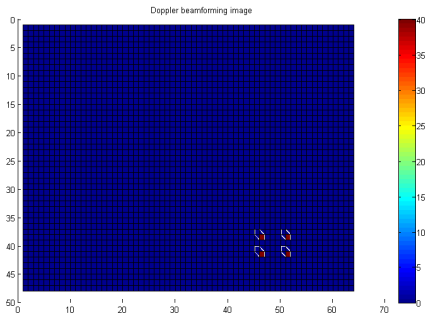


(c) The Doppler beamforming image

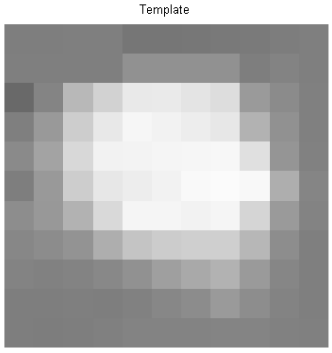
Figure 5.10: The moment when the player is hitting the ball

(a) The moment after hitting the ball

(b) Frame differencing



(c) The Doppler beamforming image



(d) Template

Figure 5.11: After hitting the ball

In this case, template matching needs to be applied to remove the ambiguities, which is shown in Figure 5.12 and Figure 5.13.

Actually, in a real golf tracking application, this kind of ambiguity does not affect the result much because the velocity of the club will reduce quickly and the ball can easily be distinguished after that. Another type of ambiguity comes when there are two players playing golf. The velocities of the balls may be the same.

The results show that frame differencing can deal with the ambiguities better than the template matching.

5.4.11 Conclusion

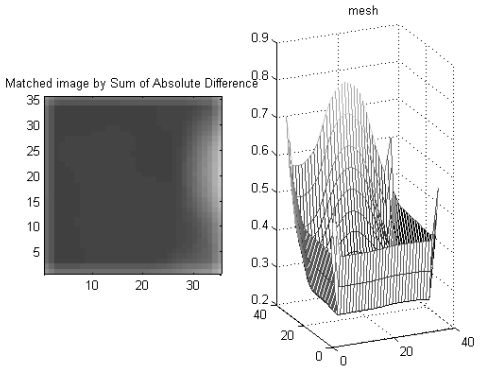
This part studies the method using imaging to reduce ambiguities of crossed array Doppler beamforming radar. The idea is that crossed array radar tracks the target first and creates a moving object image with ambiguities. Image processing methods such as template matching and frame differencing are applied in the detected area which may contain ambiguities in order to make the decision whether or not there are objects in this area. The simulation shows that images can identify the possibilities of objects. By this method the cost of Doppler beamforming radar can be reduced sharply because the data rate, computation load and memory consumption are reduced.

5.5 Scenario 3: Fusion of optical image with Doppler beamforming radar to enhance tennis ball in/out decision making

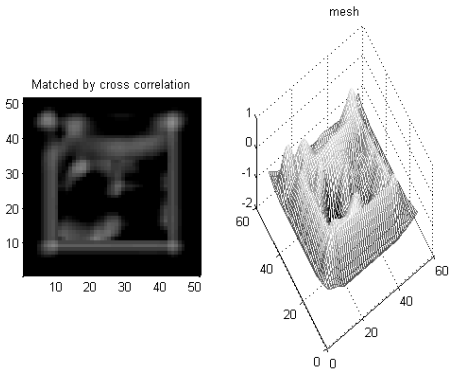
5.5.1 Introduction

Tennis and many other similar sports is fast and exciting to watch. This study aims at developing electronic tracking of the ball for judgment of in/out. Statistical

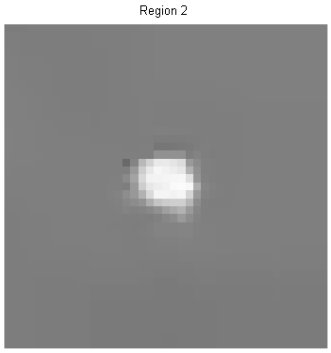
(a) Region 1



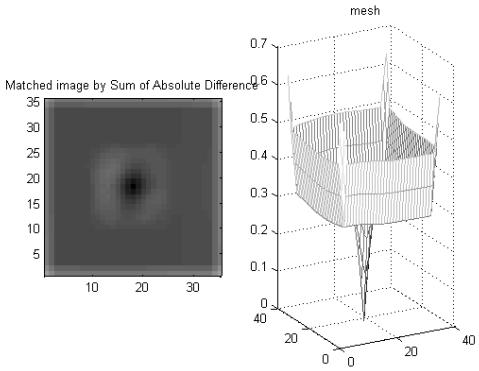
(b) Matched by Sum of Absolute Difference



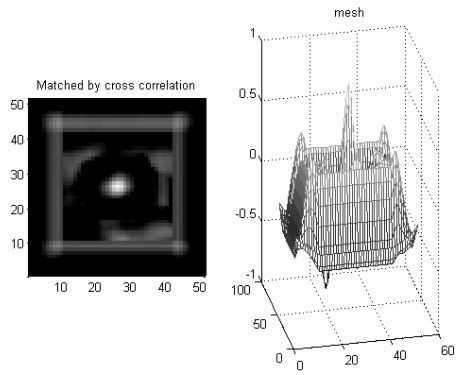
(c) Matched by Cross Correlation



(d) Region 2

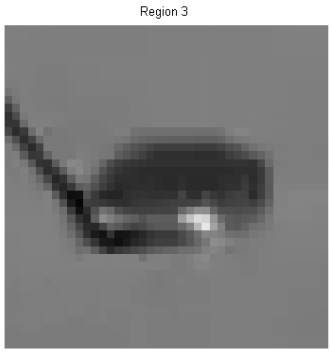


(e) Matched by Sum of Absolute Difference

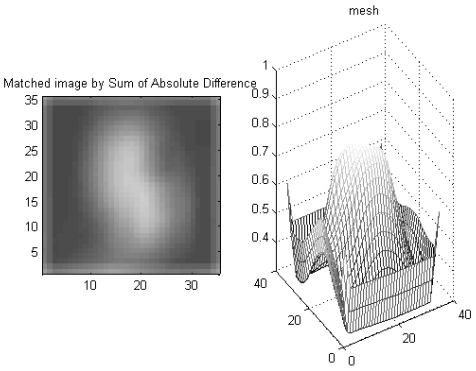


(f) Matched by Cross Correlation

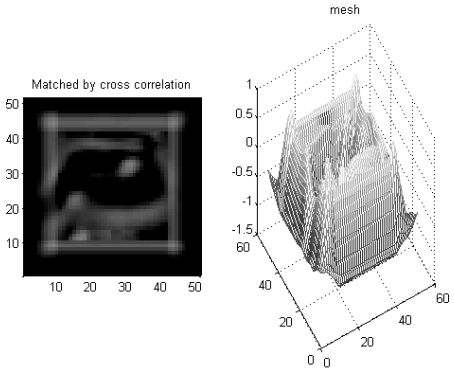
Figure 5.12: Template match



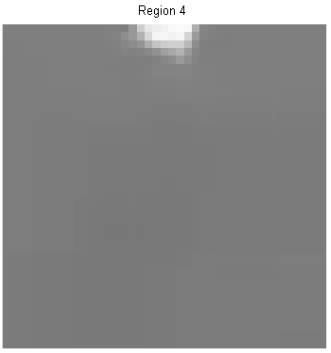
(a) Region 3



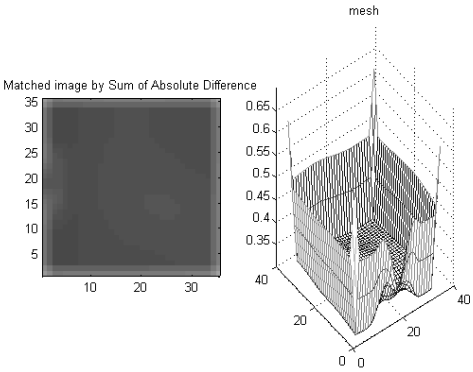
(b) Matched by Sum of Absolute Difference



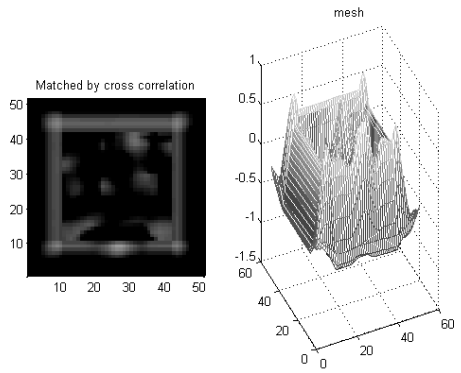
(c) Matched by Cross Correlation



(d) Region 4



(e) Matched by Sum of Absolute Difference

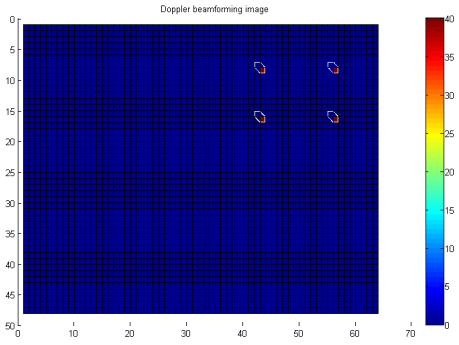


(f) Matched by Cross Correlation

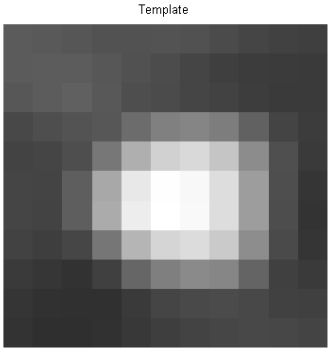
Figure 5.13: Template match 2

(a) The moment balls are flying

(b) Frame differencing

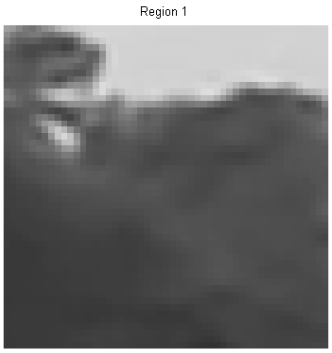


(c) The Doppler beamforming image

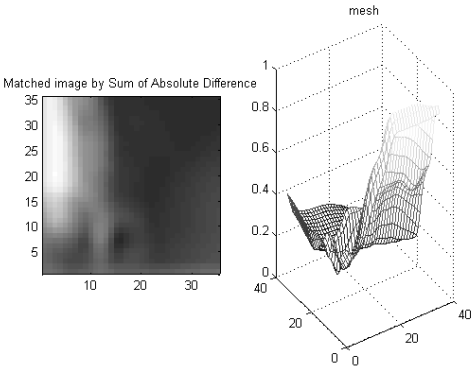


(d) Template

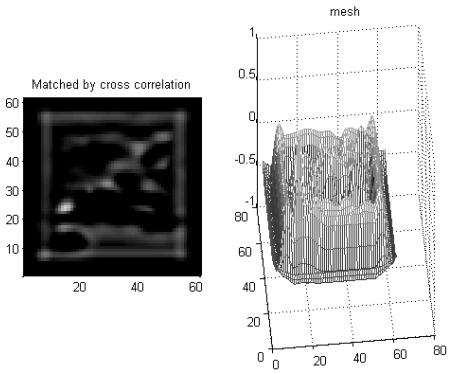
Figure 5.14: Two balls are flying



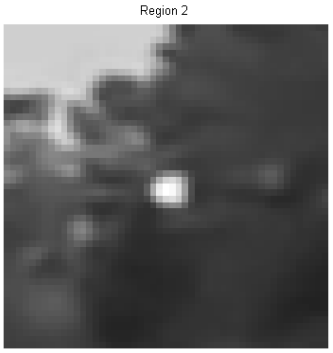
(a) Region 1



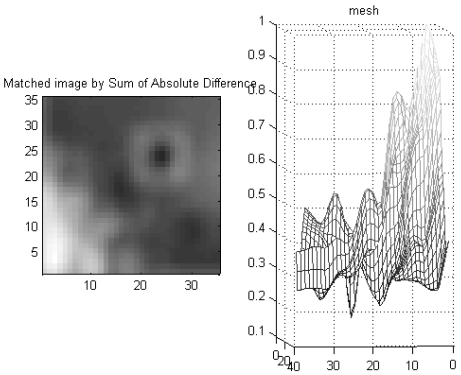
(b) Matched by Sum of Absolute Difference



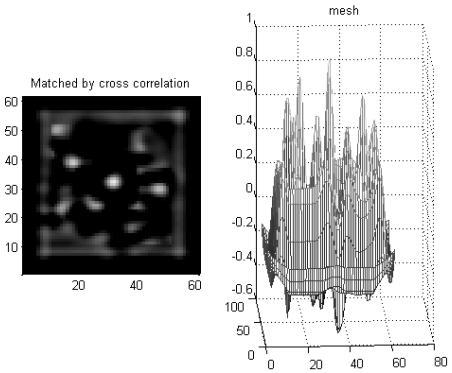
(c) Matched by Cross Correlation



(d) Region 2

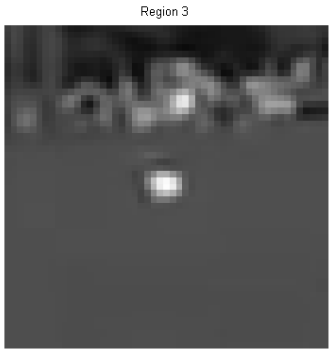


(e) Matched by Sum of Absolute Difference

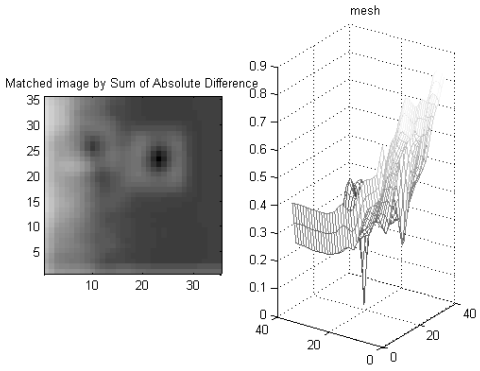


(f) Matched by Cross Correlation

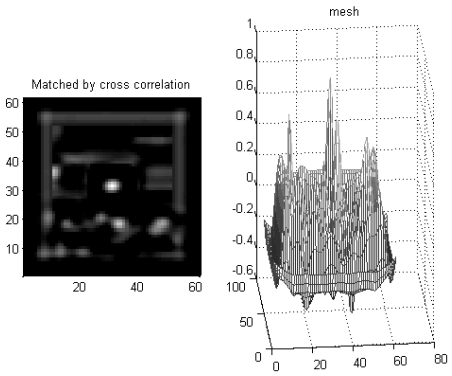
Figure 5.15: Two balls template match 1



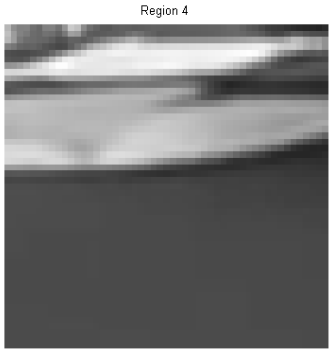
(a) Region 3



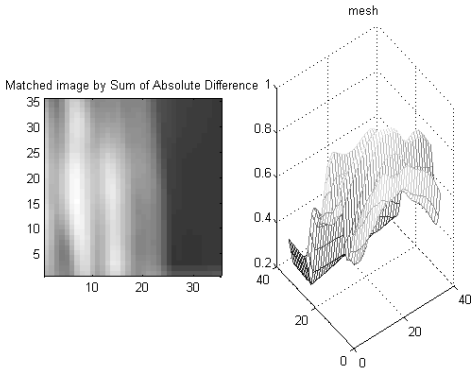
(b) Matched by Sum of Absolute Difference



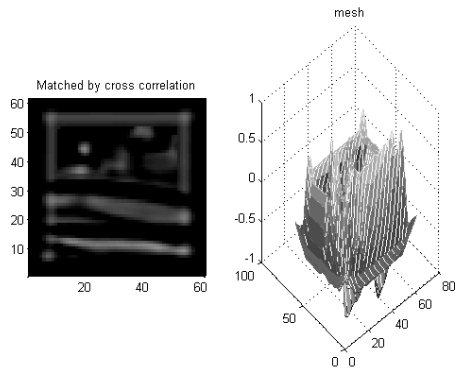
(c) Matched by Cross Correlation



(d) Region 4



(e) Matched by Sum of Absolute Difference



(f) Matched by Cross Correlation

Figure 5.16: Two balls template match 2

data for the sport will be a by-product. Also, the players can benefit from analysis of the trajectories of balls.

Extensive research has been done using computer vision to track objects, with limited success. The computing power required for rapidly moving multiple targets in the field of interest would be very large and results would be unpredictable.

Other forms of active localization have been suggested, with varying levels of success and acceptability by players and administrators who may be resistant to wearing active tags.

This scenario aims to use Doppler beamforming radar as well as optical imaging system to reduce the computational load. In detail, Doppler beamforming radar tracks the ball at a high frame rate to estimate the region of interest (RoI). Then the optical tracking system only processes the images in the area of interest. Therefore the work load of the optical system is reduced and the system avoids tracking empty scenes. If the cost reduces, this technology can be used in other sports such as volley ball, golf and cricket, as well as in surveillance systems and military systems.

5.5.2 Design constraints

- The design should be low cost with a simple structure
- The accuracy should not be reduced
- The system should be stable
- The system should use fewer cameras and have a low complexity

5.5.3 The Hawk-eye system

Hawk-eye is used to judge if a ball is in or out in tennis. This system contains several cameras located in different corners of the court. By analysis of the captured images, the system predicts the statistically most likely path of the ball, showing the results in a moving picture.

The Hawk-eye system is an expensive electronic line calling system invented by Dr Paul Hawkins. The initial goal of this system was to provide an unbiased and accurate in/out decision through a combination of electronics and a processing routine.

In general the system needs 6 to 10 cameras. In a six camera system, cameras are gen-locked into two sets of three cameras. In a ten camera system, three cameras are allocated at each end of the court and two cameras on each side of the court. These cameras are monochrome digital video cameras. All cameras should be high up in the stands and attached to something stable. Cables connect the 10 cameras

and send image data to the control system which contains 14 computers and is operated by 5 people. The system needs to be set up one week before and air conditioning is required to cool the computers.

The disadvantages of the Hawk eye system are (1) Highly expensive (2) Not 100% accurate (3) challenges the umpire's decision.

5.5.3.1 The procedure of the Hawk-eye system

1. Calibration process
    a) Intrinsic Parameters: the focal length, the pixel skew and the principal point.
    b) Extrinsic Parameters: The location and orientation of the camera.
    c) Coefficients: Radial and tangential coefficients introduced as an artifact of the camera's lens. Comparison between original and rectified image.
2. Ball detection
    a) A blob detection scheme is used to detect the ball
    b) Blob represents the group of pixels corresponding to the ball.
    c) Knowing the size and shape of the object (that is the ball) makes the detection easier.
    d) Knowing the position of the sun and position of the ball in a previous frame becomes useful for eliminating the shadow of the ball during detection.
3. Geometric Algorithm
    a) Triangulation Process: Determines accurately the position of the ball in 3D.
    b) Two sub-processes: 2D position calculation and depth calculation
4. Triangulation
    a) Determining the 3D position of the ball
    b) The cameras are mounted at ground level, positioned with their vision parallel to the ground.

5.5.4 Background

The biggest challenge of the Hawk-eye system is high cost and high computation load. The system needs both high frame rate and high resolution. The velocity of the ball is usually too high for the human eye to follow. The camera is usually fixed because it has not been possible to design a mechanical part to help the camera to follow the ball and adjust the focal length; therefore the resolution of the camera should be high. The camera could be mono color or RGB color. A typical high speed high resolution camera would have 1920 * 1080 resolution and 120Hz frame rate. If this camera is mono color, the data rate is 8bit*1920*1080*120=1990656000; that is nearly 2Gbit per second. If this camera uses RGB colors, the data rate should be nearly 6Gbit per second. Obviously even some simple image processing methods are impossible to be applied in this system.

As described in chapter 4, Doppler beamforming radar has the ability to sense only the moving object, so the fusion with Doppler beamforming may be able to greatly relieve the computing overhead required.

5.5.5 System implementation

5.5.5.1 Image registration

The way to register the radar image and the optical image is to find some control points which can both be clearly identified by the Doppler beamforming radar image and the optical image. It is hard to find such points. A way to register them is to put some points in the field of view of both the Doppler radar and the camera. The control points emit light for cameras to identify them and have small fans for Doppler beamforming radar to identify them. For a tennis court, four control points are enough. One allocation of these devices in a tennis court is shown in Figure 5.17. In practical applications, the camera/radar should be 2 to 3 meters outside the court area.

5.5.5.2 Kalman filtering

The ball tracking system is a linear system, so the linear Kalman filter can be used.

The motion model in the x direction is

$$x_{t+1} = x_t + v_t t + \frac{1}{2} a t^2 \tag{5.12}$$

$$v_{t+1} = v_t + at \tag{5.13}$$

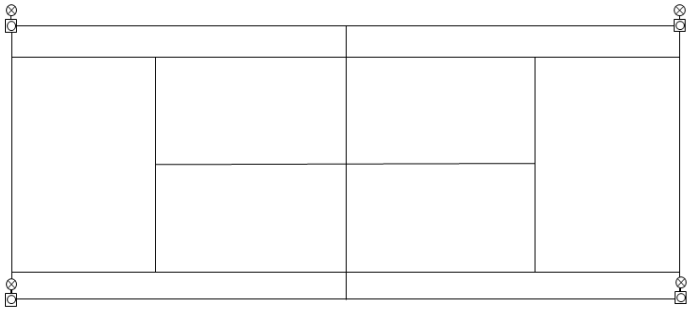


Figure 5.17: Control points in tennis courts

The motion model in the y direction is

$$y_{t+1} = y_t + v'_t t + \frac{1}{2} a t^2 \tag{5.14}$$

$$v'_{t+1} = v'_t + a' t \tag{5.15}$$

These two direction can be combined and written in other format:

$$\bar{x}_t = F x_{t-1} + G\mu_t + E_x \tag{5.16}$$

$$\bar{z}_t = C\bar{x}_t + E_\xi \tag{5.17}$$

where, $x = \begin{bmatrix} x & \dot{x} & \ddot{x} & y & \dot{y} & \ddot{y} \end{bmatrix}^T$, $F = \begin{bmatrix} 1 & T & T^2/2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & T^2/2 \\ 0 & 0 & 0 & 0 & 1 & T \end{bmatrix}$,

$E_x = \begin{bmatrix} \frac{T^4}{4} & 0 & \frac{T^3}{2} & 0 \\ 0 & \frac{T^4}{4} & 0 & \frac{T^3}{2} \\ \frac{T^3}{2} & 0 & T^2 & 0 \\ 0 & \frac{T^3}{2} & 0 & T^2 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $E_\xi = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$.

5.5.6 Region of interest

Doppler beamforming radar continuously measures and estimates the position of the tennis ball.

In detail, the radar needs to verify if a target is in the trajectory. This is to reduce processing time and keep the system working in real time. If a target is on the path of the tracked ball, the region of interest (RoI) is marked according to the estimated object position. The RoI is usually a rectangular area. The imaging searches for the ball inside the RoI to verify the position. This tracking of the RoI can be improved by Kalman filters.

A new object can be created by the radar and the imaging can be done on an existing object only. Thus, an RoI is defined around the current estimate in the image space.

5.5.7 Image processing sub-system

Basically, the image processing sub-system contains these procedures:

- Image registration: minimize the effects of camera movement;
- Line identification: determine the inner and outer edge of the lines;
- Trajectory identification: subtract the trajectory of the ball;
- Decision making: identify the location of the ball's bounce point.

The processing contains these sub sections which are shown in Figure 5.18.

5.5.7.1 Line identification

Image Preparation The first step is to remove the unwanted artifacts, which includes any foreign markings on the court and anything else that may cause pixel noise.

Thresholding converts a grey scale image to a binary image based on a set of thresholds. The optimal threshold can divide the edge of the object from the background. The value varies according to the application environment.

After the thresholding, there is still much noise, such as unwanted small objects and small gaps or holes. So these two methods are used to remove the noise.

- Unwanted small objects: an object removal method is applied on the binary picture. The object to be tracked has a minimum size, the region smaller than the minimum size should be filtered.

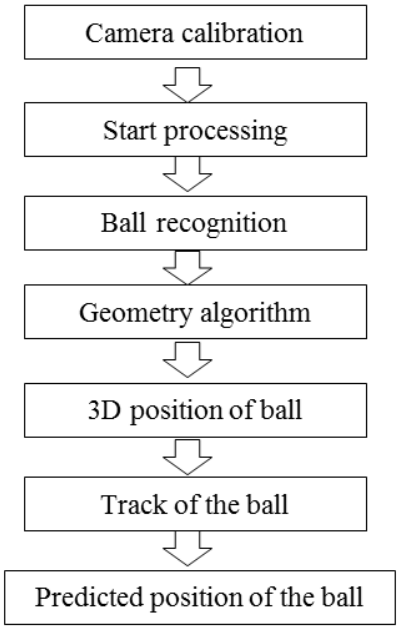


Figure 5.18: Process of the ball tracking system

- Small gaps or holes: for each pixel of a structural element of a chosen shape and size. The unconnected pixels need to be bridged. This method can remove the small gaps and holes.

Line extraction Line extraction is used to determine the locations of the inner and outer edges of each tennis court line.

The first step is to detect the edge. Canny edge detection is a method to detect the line edges. This method works by using two filters to detect both strong and weak edges. Weak edges are much harder than strong edges to be detected due to factors such as faintness.

The idea of edge detection is for the gradient of the image to be calculated. The gradient is calculated by using the derivative of a Gaussian filter. The two filters then scan the image looking for the local maxima of the gradient.

The next step is that the pixels that make up each edge need to be grouped, and hence the edges need to be defined. The edge linking and segment fitting function can be used. Kovesi's functions link edge pixels together into lists with one list created from each edge contour. Each edge list is then fitted with a line segment.

Line equation computation To simplify the computation, a standard line equation is produced for each contour using a linear equation:

$$y = ax + b \tag{5.18}$$

Each line is represented by two points.

5.5.7.2 Trajectory identification

The purpose of this section is to calculate the equations for the path that the ball has traveled. One is for the pre-bounce trajectory while the other is for the post-bounce trajectory.

Image differencing As described before, one way to subtract the moving object is image registration and image differencing. The input image should be two dimensional grey scale.

Image preparation The pre- and post- bounce tube images need to be first cleared. The thresholding can remove a majority of the unwanted artifacts in the image. The threshold value is computed using the Otsu's method. The Otsu's method is a threshold technique that works by assuming that every image has two classes of pixels. The two classes are foreground pixels and background pixels. This thresholding method computes a histogram of the input image and then selects a threshold based on that histogram.

Trajectory equation computation The erosion method is used to erode the tubes into a line which only contains the center pixels. Then the coordinates of the pixels left over after erosion need to be obtained by searching the binary image for all white pixels. A best-fit function is used to calculate the trajectory. Usually, for a moving ball, the trajectory function is second order and non-linear:

$$y = ax^2 + bx + c \tag{5.19}$$

5.5.7.3 Decision making

Bounce point calculation The point at which the ball has made contact with the court needs to calculated. One method is by calculating the incoming and outgoing trajectory; the cross-over point is the bounce point location.

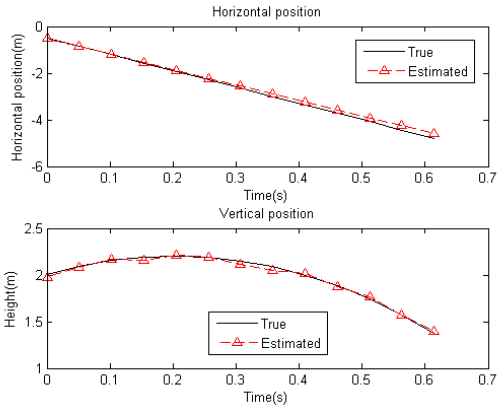
Closest line measurement The tennis court line that is closest to the bounce point of the ball needs to be found. Once the collection of the white pixels has been completed, the next step is to calculate the distance from the center of the bounce point to the nearest white pixel. This calculation computes the shortest distance from the center of the bounce point to the closest tennis court line.

Decision engine The decision engine is used to judge if the ball is in or out:

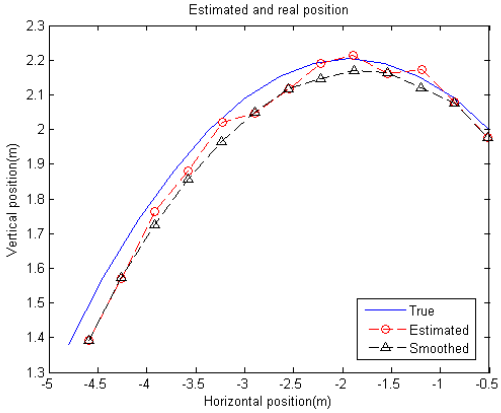
- In decision: when the ball is either inside the boundary lines, or is touching a boundary line.
- Out decision: when the ball is outside the boundary lines.

5.5.8 Simulation and implementation

Figure 5.19 shows the Kalman filter tracking of the RoI.



(a) Trajectories in two directions

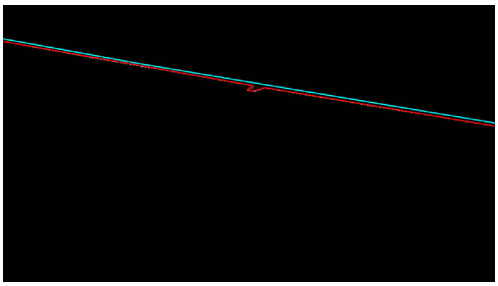


(b) Estimated and real position

Figure 5.19: Kalman filtering

Figure 5.20: Region of interest

(a) Sample Region of Interest



(b) Detected line edge

Figure 5.21: Sample RoI and the detected line edges

In Figure 5.20, according to the radar image, the estimated trajectory is projected into the optical image, then the RoI can be marked.

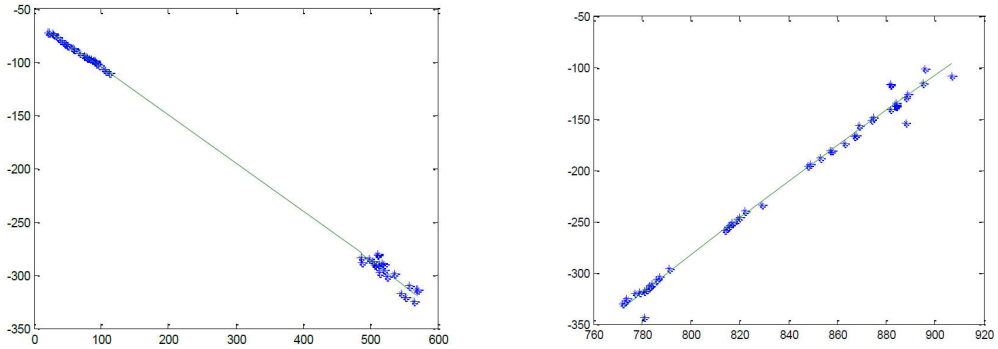
After the Region of interest is detected, when the RoI is approaching the ground, the image processing system needs to find the trajectory and make a decision. Figure 5.21 shows the sample of RoI and the detected line edges.

There are two lines to segment the region. The equations which represent these two lines are

$$y = (-0.1667)\,x + (-229.1667) \tag{5.20}$$

$$y = (-0.1825)\,x + (-131.1825) \tag{5.21}$$

Then the trajectory of the ball needs to be identified.



(a) Pre-bounce trajectory **(b)** Post-bounce trajectory

Figure 5.22: Pre-bounce and post-bounce trajectory

The equations of these two lines are

$$y = (0.0009539)\,x^2 + (-0.4972)\,x + (-62.2122) \tag{5.22}$$
$$y = (-0.000842)\,x^2 + (3.1737)\,x + (-2282.4667) \tag{5.23}$$

Finally the decision should be made by the estimated ball position.

The final position to be detected is $(735.9601, -402.7266)$

5.5.9 Further work and conclusions

Recommendations for further work It is recommended that further investigation takes place in regard to applying machine learning to automating the in/out decision process. Further, the allocation of cameras should be finalized. Static location of cameras will introduce a level of uniformity to the images that are captured and consequently processed by the image processing sub-system. Uniformity simplifies the processing aspect of the overall system. Third, further testing should be carried out with images where on-court illumination varies. That is, testing with images that have been captured on a court lit only by daylight, an indoor court lit only by indoor lighting and an outdoor court lit by court lighting and any ambient light.

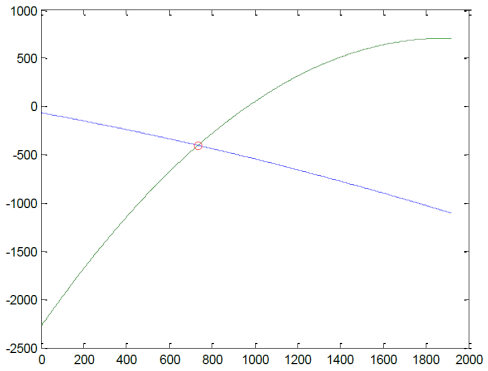


Figure 5.23: The predicted position

Conclusions The simulation shows that it is possible to effectively use imaging principles and techniques to determine the location at which the ball has landed on the tennis court's surface.

5.6 Scenario 4: fusion of optical image and beamforming radar to track target in 3D position

5.6.1 Introduction

Both the radar and optical imaging have been used to track a target separately. The information provided by the radar image and the optical image is complementary. Radars measure the range well but measure the AoA with a low resolution. Optical images measure the AoA well but cannot measure the range. Instead of the fusing of these two types of sensors at pixel level, in this scenario, these two types of sensors work separately and the kinematic information can be fused to reduce the uncertainty.

The goal of Kalman filtering is to estimate state variables of a system from incomplete noisy measurements. The Kalman fusion is used to fuse data from noisy sensors to improve the estimation of the present value of the state of variables of a system.

5.6.2 Design constraints

The system should be low cost.

The system should be able to track in real time.

The accuracy of the fusion system should be improved and the noise should be reduced.

5.6.3 Background

In a tracking system, there are all types of noise which affect the accuracy of the tracking. The uncertainty of the range and position can be reduced by correctly fusing several types of sensors. A Kalman filter and an extend Kalman filter can be used in linear systems and non-linear systems, respectively.

In a Kalman filter, it is supposed that the real state k in the time series is derived from the state in time $k-1$:

$$x(k+1) = Fx(k) + Bu(k) + G\omega(k) \tag{5.24}$$

where F is the transform model in state $x(k)$, B is the input control model of control vector $u(k)$, and $\omega(k)$ is the process noise with zero mean and covariance Q_k.

The measurement of $x(k)$ in time k is

$$z(k) = Hx(k) + \upsilon(k) \tag{5.25}$$

where H is the observation model, $\upsilon(k)$ is the measurement noise with zero mean and covariance R_k.

According to the stage difference, the Kalman filtering fusion of two sensors can be divided into 2 methods: the state-vector and the measurement-level fusion. In the state-vector fusion, state vectors of the two methods are different, and a weighted sum of the independent state estimates is used. In the measurement fusion, the two sensors share the same state vector and use their own measurement equations. The differences are described below.

5.6.3.1 Fusion methods

State vector fusion In the state vector fusion, each sensor is processed separately. The sensor estimates the state vector separately and the state vectors are fused by their associated covariance matrices.

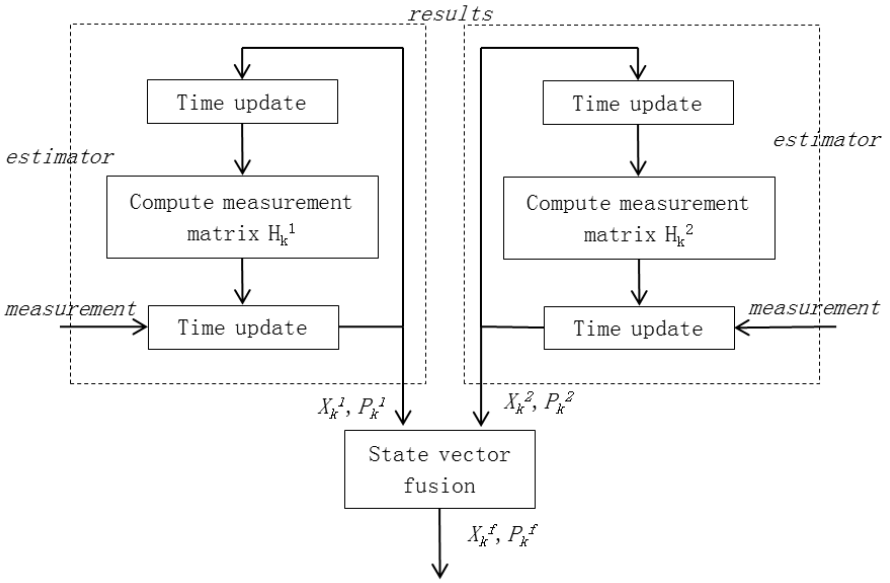


Figure 5.24: State vector fusion

$$\hat{X}^f = P^2\left(P^1+P^2\right)^{-1}x^1 + P^1\left(P^1+P^2\right)^{-1}x^2 \tag{5.26}$$

$$\hat{P}^f = \hat{P}^1 - \hat{P}^1\left(\hat{P}^1+\hat{P}^2\right)^{-1}\hat{P}^{1T} \tag{5.27}$$

The procedure of the state vector fusion is shown in Figure 5.24.

Measurement fusion In the measurement fusion, all the measurements are combined and the estimation procedure is the same. The fusion is only in the measurement equation

$$z_k = H_k x_k + v_k \tag{5.28}$$

If there is one sensor, $H = \begin{bmatrix} 1 & 0 \end{bmatrix}$, covariance is R, and if there are two sensors, $H = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, covariance is $\begin{bmatrix} R1 & 0 \\ 0 & R2 \end{bmatrix}$.
The procedure of the measurement fusion is shown in Figure 5.25.

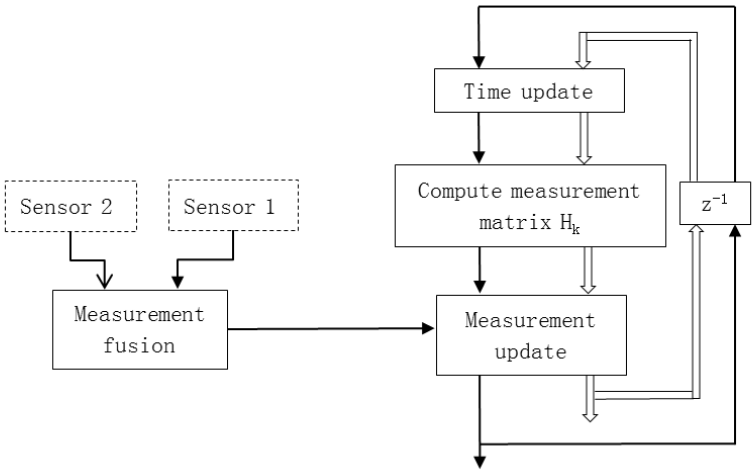


Figure 5.25: Measurement fusion

5.6.4 Optical image and radar image model

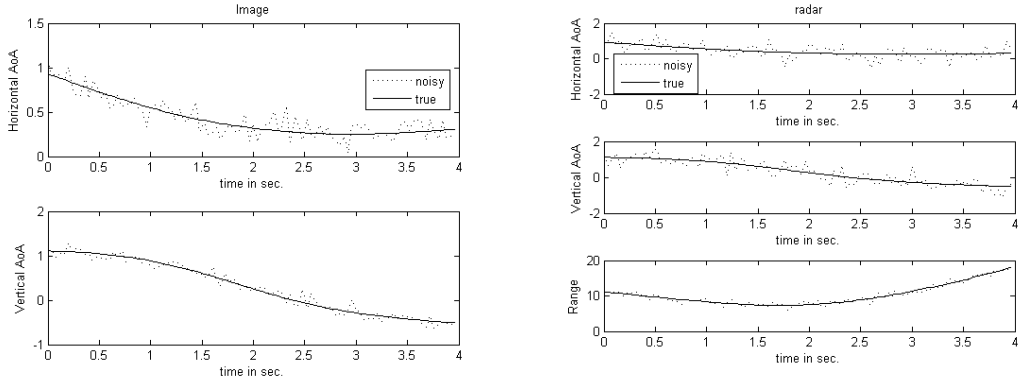
The optical images measure the horizontal angle α and vertical angle β of the target, while radars measure the horizontal angle α and vertical angle β, as well as the range r of the target. In a Doppler beamforming image, the radial velocity v can also be measured.

$$z_o(k) = \begin{bmatrix} \alpha & \beta \end{bmatrix}^T \tag{5.29}$$

$$z_r(k) = \begin{bmatrix} \alpha & \beta & r \end{bmatrix}^T \tag{5.30}$$

The predicted state of the targets is $\begin{bmatrix} \tilde{x} & \dot{\tilde{x}} & \ddot{\tilde{x}} & \tilde{y} & \dot{\tilde{y}} & \ddot{\tilde{y}} & \tilde{z} & \dot{\tilde{z}} & \ddot{\tilde{z}} \end{bmatrix}$.

According to the coordinate transform in chapter 3, the relationship between the camera and radar measurement and the position in the Cartesian coordinate system is as follows:



(a) Measurement from the optical image

(b) Measurement from the radar

Figure 5.26: Measurements from image and Radar

$$\tilde{\alpha} = \arctan \frac{x}{z} \tag{5.31}$$

$$\tilde{\beta} = \arctan \frac{y}{z} \tag{5.32}$$

$$\tilde{r} = \sqrt{x^2 + y^2 + z^2} \tag{5.33}$$

5.6.5 Simulation and implementation

The simulation uses the position, velocity, and acceleration components in each of the three Cartesian coordinates x, y, and z. The simulation is based on the simulation from [20]. The sample frame rate is set to be 25 f/s. The horizontal and vertical AoA noise of the optical image is 0.01. The horizontal and vertical noise of the radar is 0.1, and the noise of the radar range is 0.5 meters. The number of Monte Carlo runs is 50. The initial state is $\left[\begin{array}{ccccccccc} x & \dot{x} & \ddot{x} & y & \dot{y} & \ddot{y} & z & \dot{z} & \ddot{z} \end{array}\right] = \left[\begin{array}{ccccccccc} 3 & 1 & 1 & 4 & -2 & 2 & 10 & -3 & -1 \end{array}\right]$.

The measurements of the optical image and the radar image are shown in Figure 5.26.

The root mean square position error (RMSPE), root mean square velocity error (RMSVE) and root mean square acceleration error (RMSAE) are shown in Table 5.3

Table 5.3: Comparison of state vector fusion with measurement fusion

	State vector fusion	Measurement fusion
RMSPE	0.2616	0.2591
RMSVE	0.5018	0.4844
RMSAE	0.6466	0.6400

The measurement fusion is a bit better than the state vector fusion according to the simulation.

5.6.6 Conclusion

According to the stage of the fusion, there are two methods to use a Kalman filter to fuse the optical imaging and radar imaging. The procedures of both state vector fusion and measurement fusion are provided. The RMSPE, RMSVE and RMSAE of the simulation using these two methods are provided.

5.7 Scenario 5: Doppler beamforming radar to improve gait analysis

5.7.1 Introduction

Gait recognition, which captures the manner of human walking, has been considered to be the biometric method with the most potential in the area of intelligent surveillance at a distance. Existing gait recognition methods usually use a single sensor such as a camera, a wearable sensor and a radar. Much research has been done and some results have been achieved. Some research found that humans can be identified by the Doppler effect produced by different parts of the human body.

By investigating the performance of each gait recognition method, it has been found that the characteristics of optical imaging and Doppler radar in gait analysis are complementary. For example, the optical imaging senses better in lateral distance while the radar senses better in radial distance. The radar has low resolution but can work in the dark while the optical imaging has higher resolution but is highly dependent on illumination. Therefore, it is proposed that by combining gait feature extraction from the radar with Doppler beamforming, not only the feature can be extracted, but also the position of the feature can be detected. The position information can facilitate the fusion of the radar image with the optical image.

Table 5.4: Real possible affecting factors [8]

Typical assumptions	Real possible affecting factors
No camera motion	Changes of the clothing style
Only one person in the view field	Distance between the camera and the walker
No occlusion	Background or environment
No carried objects	Carried objects such as briefcase
Normal walking motion	Abnormal walking style
Moving on a flat ground plane	Variations in the camera viewing angles
Constrained walking path	Walking surface such as flat ground, grass or slope
Plain background or environment	Mood
Specific viewing angle	Walking speed
Data recorded over one time span	Change with time

5.7.1.1 Difficulties in existing gait analysis methods

- Detection in complicated environments. Most of the prior research has been done in simplified and controlled situations: static cameras, single person walking, simple and static background, no weather and light changing. Although performance is reported to be good, most of the work has been done on small or non-realistic databases. In a real environment, there is much more interference; many moving objects exist and the objects are very difficult to extract. Table 5.4 shows the real possible affecting factors summarized by Nixon [8]:

- Computation is very high. Real-time processing is important to a surveillance application. By now high performance optical imaging based detection methods require more computation. They need to analyze each picture, extract features and classify the features in real time.
- High cost of the equipment.

5.7.2 Design constraints

- The system should be low cost.
- The system should be reliable.

5.7.3 Micro Doppler characteristics of human movements

All the swinging parts of people like arms and legs can produce Doppler effects. Many models of the human body are used in the experiments. These models divide the human body into several rigid parts.

The torso can represent the movement of the whole person. The movement of other parts can be considered to be relative swinging movements around the body. The lower hand movements and lower leg movement can be considered as swinging movements with respect to the elbow joints and knee joints in a time varying function, separately. The foot movements can be considered as swinging movements with respect to the ankles in a time varying function.

The short time Fourier transform (STFT) is used to subtract the movements. The window function needs to be carefully chosen because if the window is too narrow, the length of samples for Doppler beamforming processing is not sufficient, and similar frequencies cannot be separated. Otherwise, if the window is too wide, the frame rate of the Doppler beamforming image is too low.

5.7.4 Doppler beamforming

Doppler beamforming is described in chapter 4. Different frequencies should be separated and the Doppler beamforming is processed based on different frequencies.

5.7.5 Simulation and implementation

The Doppler processing of the data from one sensor can detect all the Doppler frequencies but cannot determine the position.

Figure 5.27 shows STFT of the signals from one sensor reflected by a walking person.

Figure 5.28 demonstrates an example of how to use Doppler beamforming images to locate the position of the movements of the human body. If the image is fused with the optical image, the movement characteristics of the human body can be detected.

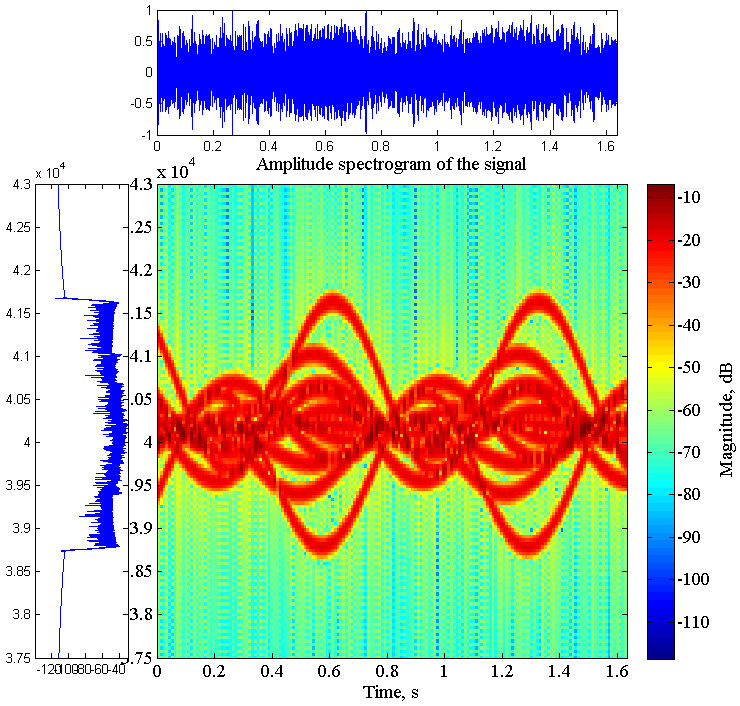


Figure 5.27: Short-time-fourier-transform spectrum of a human reflection

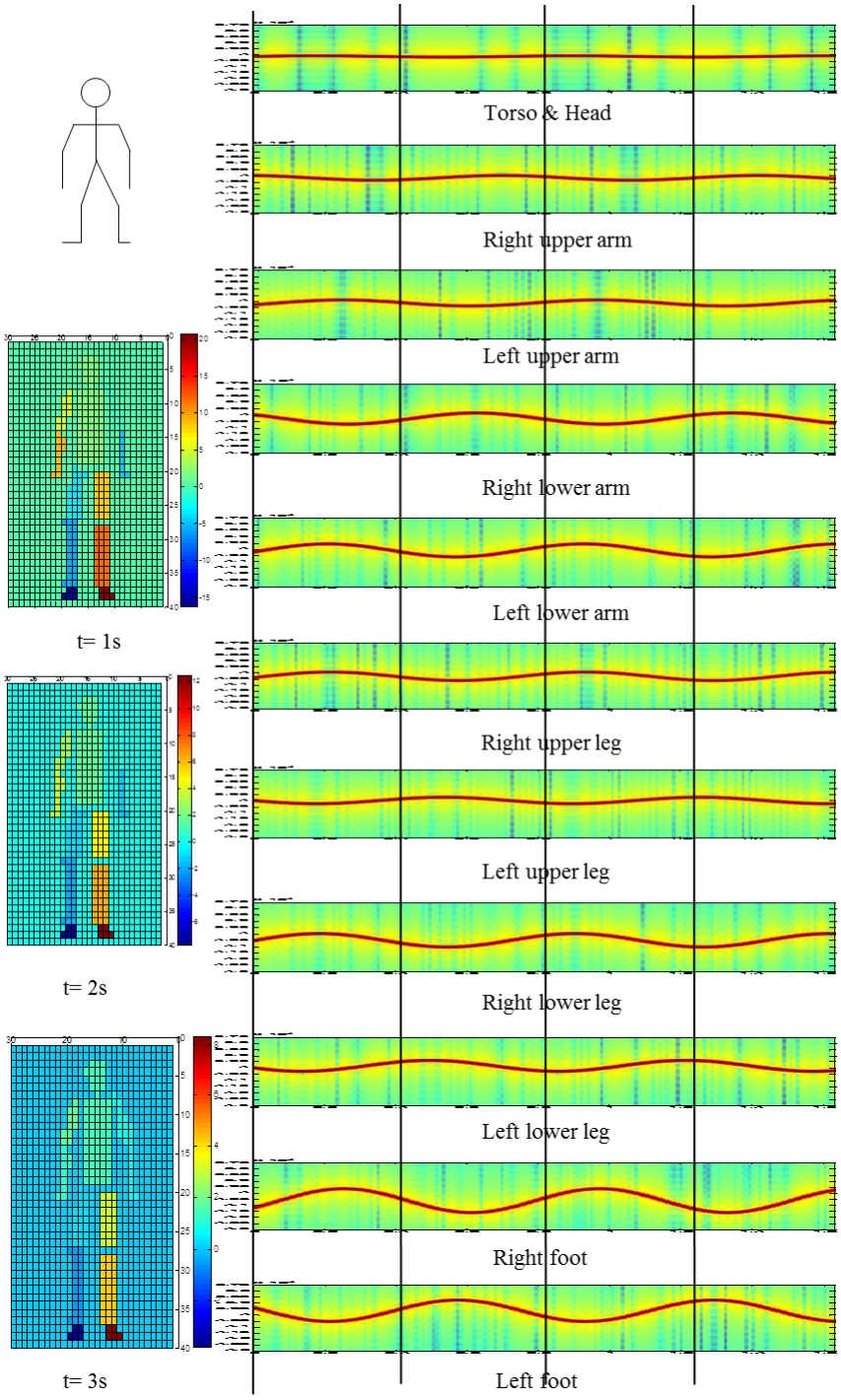


Figure 5.28: Spatial-time analysis

5.7.6 Conclusion

The simulation shows that not only the frequency changing in the Doppler signals can be known, but also the position of the moving objects can be known. The position is useful for the fusion of the Doppler gait features with the optical imaging gait features.

6 Hardware construction and experiments

6.1 Introduction

The mathematical explanation and the simulation in previous chapters have demonstrated the creation of the pulsed beamforming and Doppler beamforming image, and the fusion algorithms about these new image layers with traditional optical image RGB layers. All the assumptions are based on good design of the hardware system and the results are highly affected by the hardware design and capability.

The system structure is described in this chapter. As the optical images can be captured by a low cost high quality web-cam, most of the hardware job is to design a radar system which can capture the data required. The radar system mainly contains the sensor board, the controller board and a PC.

As demonstrated before, the computation load and memory consumption increase quickly with the increasing of the sensor number. FPGA has the benefits in parallel processing and is suitable for the controller board design. In the first design, normal size of the transducers were adopted. As for the sensor spacing consideration, the MEMS transducers with very small sizes were employed. If the MEMS chips contain ADCs inside the packages, the structure of the Printed Circuit Board (PCB) can be much simpler. Like image CCD sensors, some specified transmitter and receiver sensor array boards can be manufactured.

Though the hardware cannot be extended to a large sensor array, which can create a high-resolution image, the experiment has shown that the hardware can work. The experimental facility is limited by costs. Given the financial resources, development of a FPGA based computational solution could take a relatively short time.

6.2 Hardware implementation

6.2.1 System structure

A web-cam and a pulsed Doppler beamforming radar were allocated side by side. The camera provided light RGB values and pulsed Doppler beamforming radar

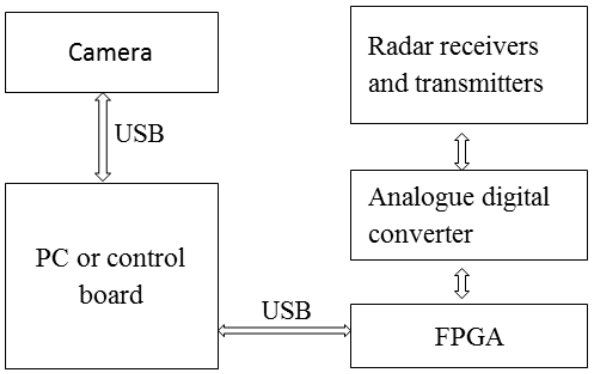


Figure 6.1: System structure the proposed system

provided the range and the radial velocity of the target. A PC and a FPGA board were used to control the system and process the data.

An electromagnetic wave is able to work over long distance. The scan rate can be high due to the high propagation speed. However, the computational load is also high and the system structure is complicated. In this case, an acoustic radar was preferred because:

- For the vision range detection, the detection range of acoustic radar is long enough compared to cameras.
- The structure is simple, therefore the system is robust and low in cost. According to the Nyquist sampling equation, $f_s \geq 2f$; therefore, for research purposes, all the signals can be sampled and recovered without pre-processing, and all the raw data can be stored on a PC, which is convenient for research.
- The cost of ADCs increase sharply as the sampling rate increases.

The system structure is shown in Figure 6.1.

The sensor number of the array is M by N , and the spacing between two adjacent sensors is 0.00425 meters. For a 12 by 12 sensor array, the array size is about 6 cm by 6 cm, and even with other circuits, the radar board is very portable. The camera used is a Logitech C920 HD Pro Webcam, which is a portable webcam with a high image quality. The FPGA board is the Altera DE1 board, with size

Table 6.1: Cost estimation of each part of the system

Part name	Estimated cost
FPGA Board	150 AUD for development Kits, 60 AUD for the core module
High quality web-cam	50 AUD
Micro-CPU control board, such as Raspberry PI	40 AUD
Pulsed Doppler beamforming radar PCB board manufacturing	30 AUD (the more, the cheaper)
Components such as transducer and chips for the radar	30 AUD
Accessories such as SD card, powers and connects for FPGA and Micro-CPU control board	30 AUD
Assembling box	20 AUD
Total estimated cost	350 AUD

around 10 cm by 10 cm. This design is for research to verify the propositions, and for the purpose of building an industrial product; the FPGA board and the micro controller can be integrated into a sensor board.

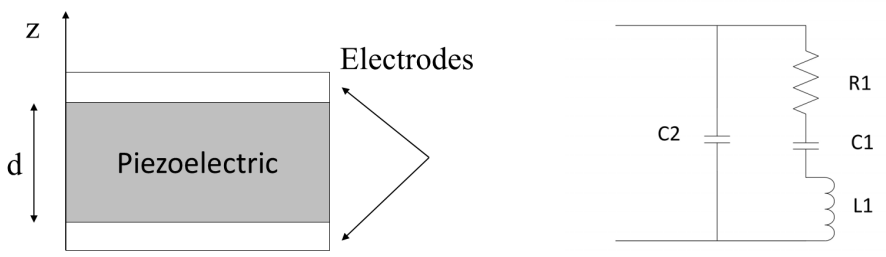
In the research stage, a PC was used to collect and process the data. In the future, a micro controller can be adopted to control the whole system as most of the computation is carried by a FPGA. In this stage, the FPGA is just used to control the ADCs, to generate transmitted signals and to transmit the captured data to the PC.

The cost of each part is estimated in Table 6.1. Generally, the cost of the vision-radar system is very low compared with a traditional radar.

6.2.2 Ultrasonic transducer

There are two types of ultrasonic transducers available: Magnetostrictive and Piezoelectric.

Piezoelectric transducers use the piezoelectric effect that certain materials possess to convert electrical energy into mechanical energy. Magnetostrictive transducers work on the magnetostrictive effect and convert electrical energy into a magnetic field and then into mechanical energy. Piezoelectric transducers are now the most common transducers used to create ultrasonic signals. Both open and closed design transducers are available in the market. Piezoelectric transducers convert electrical energy into acoustic energy and vice versa. The active elements used in most



(a) Piezoelectric transducer **(b)** The equivalent circuit

Figure 6.2: Piezoelectric equivalent circuits [4]

piezoelectric transducers are the piezoelectric ceramic material, with the thickness of the wafer material determining the frequency emitted by the transducer.

The electrical behavior of ultrasonic ceramic piezoelectric transducers may be represented by an equivalent circuit made up of inductors, capacitors and resistors and may be represented by the simplified equivalent circuit in Figure 6.2.

According to the requirement of this project, 400SR120 from Midas Components Ltd was chosen as the receiver and 400PT120 was chosen as the transmitter in the prototype. They are transducer pairs with aluminum housing. However the size of the transducer was too big compared with the required sensor spacing. Then the MEMS transducer SPM0404UD5 was adopted as the receiver with size 4.16mm by 3.10mm. As this is a surface mount chip, the design of the PCB is also simplified.

6.3 Controller and programing

6.3.1 Altera DE1 board and Cyclone II FPGA chips

FPGA is the acronym for Field-Programmable Gate Array. For cost and performance balance, Altera DE1 with Cyclone II FPGA chip: EP2C20F484C7N was chosen. The board is built in with 512-kbyte SRAM and 8-Mbyte SDRAM and the resources are shown in Figure 6.3, in detail.

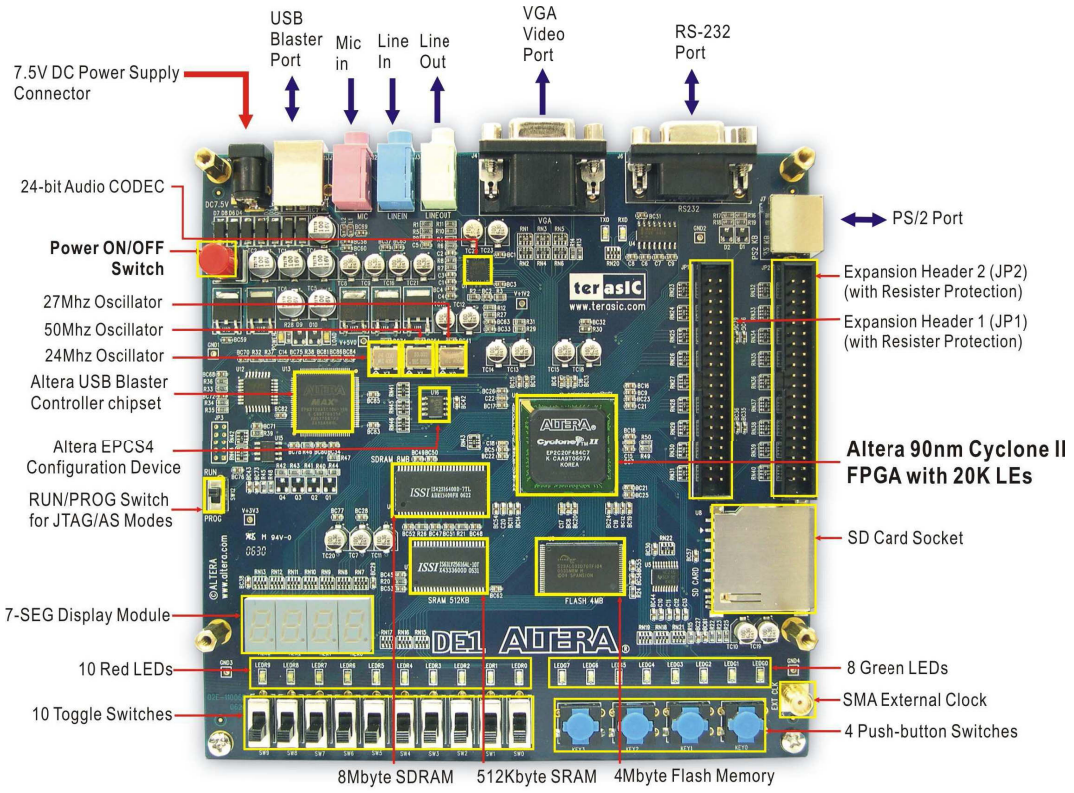


Figure 6.3: Altera DE1 Board [5]

6.3.2 Signal generator and ADC system diagram

The basic clock was set to be 50MHz and the 50MHz clock was divided into all the clocks required. In the FPGA, some hardware and a 32bits Nios II microprocessor were built to control the whole system. The micro controller connects to JTAG port and communicates with PC through USB-blaster. The USB-Blaster circuit is a FT245BM chip and a CPLD. The parallel output data from the FT245BM FIFO was serialized by the CPLD to drive the JTAG pins on the attached FPGA. The highest performance mode is when the packets are 'byte-mode' packets, where each byte from the host generates 8bits of JTAG TDI activity. JTAG bit rate is 6 MHz with USB-Blaster. The best data rate is 1MB/s divided by 8, or about 1Mbps. Therefore, in the first design, the data could not be sent in real-time. One solution was to store the data in SRAM, then write the SRAM to the PC. The other method was that when the optimized algorithm was found for the radar signal processing, the FPGA was used to process these data in advance, therefore very limited data were required to send to the PC. The 3.3V 40kHz square wave to be transmitted was created. To make the ultrasonic wave more powerful, a XOR gate was used to generate two inverse 5V signals. The inverse signals were connected to the two pins of the ultrasonic transmitter. As the characteristic of the ultrasonic transducer is similar to a capacitor, when the signal was sent to the ultrasonic transducer, the actual wave was 40 kHz sine wave at 10V. The programing blocks can be seen in Figure 6.4.

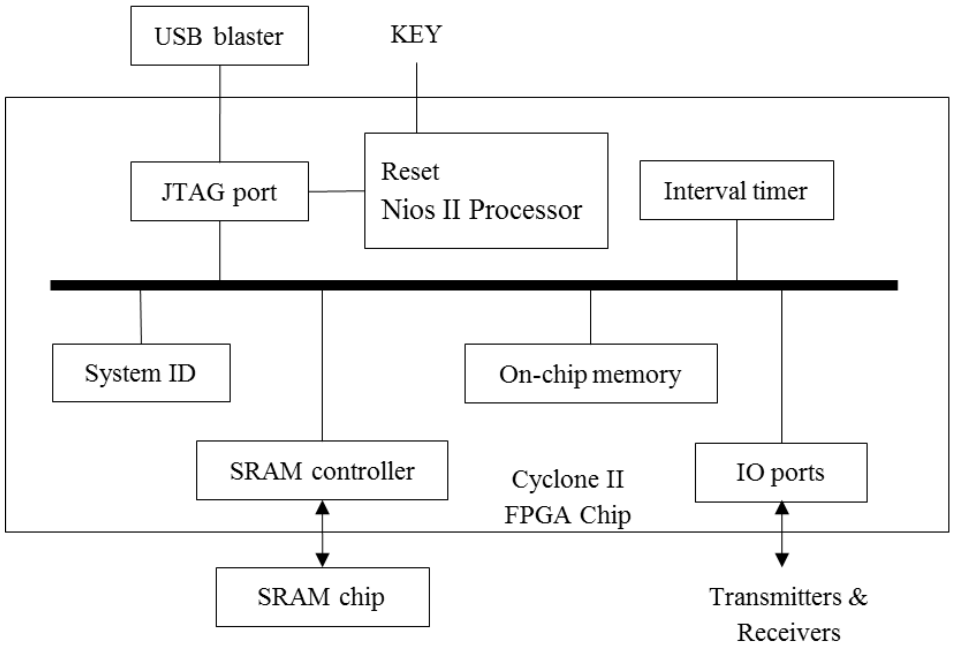


Figure 6.4: Blocks inside FPGA

6.3.3 FPGA Programming

The signal generator and ADC system diagram are shown in Figure 6.5. The explanation of the function is as follows:

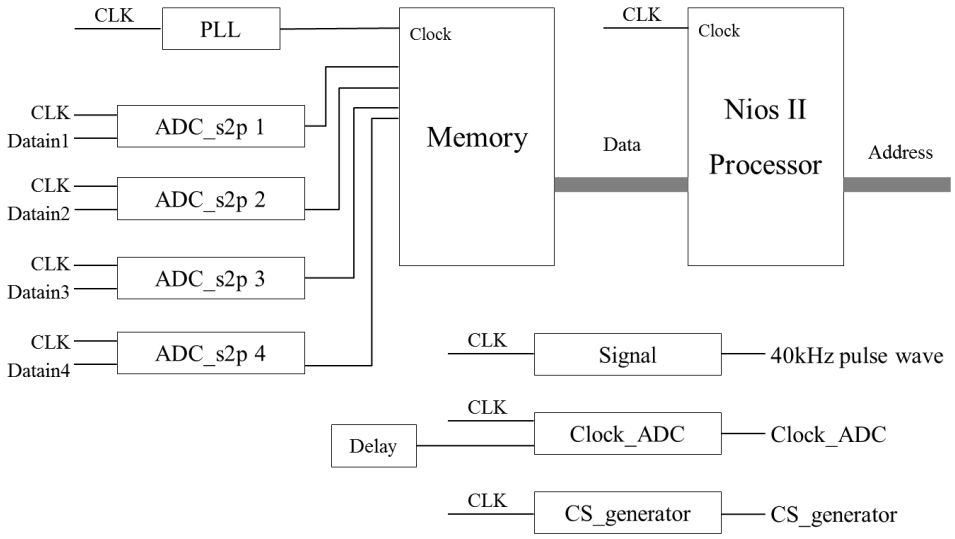


Figure 6.5: Signal generator and ADC system diagram

- DAC chip: The DAC chip MCP3001 has 10 bit resolution, 200ksps sampling rate at 5V, a SPI serial interface (modes 0,0 and 1,1). This chip requires three inputs: Data, CS and CLK to convert and transport data.
- Signal: Create the 40 kHz pulsed wave or continuous wave as specified in the system description.
- Clock_ADC: The ADC clock is to synchronize with the clock of the ADC. This clock is also the data transport clock between the ADC and DE1 board; as the clock must more than 14 times of the ADC sample rate. In this thesis, as the sample frequency is 160kHz, the ADC clock was set to 2.5MHz.
- CS_generator: The Chip Select (CS) clock was sent to the ADCs to control the sample rate. A 92% duty cycle was required to allow the ADCs to send data. Also, this signal is to enable data conversion and memory storage. The CS_generator was set to 160 kHz.
- Delay: According to the data sheet, delays were used to make the waveform more accurate.
- ADC_s2p: Convert the serial data from the ADCs to parallel data according to data format.

- Timer: Count the time from each pulse and record the time when an object is detected.

- Memory: Store the sample data to SRAM and use SOPC to access the data and send them to the host computer through JTAG and USB_blaster. A static RAM (SRAM) Controller provides a 32-bit interface to the SRAM chip on the DE1 board. The SRAM memory is mapped to the address space 0x08000000 to 0x0807FFFF.

- Nios II microprocessor: The Altera Nios II processor is a 32-bit CPU. In the first stage of this project, the processor contained 1 input: 10bit data, which receives the data from ram; 2 outputs: 1 bit enable, which enables the whole system and controls when to write to and read from SRAM; 19 bit address, which help the host computer to read data from SRAM.

- PLL block: according to the data sheet, the SRAM clock should be double the data rate, so a PLL block is required.

6.4 The PCB board layout

The PCB board was designed with Altiumn Protel DXP. The schematic diagram is shown in Figure 6.6.

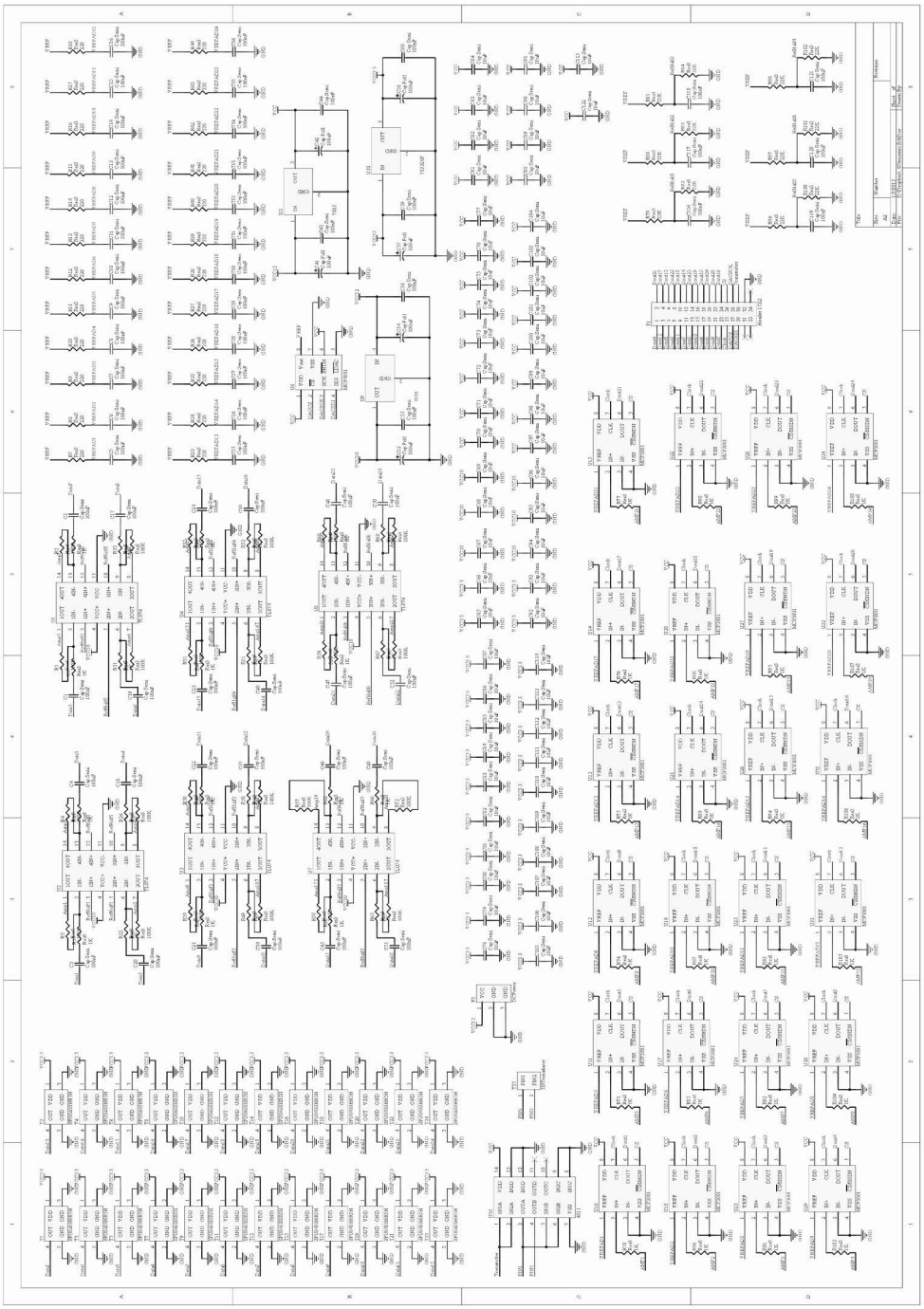


Figure 6.6: Schematic diagram of the designed board

To assemble the board, the components were carefully laid out on the PCB board. The solder tin was put on the board in advance and a toasting oven was used to heat and wait for the cooling of soldering tin. Then the components were fixed onto the board. Further tests on each channel were required. The layouts of both the top layer and bottom layer of the PCB board are shown in Figure 6.7 and Figure 6.8.

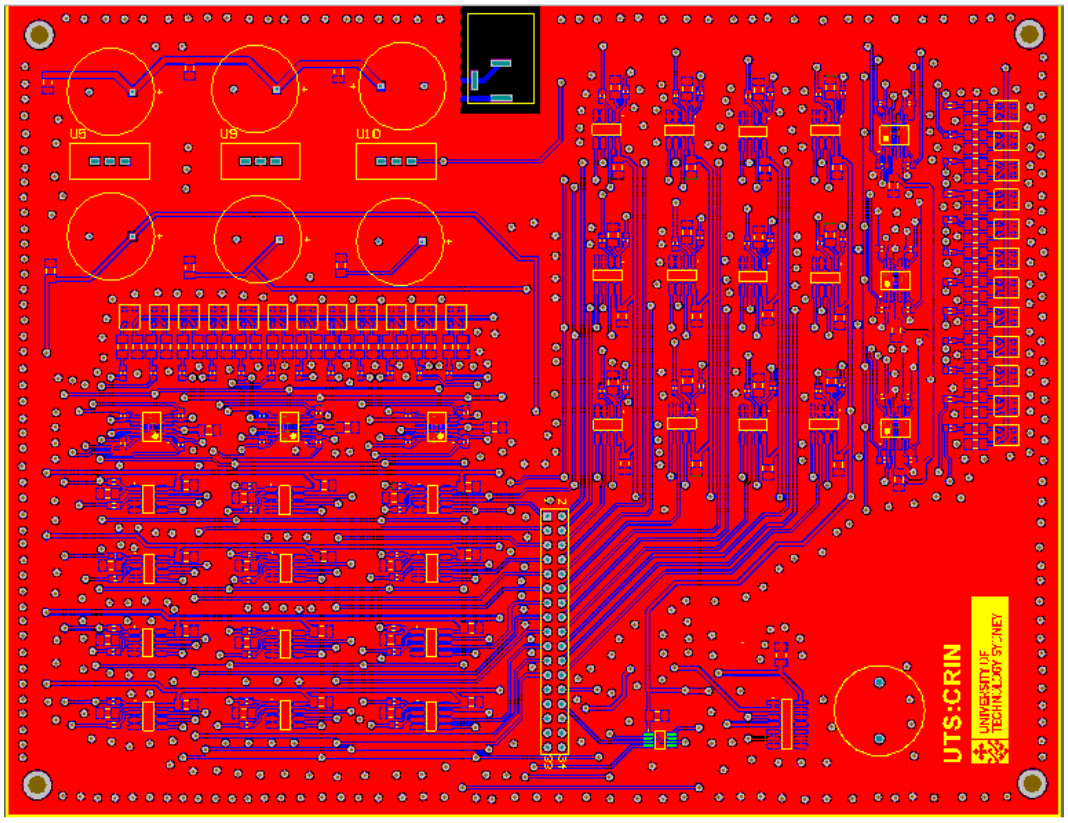


Figure 6.7: Top view of the PCB

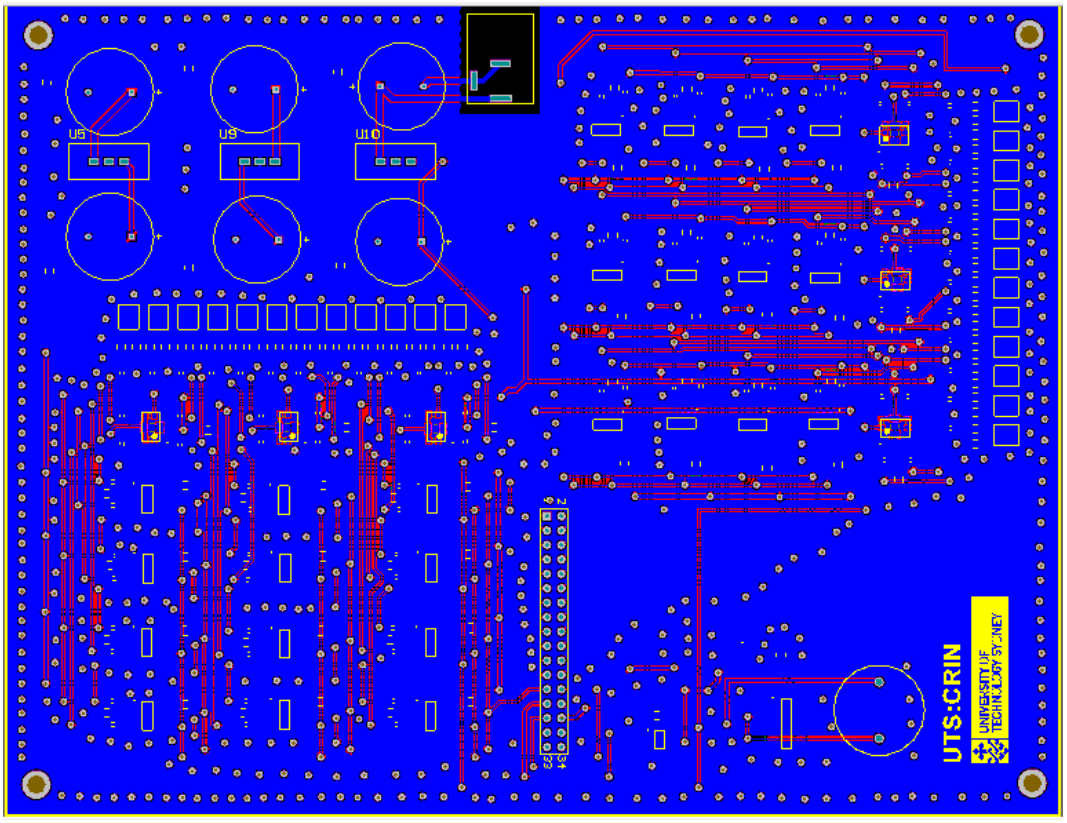
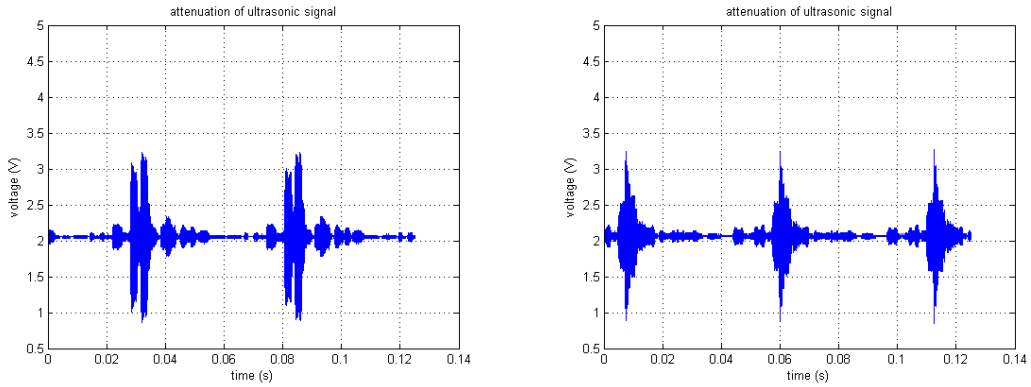


Figure 6.8: Bottom view of the PCB

6.5 Experiments

As mentioned before, for the purpose of research, all the raw data needed to be imported to the PC through the FPGA. If the sample rate is 160 kHz, sample data are too large for the USB 2.0 to transport in real time. The FPGA board does not have a USB 3.0 interface. Here four channels are taken as an example to demonstrate the linear beamforming. The experiment of two dimensional beamforming will be done in the future after the USB 3.0 interface can be used. Alternatively, the algorithms can be executed inside the FPGA to reduce the data rate. In this case, a FPGA with more logic gate resources is required.



(a) Received signal 1

(b) Received signal 2

Figure 6.9: Attenuation of 40kHz ultrasonic

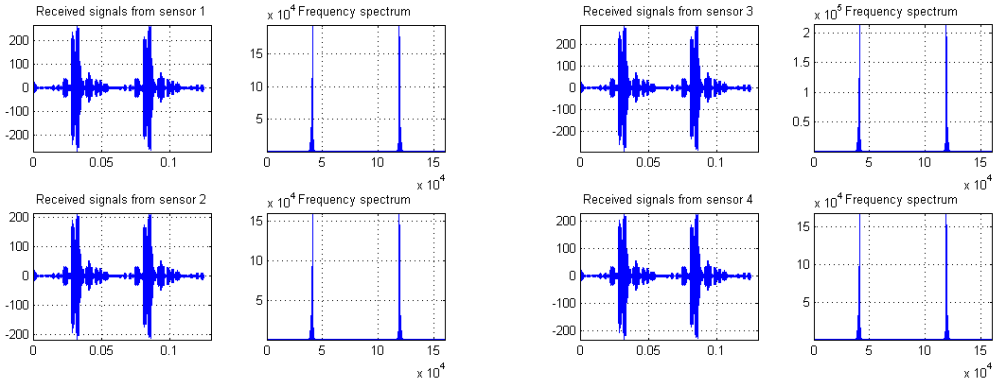
6.5.1 Attenuation of ultrasonic sound

The acoustic radar system was tested. The range of the targets was between 5 to 15 meters. The amplified ratio of the amplifying circuit based on op-amps was approximately 100. The results are shown in Figure 6.9.

A typical ADC can only work at a voltage of 0-5 Volts. An offset voltage needs to be added to the received signals. In the experiments, the offset voltage was 2.1 V. The maximum amplitude of the voltage needs to be limited at 5 V.

6.5.2 Experiments on the linear array - pulsed beamforming

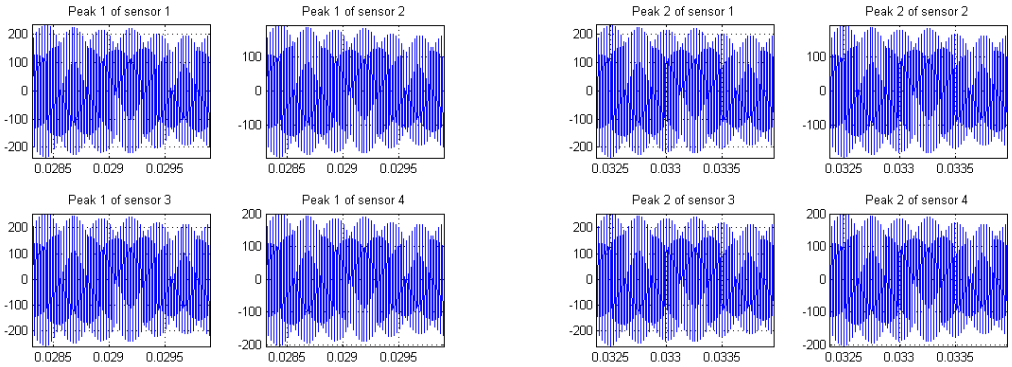
Figure 6.10 shows the received signals from four channels.



(a) Sensor 1 **(b)** Sensor 3

Figure 6.10: Received signals

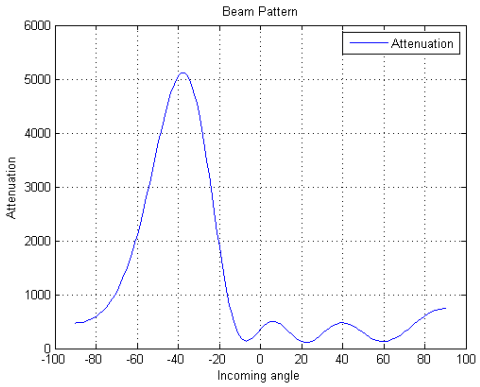
There were two peaks in the signals. The two peaks were subtracted as follows in Figure 6.11.



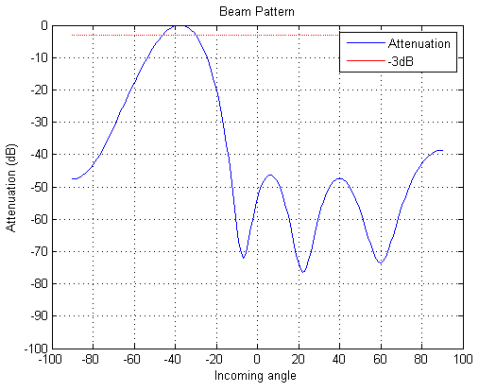
(a) Peak 1 **(b)** Peak 2

Figure 6.11: Two peaks of the received signals

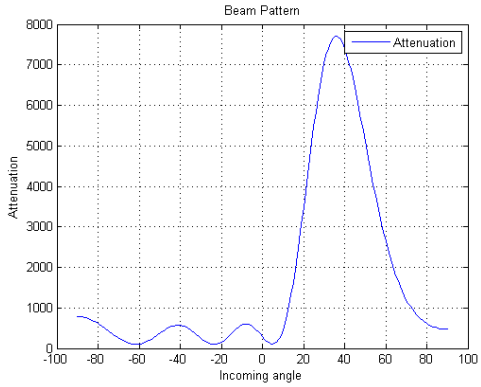
The beamforming processing results are shown in Figure 6.12.



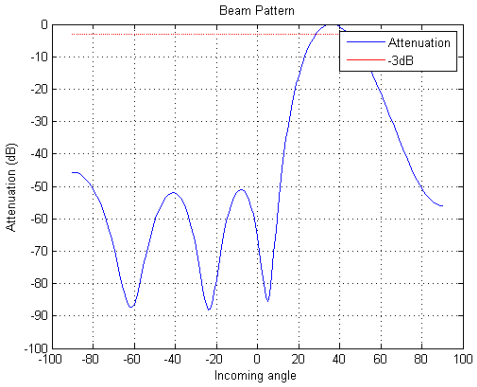
(a) -38º - Linear scale



(b) -38º - Log scale



(c) 38º - Linear scale



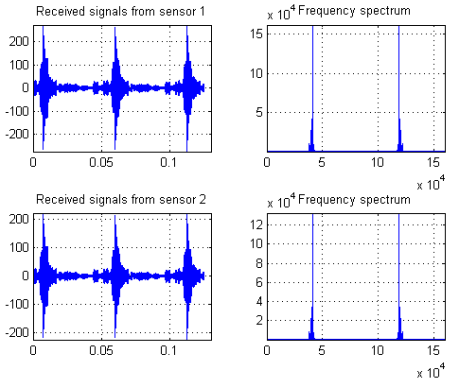
(d) 38º - Log scale

Figure 6.12: AoA of two objects

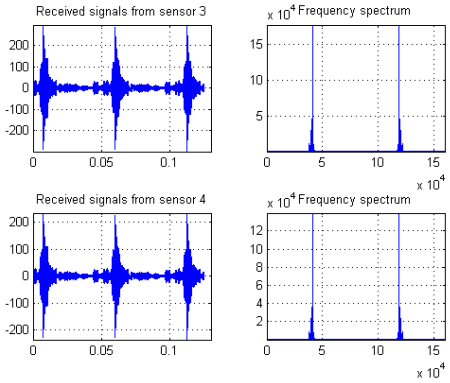
6.5.3 Experiments on the linear array - Doppler beamforming

Figure 6.13 shows the received signals from four channels.

There were two peaks in the signals. The two peaks were subtracted in Figure 6.14.

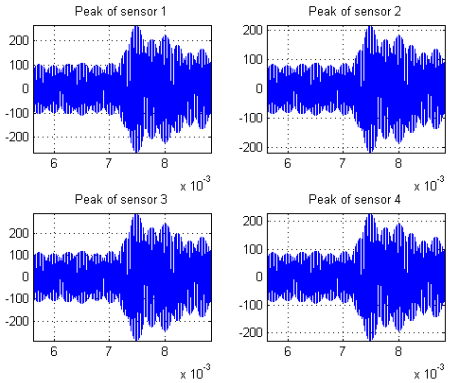


(a) Received signals from sensor 1 and 2

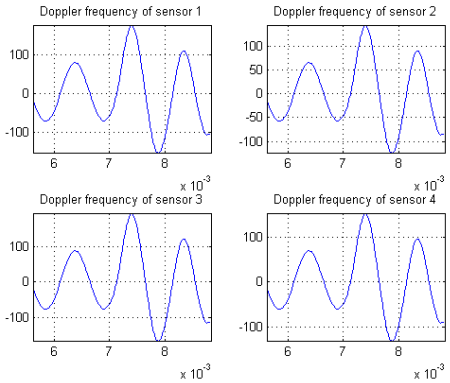


(b) Received signals from sensor 3 and 4

Figure 6.13: Received signals



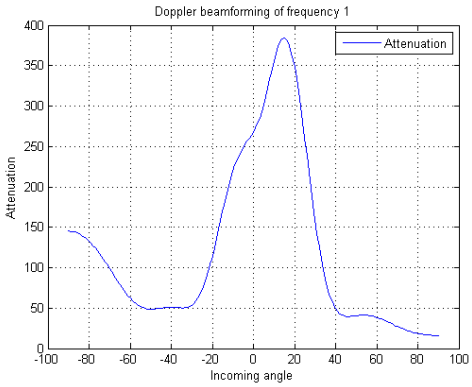
(a) The peak



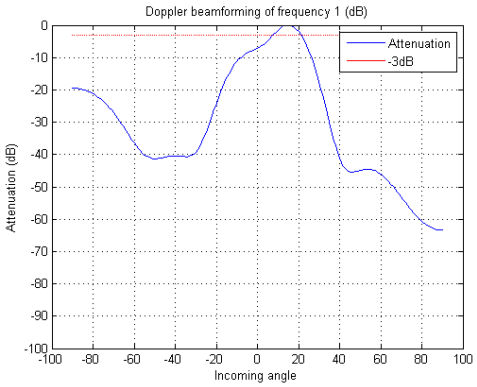
(b) Doppler frequency

Figure 6.14: Two peaks of the received signals

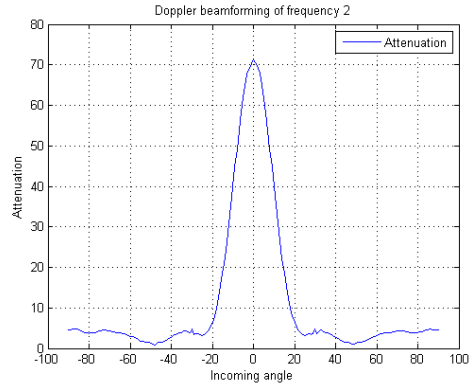
The beamforming processing results are shown as in Figure 6.15.



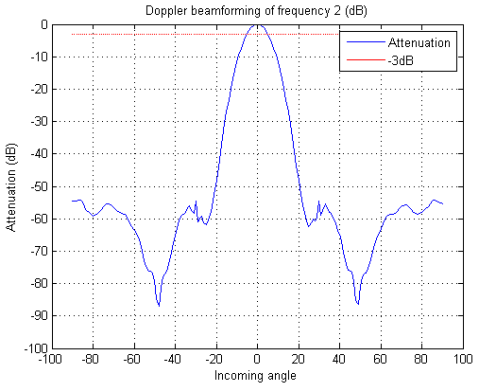
(a) Doppler beamforming of frequency 1



(b) Doppler beamforming of frequency 1 (dB)



(c) Doppler beamforming of frequency 2



(d) Doppler beamforming of frequency 2 (dB)

Figure 6.15: Beam pattern of linear array Doppler beamforming

6.5.4 Conclusion of the experiment

The experiment showed that both linear array pulsed beamforming and Doppler beamforming can work. However, due to the limitation of the hardware, the experiment cannot be extended to two dimensions. That work will be done in the future. With the development of the FPGA processing capability, both the logic gate number and processing speed can be improved. Also, the transmission between the FPGA and the PC can be improved by the use of USB 3.0. If the data are processed inside the FPGA, the data rate can be reduced to a very low

level. In chapter 3 and chapter 4, the beamforming algorithm is extended to two dimensions in simulation based on the result of the experiment.

7 Conclusion and future work

7.1 Thesis summary and validation of research propositions

This thesis mainly contains 4 parts: (1) the system design and hardware construction, (2) the construction of the range layer of the radar image, (3) the construction of the radial velocity layer of the radar image and (4) the applications of fusing these two extra layers of radar image with traditional RGB layers. The system design and hardware construction are the basis of the following parts and the next three parts are to solve the Propositions.

The thesis is all about the creation of new systems that have not been created before, and have resulted in a provisional patent, with sub patents. Therefore, there is no existing literature describing a similar system or its performance. So, there is no literature to compare against.

System design and hardware construction

The system contains the optical imaging subsystem and the radar imaging subsystem. The optical imaging subsystem contains a high quality web-cam and the radar imaging subsystem needed to be designed for specific applications. According to the supposed application environment, the radar is a vision range airborne acoustic radar. The radar subsystem contains a transmitting and receiving board, a FPGA board and a PC. All raw data are transmitted to the PC to process. There are some key points in designing the hardware.

One difficulty was that all the channels needed to work simultaneously. This is part of the reason why the hardware needed to be built instead of using the existed multiple channel devices. In most of these devices, some channels share the same ADC core and use the core in the time division way. In most ADC chips with multiple channels, there is actually only one ADC core. The beamforming processing needs to compare the phase difference between sensors, so even though the switching between channels is very fast, the switching time will affect the phase difference. In the design of the appliance in this thesis, for each receiver, an one channel ADC chip is adopted. Therefore the circuits design is more complicated and the cost is higher when there are more sensors.

Another difficulty was to optimize the resource consumption. As discussed before, the data rate and computation load were very high when the number of sensors was increased. For a low cost development, the resources are very limited. One way to process more data is to pre-process the data. The pre-processing methods include the following. First, the presence of targets need to be detected and the signals without targets need to be discarded. Only one channel is needed to verify the presence of the target because the phase difference between adjacent sensors is less than a cycle. Then not all the sampled data need to be transmitted to the PC. Second, the amplitude of the signals needs to be adjusted to the same level and fewer bits in each sample are required. According to the simulation in **Proposition 1** and **Proposition 2**, the beamforming does not need highly accurate sample value. Maybe 8 bits or even 6 bits are high enough. Third, in some applications such as golf ball tracking, the system does not need to work all the time. The sampled data can be stored in the FLASH memories which have very high storage speed and the capacity of the memories may be up to 64 GByte. After capturing the data, the system has enough time to transfer and process the data.

The signal to noise ratio is important to the system and the longer the distance the ultrasonic waves travel, the weaker the received signal will be. The attenuation was calculated and the experiment results showed the received signals can clearly reflect the targets. As the signals adopted are 40 kHz narrow band signals, the additive noise can hardly affect the beamforming image, which was discussed in chapter 3 and chapter 4.

Though the system construction is not in the propositions, it is the basis of all the propositions. The hardware is demonstrated to work for the creation of the image layers.

Proposition 1

Proposition 1 is whether or not a system can be built to create an image using a pulsed beamforming algorithm that has ranges as the pixel values. Beamforming is widely used in target tracking, but it is not popularly used for creation of an image. By investigating the principles of linear array time domain and frequency domain beamforming, the two dimensional beamforming is studied. The two dimensional AoA can represent the position of pixels which reflect the real objects. Some simulations have been done to verify the resolution when there are multiple targets. The results show that the resolution depends on the phase difference of the two objects. If the phase is similar, the two objects interfere with each other, and if the phase is inverse, the two objects do not interfere with each other. As the phase difference between multiple targets cannot be controlled, the way to increase resolution is to increase the sensor number. In two dimensional beamforming, the full rectangular sensor array has the best performance. The main lobe to side lobe ratio is very high, which means the targets can easily be subtracted from the image.

As discussed in the hardware construction part, the computation load and memory consumption costs increases with the sensor number. At least, the current FPGA and micro controller can not process such an amount of data in real time. One way is to reduce the unit cost of sensors; the other way is to reduce the sensor number. Therefore, performance of the two line multiplication beamforming, crossed array beamforming and other structures of two dimensional beamforming radar array was tested. The results show that two line multiplication beamforming has a high main lobe to side lobe ratio, but it can not deal with ambiguities when there was more than one object. However the ambiguities can be removed by the optical imaging. Among other shapes of sensor arrays, some types of sparse rectangular sensor arrays have acceptable results. To evaluate the results, an example image is set with one person and two birds and the simulation results are listed. The ability to resist noise is tested and the results show that the system is very robust to noise. For the range information detection, the triangular method and pulsed radar are two popular ways. The triangular method is not suitable because it cannot deal with the situation when there are multiple targets. The pulsed radar is more accurate in distance detection. However as the AoA needs to be detected in the same time and a N series of samples need to be processed, therefore, the resolution of the distance is about half a meter. In electromagnetic wave radar, the distance resolution can be high. According to the results above, the final algorithm is proposed and an example of the proposed image layer is provided. Therefore, the image layer with range information as well as the hardware can be constructed.

Proposition 2

Proposition 2 is whether or not a system can be built to create an image using beamforming with Doppler principles that has radial velocity as the pixel values. The benefit of the Doppler radar is that it can only sense the moving objects. In most applications, the targets to be tracked are moving objects. Also, Doppler beamforming can separate objects with different velocities. In most applications, the targets are fast moving objects, which can be easily subtracted from the background. Therefore, in this case when optical imaging does not work well, a Doppler beamforming image is more useful than a pulsed beamforming image. In the first version of the simulation, the Doppler beamforming image is not clear and the zero frequency components affect the results. When the high pass filters are used to filter the zero frequency, the Doppler beamforming image is much more clear. The ideal way to separate the targets with different velocities is using FIR filter banks. This method consumes a considerable amount of hardware resources. For the cost consideration, the FFT filtering method was tested and the results show the difference between FIR filter bank and FFT filtering is very small, and the computation cost of FFT filtering is much lower. The performance when there were multiple targets was tested. Doppler beamforming is also very robust to noise as narrow band 40 kHz waves were used. Because pulsed beamforming and

Doppler beamforming share the same hardware, the performance of two line multiplication beamforming, crossed sensor array beamforming and other shapes of sparse rectangular array beamforming were tested. Finally, the results were concluded and an example of the radar image layer with radial velocity information was provided. Therefore, the image layer with radial velocity information as well as the hardware can be constructed.

Proposition 3

Proposition 3 is that whether or not the relationship between low resolution radar image layers and high resolution optical image layers can be found, and the target tracking can be enhanced by fusing the two additional layers with the optical image. The parameter differences between the proposed beamforming radar image and the optical image have be investigated in chapter 3. The difference between the radar image and the optical image is that the pixel values in the radar image represent ranges and radial velocity while the pixel values in the optical image represent intensity of reflection light. The pixels with the same AoA in the radar image and the optical image should reflect the same objects, though there may be a slightly difference. The calibration of the radar and the camera and the registration of the radar image and the optical image are required to match pixels. Due to the resolution difference, one pixel in the radar image relates to a region of pixels in the optical image.

In different applications, the environment and the requirements vary considerably. The fusion algorithms should be based on the applications. Therefore five scenarios were studied.

Scenario 1

The scenario 1 is the fusion of the video camera frames with Doppler beamforming radar to enhance moving object subtraction. The idea arises that both the optical frame differencing image and the Doppler beamforming image can mark the moving objects. The shortcoming of frame differencing of optical images is that this method can be affected by illumination changing, small moving objects such as tree leaves, and only the edge of the objects can be detected. The shortcoming of Doppler beamforming radar is that the image resolution is very low, so the edge of the objects cannot be detected. However Doppler beamforming does not require illumination. Based on the characteristics, the purpose of this scenario is to fuse these two types of image to remove unwanted effects and improve the subtraction of moving objects.

The simulation shows that fusion can remove or reduce ambiguity such as tree leaves, shadows and also improve the subtraction of far distance objects with high velocity.

Scenario 2

The scenario 2 is the fusion of the crossed array Doppler radar with the optical image to enhance the golf ball tracking. The idea is that the crossed array radar tracks the target first and creates the moving object image with ambiguity. Image processing methods such as template matching and frame differencing are applied in the detected area which may contain ambiguity to make the decision whether there are objects in these areas or not. The results show that images can identify the possibilities of objects. By this method the cost of Doppler beamforming radar can be reduced sharply because the data rate, computation load and memory consumption are reduced. The simulation shows that the frame differencing can easily detect the moving golf ball and remove ambiguity.

Scenario 3

The scenario 3 is the fusion of the optical image with Doppler beamforming radar to enhance tennis ball tracking. The radar system and optical imaging system work separately. Doppler beamforming radar detects the position of the tennis ball. The region of interest is marked according to the position of the ball. Kalman filtering is used to reduce the errors in the trajectory. When the tennis ball approaches the ground, the optical imaging searches for the ball only in the region of interest. The trajectory is estimated and the decision is made whether the ball is in or out. The simulation demonstrated that it is possible to effectively use imaging principles and techniques to determine the location at which the ball has landed on the tennis court surface.

It is recommended that further investigation takes place in regards to applying machine learning to automate the in/out decision process. Further, the allocation of cameras should be finalized. Static location of cameras will introduce a level of uniformity to the images that are captured and consequently processed by the imaging sub-system. Uniformity simplifies the processing aspect of the overall system. Third, further testing needs to be carried out with the images where the court lighting varies. That is, testing with images that have been captured on a court lit only by the day light, an indoor court lit only by the indoor lighting and outdoor court lit by court lighting and any ambient light.

Scenario 4

The scenario 4 is the fusion of the optical imaging and the beamforming radar to track targets in 3D position. The Kalman filter is applied to minimize the measurement errors. Two fusion algorithms, state vector fusion and measurement fusion are compared. The RMS errors of both the state vector fusion and measurement fusion are provided.

Scenario 5

The scenario 5 is using Doppler beamforming radars to improve gait analysis. The micro-Doppler characteristics are used to analyze the human movements. Then the Doppler processing is combined with Doppler beamforming to relate the Doppler frequency to the body parts, which can be projected into the optical image. The simulation shows that not only the frequency changing in the Doppler signals, but also the position of the moving objects can be known.

7.2 Thesis limitations

Limitation in hardware

As discussed before, the biggest limitation of the hardware is the conflicts between hardware resources and system requirements. Due to this limitation, in experiments, only four receivers were tested. The complicated structures with more sensors were done in simulation.

Also for research purposes, the final hardware was not assembled. The FPGA development board was used in the experiment. To make an industrial product, the FPGA, sensors and sample circuit should be put on one board. Then the cost can be reduced and less noise should be made. Also the system would be more robust.

Limitation in theoretical algorithms

The computation load of the beamforming algorithm is high. For a two dimensional frequency domain beamforming, the assumed AoA needs to be scanned pixel by pixel. For a high resolution beamforming image, there are more than 1 mega pixels, and for each pixel, the sum of the estimated delay from each sensor needs to be calculated. If all the data are processed in the PC, the computation ability is not sufficient to track in real time. This problem can be solved in the hardware side. There may be also many methods to reduce the computation load in algorithms.

The resolution of the beamforming image is low. This is due to the wide beamwidth of the beamforming spectrum. One way to narrow the beamwidth is to increase the sensor number. This method is not cost effective and there may be some ways to increase resolution with fewer sensors.

Pulsed beamforming and Doppler beamforming share the same hardware. The radar needs to switch between the pulsed mode and the Doppler mode in a time division way. One way to get the clear beamforming image is to increase the FFT series length, but this also limits the frame rate and range resolution. In

acoustic radars, one method is to increase the carrier frequency, but the method also increases the sampling load and the attenuation.

The other limitation is the ambiguity problem in the sparse rectangular sensor array beamforming. Some solutions have been proposed in this thesis.

Limitation in fusion methods

The fusion methods were tested in simple environments. In real application, the environment is much more complicated. This is a problem in most optical imaging algorithms. Also, there may be some interference. How to deal with the interferences needs to be studied.

In this thesis, most results are evaluated by people. The final decision making algorithms have not been investigated. The machine learning algorithms need to be applied to evaluate the system and the fit rate needs to be compared.

7.3 Recommendation and further work

The thesis proposed a new image format which contains traditional red, green and blue light intensity layers and range and radial velocity layers. Some fusion algorithms for the use of the new image format are investigated. Based on the new proposed system, much future work can be done, such as improving the system accuracy, reducing the computation load and the memory consumption. Also, the most important thing is to find more proper applications for the proposed system.

Improve the accuracy of the beamforming

The accuracy can be improved in the following ways: the first is to optimize the hardware design to minimize the noise and measurement errors. The second is to increase the sensor number as well as the capabilities of the hardware. The third is to optimize the algorithms.

Reduce the computation load and memory consumption

The conflicts between performance and computation load as well as memory consumption need to be solved. One way is to apply the final algorithm into a FPGA. The second way is to reduce the sensor number, in the meantime improving the resolution and reducing the ambiguities.

Improvement in the beamforming algorithms

The beamforming algorithm can be improved in these ways: first the adaptive beamforming can be used to inhibit the ambiguity level; second other shapes of sensor arrays need to be tested such as the honeycomb type; third the multiple beam beamforming and the MUSIC algorithm can be tested; fourth some methods to narrow the beamwidth can be investigated.

Apply the system in more applications

Both the range layer and radial velocity layer of the image may be very useful for optical imaging such as in video compressions. Knowing the moving parts can optimize the compression results. More applications can be explored and the hardware can be adjusted according to the tracking purpose.

Setup the hardware and test in complicated environments

The hardware used was tested in simple environments. In complicated environments, they have not been tested. In the future, more experiments need to be performed according to the application.

Acknowledgments

I would like to acknowledge all the help and encouragement received during my PhD research. My special thanks go to my supervisor: Prof. Robin Braun. I learned a lot from his mature view on topics from a high level and excellent computer skills. I am gratefully thanking for his guidance to me throughout the years. His kindness and encouragement made me confident when I face difficulties.

I also would like to thank China Scholarship Council (CSC) and University of Technology, Sydney (UTS) who provide the CSC and UTS PhD Scholarship to support my PhD research financially.

My parents gave me endless care and encouragement. They always support me and try to get me the best education they can afford throughout the years I grew up. They have been the important driving force to encourage me behind my PhD research. I am sincerely thanking them.

Some contents in Chapter 1 and Chapter 2 have been submitted to a UTS subject (32144: Technology Research Preparation). This subject aims to guide research students to prepare literature reviews. As these contents have not be published, they cannot be referenced but can be searched by some software such as Turnitin.

Bibliography

[1] S.-E. Hamran, *Radar performance of ultra wideband waveforms*. INTECH Open Access Publisher, 2010.

[2] M. C. Sullivan, *Practical Array Processing*. McGraw-Hill, 2009.

[3] B. Gary and K. Adrian, "Learning OpenCV: computer vision with the OpenCV library," *O'Reilly USA*, 2008.

[4] C. Buccella, V. De Santis, M. Feliziani, and P. Tognolatti, "Finite element modelling of a thin-film bulk acoustic resonator (FBAR)," *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 6, pp. 1296 – 1306, 2008.

[5] "Altera DE1 Board Resources for Students." [Online]. Available: http://users.ece.gatech.edu/~hamblen/DE1/

[6] T. E. Curtis and R. J. Ward, "Digital beam forming for sonar systems," in *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 127, no. 4. IET, 1980, pp. 257–265.

[7] U. Hamid, R. A. Qamar, and K. Waqas, "Performance comparison of time-domain and frequency-domain beamforming techniques for sensor array processing," in *Applied Sciences and Technology (IBCAST), 2014 11th International Bhurban Conference on*. IEEE, 2014, pp. 379–385.

[8] M. S. Nixon, T. Tan, and R. Chellappa, *Human identification based on gait*. Springer-Verlag New York Inc, 2006, vol. 4.

[9] M. Azghani, A. Aghagolzadeh, S. Ghaemi, and M. Kouzehgar, "Intelligent modified mean shift tracking using genetic algorithm," in *Telecommunications (IST), 2010 5th International Symposium on*, 2010, no. x, pp. 806–811.

[10] S. S. Ram, Y. Li, A. Lin, and H. Ling, "Doppler-based detection and tracking of humans in indoor environments," *Journal of the Franklin Institute*, vol. 345, no. 6, pp. 679–699, 2008.

[11] K. G. Foote, R. Patel, and E. Tenningen, "Target-tracking with a multiple-frequency split-beam echo sounder," in *OCEANS 2011*, 2011, pp. 1–9.

[12] R. S. Blum and Z. Liu, *Multi-sensor image fusion and its applications*. CRC press, 2005.

[13] Z. Hou and C. Han, "A target tracking system based on radar and image fusion," *Proc. of the 6 ISIF*, pp. 1426–1432, 2003. [Online]. Available: http://ftp.isif.org/fusion/proceedings/fusion03CD/poster/p79.pdf

[14] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and guard rail detection using radar and vision data fusion," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 95–105, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4114330

[15] A. Bonci, G. Ippoliti, A. L. Manna, S. Longhi, and L. Sartini, "Sonar and Video Data Fusion for Robot Localization and Environment Feature Estimation," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, 2005, vol. 5, pp. 8337–8342.

[16] D. Muller, J. Pauli, M. Meuter, L. Ghosh, and S. Muller-Schneiders, "A generic video and radar data fusion system for improved target selection," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, 2011, no. Iv, pp. 679–684.

[17] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier, "Fade: a vehicle detection and tracking system featuring monocular color vision and radar data fusion," in *Intelligent Vehicle Symposium, 2002. IEEE*, 2002, pp. 632—-639.

[18] S. L. Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover," *Pattern recognition*, vol. 31, no. 11, pp. 1811–1823, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132039800051X

[19] R. Labayrade and M. Perrollaz, "Sensor Data Fusion for Road Obstacle Detection : A Validation Framework," pp. 375–394, 2006.

[20] J. R. Raol, *Multi-Sensor Data Fusion with MATLAB*. CRC Press, 2009.

[21] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *Image Processing, IEEE Transactions on*, vol. 10, no. 4, pp. 623–631, 2001.

[22] K. Hasselmann and S. Hasselmann, "On the nonlinear mapping of an ocean wave spectrum into a synthetic aperture radar image spectrum and its inversion," *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 96, no. C6, pp. 10 713–10 729, 1991.

[23] S. Hasselmann, C. Brüning, K. Hasselmann, and P. Heimbach, "An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra," *Journal of Geophysical Research: Oceans (1978–2012)*, vol. 101, no. C7, pp. 16 615–16 629, 1996.

[24] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. L. Williams, "Textural infornation in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 24, no. 2, pp. 235–245, 1986.

[25] H. A. Zebker and R. M. Goldstein, "Topographic mapping from interferometric synthetic aperture radar observations," *Journal of Geophysical Research: Solid Earth (1978–2012)*, vol. 91, no. B5, pp. 4993–4999, 1986.

[26] R. A. Moursund, T. J. Carlson, and R. D. Peters, "A fisheries application of a dual-frequency identification sonar acoustic camera," *ICES Journal of Marine Science: Journal du Conseil*, vol. 60, no. 3, pp. 678–683, 2003.

[27] A. Meyer and D. Döbler, "Noise source localization within a car interior using 3D-microphone arrays," *Proceedings of the BeBeC*, 2006.

[28] Wikipedia, "Radar," p. 1, 2012.

[29] K. Kalgaonkar and B. Raj, "Acoustic doppler sonar for gait recogination," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on.* IEEE, 2007, pp. 27–32.

[30] Z. Zhang, P. O. Pouliquen, A. Waxmant, and A. G. Andreou, "Acoustic micro-Doppler gait signatures of humans and animals," *Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference on*, no. 14-16 March 2007, pp. 627–630, 2007.

[31] Z. Zhang, P. O. Pouliquen, A. Waxman, and A. G. Andreou, "Acoustic micro-Doppler radar for human gait imaging," *The Journal of the Acoustical Society of America*, vol. 121, no. 3, pp. 110–113, 2007. [Online]. Available: http://link.aip.org/link/JASMAN/v121/i3/pEL110/s1&Agg=doi

[32] D. S. Childress and R. F. Weir, "A new method of characterising gait using a portable, real-time, ultrasound ranging device," in *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, 1997, vol. 1810, no. C, pp. 1810–1812.

[33] A. Sabatini and V. Colla, "A method for sonar based recognition of walking people," *Robotics and Autonomous Systems*, vol. 25, no. 1-2, pp. 117–126, 1998.

[34] T. Karhik and A. Kralj, "Gait dynamic stability assessment in a sagittal plane," in *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, 1996, pp. 467–468.

[35] T. Bajd, A. Kralj, and T. Karcnik, "Unstable states in four-legged locomotion," *Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94. Proceedings of the IEEE/RSJ/GI International Conference on*, vol. 2, pp. 1019–1025, 1994.

[36] H. Peremans, K. Audenaert, and J. Van Campenhout, "A high-resolution sensor based on tri-aural perception," *IEEE Transactions on Robotics and Automation*, vol. 9, no. 1, pp. 36–48, 1993. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=210793

[37] M. Hewish, "The last line of defense: Reducing the vulnerability of fixed installations," *Jane's International Defense Review*, vol. 32, no. 10, pp. 28–30, 1999.

[38] V. G. Nebabin and V. V. Sergeev, "Methods and techniques of radar recognition," *Moscow Izdatel Radio Sviaz*, vol. 1, 1984.

[39] J. L. Geisheimer, W. S. Marshall, and E. Greneker, "A Continuous-Wave (CW) Radar for Gait Analysis," *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 1, pp. 834–838, 2001.

[40] P. van Dorp and F. C. A. Groen, "Human walking estimation with radar," in *Radar, Sonar and Navigation, IEE Proceedings-*, vol. 150, no. 5, 2003, pp. 356 – 365. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1249155

[41] R. Boulic, N. M. Thalmann, and D. Thalmann, "A global human walking model with real-time kinematic personification," *The visual computer*, vol. 6, no. 6, pp. 344–358, 1990.

[42] S. Gurbuz, W. Melvin, and D. Williams, "Detection and Identification of Human Targets in Radar Data," *Signal Processing, Sensor Fusion, and Target Recognition XVI. Edited by Kadar, Ivan. Proceedings of the SPIE*, vol. 6567, p. 65670I, 2007.

[43] M. Otero, "Application of a continuous wave radar for human gait recognition," in *Signal Processing, Sensor Fusion, and Target Recgonition XIV. Edited by Kadar, Ivan. Proceed- ings of the SPIE*, 2005, pp. 538–548. [Online]. Available: http://72.52.208.92/~gbpprorg/mil/cavity/05_0330.pdf

[44] S. Z. Gurbuz, W. L. Melvin, and D. B. Williams, "Comparison of radar-based human detection techniques," in *41st Asilomar Conf*, 2007, pp. 4–7.

[45] S. Gurbuz, W. Melvin, and D. Williams, "Radar-based human detection via orthogonal matching pursuit," *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, vol. 1, no. 11, pp. 1829–1841, 2010. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstracthttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5496209

[46] C. P. Lai, Q. Ruan, and R. M. Narayanan, "Hilbert-Huang transform (HHT) analysis of human activities using through-wall noise radar," in *Signals, Systems and Electronics, 2007. ISSSE'07. International Symposium on.* IEEE, 2007, pp. 115–118.

[47] C. P. Lai, R. M. Narayanan, Q. Ruan, and A. Davydov, "Hilbert-Huang transform analysis of human activities using through-wall noise and noise-like radar," *Radar, Sonar & Navigation, IET*, vol. 2, no. 4, pp. 244–255, 2008.

[48] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[49] K. Kalgaonkar, P. Smaragdis, and B. Raj, "Sensor and data systems, audio-assisted cameras and acoustic doppler sensors," in *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 2–3.

[50] J. Georgiou, P. Pouliquen, and A. Cassidy, "A multimodal-corpus data collection system for cognitive acoustic scene analysis," in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, 2011, pp. 1 – 6.

[51] M. Piccardi, "Background subtraction techniques: a review," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 4. Ieee, 2004, pp. 3099–3104.

[52] B. P. L. Lo and S. A. Velastin, "Automatic congestion detection system for underground platforms," in *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on.* IEEE, 2001, pp. 158–161.

[53] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1337–1342, 2003.

[54] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis." *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 2, pp. 162–77, Feb. 2005. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1374864

[55] D. Tan, K. Huang, S. Yu, and T. Tan, "Uniprojective features for gait recognition," *Advances in Biometrics*, pp. 673–682, 2007.

[56] D. Tan, S. Yu, K. Huang, and T. Tan, "Walker recognition without gait cycle estimation," *Advances in Biometrics*, pp. 222–231, 2007.

[57] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II—-842.

[58] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *Image Processing, IEEE Transactions on*, vol. 12, no. 9, pp. 1120–1131, 2003.

[59] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-based Gait Recognition for Human Identification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1505–1518, 2003.

[60] S. Yu, D. Tan, K. Huang, and T. Tan, "Reducing the effect of noise on human contour in gait recognition," *Advances in Biometrics*, pp. 338–346, 2007.

[61] C. R. Wren, A. Azarbayejani, and T. Darrell, "Pfinder: real-time tracking of the human body," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 780–785, 1997. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=598236

[62] L. Bombini, P. Cerri, and P. Medici, "Radar-vision fusion for vehicle detection," in *Proceedings of International Workshop on Intelligent Transportation*, 2006, pp. 65–70. [Online]. Available: http://www.ce.unipr.it/people/bertozzi/publications/cr/wit2006-crf-radar.pdf

[63] A. Roy, N. Gale, and L. Hong, "Fusion of Doppler radar and video information for automated traffic surveillance," in *Information Fusion, 2009. FUSION '09. 12th International Conference on*, 2009, pp. 1989–1996.

[64] P. Pathirana, A. Lim, A. Savkin, and P. Hodgson, "Robust video/ultrasonic fusion-based estimation for automotive applications," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 4, pp. 1631–1639, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4273748

[65] R. Kažys, L. Jakevičius, and L. Mažeika, "Beamforming by means of 2D phased ultrasonic arrays," *Ultragarsas,(Ultrasound), Kaunas, Technologija*, no. 1, p. 29, 1998.

[66] C. Pohl and J. L. V. A. N. Genderen, "Multisensor image fusion in remote sensing : concepts , methods and applications," vol. 19, no. 5, pp. 823–854, 1998.

[67] R. Duda, E. Hart, and D. Stock, *Pattern Classification second edition*, 2001.