

# Understanding Pedestrians Behaviors and Properties Using 2D LiDARs based on Gait

(2次元 LiDAR を用いた歩容に基づく  
歩行者の行動と属性の理解)



Mahmudul Hasan

マハムドウル ハサン  
(19DM054)

Major Dept. Information and Computer Sciences  
Graduate School of Science and Engineering  
Saitama University, Japan

***Supervisor:*** Professor Yoshinori Kobayashi

A thesis submitted in fulfillment of the requirements for the degree of  
*Doctor of Philosophy*

September 2022

*To my grandparents, parents, wife, and children  
for their unconditional love and affection*

# Acknowledgements

This journey would not have been possible without the support of many people. Thank you to my supervisor, *Professor Yoshinori Kobayashi*, for your patience, guidance, and support. I have significantly benefited from your wealth of knowledge and meticulous editing. I am incredibly grateful that you took me on as a student and continued to have faith in me over the years. I want to express gratitude to *Professor Yoshinori Kuno* for his treasured support, essential to starting my Ph.D. research. I also thank Assoc. Prof. *Hisato Fukuda*, and Asst. Prof. *Ryota Suzuki* for their mentorship and technical support in my works.

The commendable thesis evaluation committee members, *Professor Tetsuya Shimamura*, *Professor Takashi Kumoro*, *Professor Takuya Azumi*, and my supervisor, provided several scholarly comments and helpful suggestions. I am so grateful to the committee for having those to improve my dissertation.

I want to thank my friends (*Matiqul* and *Kamal*), lab mates, and research team - *Junichi Hanawa* and *Riku Goto*, for a cherished time spent together in the lab and social settings. My appreciation also goes to all the participants in the experiments from Japan, the Philippines, and Bangladesh for helping me create a robust dataset.

I am ever grateful to Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for the scholarship (Monbukagakusho) that made it possible to continue my uninterpretable studies. I also want to thank the government of Bangladesh and Comilla University for granting me study leave during this tenure.

Finally, I would like to express my gratitude to my parents and sister, who envisaged seeing me as a doctor since the beginning of my life. The adorable faces of my children (*Mimar Hasan* and *Manaira Hasan*) recharged me fully every day. Thank you to my lovely wife, *Sultana M. Hasan*, for your tremendous sacrifices for me to pursue a Ph.D. degree. Without their enormous understanding and encouragement over the past few years, it would be impossible for me to complete my study.

# Abstract

Nowadays, criminal offenses increased rapidly, which intended people to deploy a surveillance system on the premises. An effective monitoring system dealing with many rigorous factors might also be solved in real-time. Most of the automatic security measuring techniques are based on video cameras. Here the issue of privacy arises parallelly with security. None of these can be ignored in a civilized society. Anyone may oppose being exposed to a video camera. Then, how to ensure the recording of suspicious activities. Even safety monitoring is essential for the elderly, disabled, or children. Neither a video camera nor a three-dimensional sensor provides proper privacy in public monitoring activity. Besides, there is always a possibility of leaking videos. Furthermore, video cameras are also prejudiced with their capturing resolutions and even lighting conditions. A two-dimensional LiDAR (Light Detection And Ranging) sensor could be the best alternative for cameras that can deal with all these shortcomings. It has a wide angular range of 270-degree and a scan covering about 30 meters. LiDAR sensors only measure the distance by timing the travel of the light pulse. Though there is no visual image created by it, there is no concern for privacy. Any lighting condition is appropriate for LiDAR scanning. On the other hand, LiDAR data processing is eventually straightforward, cheap, and easy to process in real-time. We used a 2D LiDAR sensor for our study and tried to determine a person's behavior in different segmentations.

People's behavior understanding is essential in every surveillance system. In this dissertation, we categorized this into four phases. The first one is person tracking. Following an individual in a public or private event is essential to determine meaningful events and suspicious activities. We grabbed the question of undisclosed people tracking where any recognition is deliberately omitted to maintain the privacy of spectators. In this study, an ankle level LiDAR positioning was a hilarious enhancement for sensing appropriate trajectories of persons. The time-series data Motion History Image (MHI) was created to plot positional values on the image and leave past images as afterimages. These MHI were used as input to the system. The conventional density-based algorithm DBSCAN was used to determine the clusters for an individual moving in the image. Counting the number of pixels in the ankle image Kalman Filter determines the person's path. An in-house dataset appraises the method in different

gestures, but the performances are yet to be improved with prior model extension. We found some shortcomings in density-based algorithms in a rigorous analysis with LiDAR data density, especially in highly dense data and occlusion. Our modified algorithms: EDBSCAN and EOPTICS, effectively resolved the issues and are compatible with LiDAR data density. Their viable precisions triumph over traditional DBSCAN and OPTICS audaciously.

Secondly, we emphasized Person Property Estimation (PPE) as a people's behavior using compulsive data. Identifying individual measurements of a person (Height, Age, etc.) is essential in recruiting and clothing or any commercial agencies. Besides cameras, LiDAR-based estimation is a solemnly unknown study. In this research, multiple LiDARs were set in Multi-Layer and Multi-Angle positions to get much sensing detail in LiDAR images. Discovering hidden patterns of data and ensuring supremacy in terms of accuracy Deep Neural Network (DNN) model were applied to find the properties. A composite dataset was developed to execute comprehensive experiments with the utmost accuracy—augmented training data, diversified trials, and cross-checking profoundly enhanced real-time appreciations. Thirdly, identifying an individual in different positions within the scope was considered in this study carefully. We analyzed Gait; a pattern performed by walking the individuals to facilitate it. A new modality with gait data analysis by ankle level 2D LiDAR shows versatile usage of the sensors. A comprehensive dataset KoLaSU (Kobayashi Laboratory of Saitama University), was developed with numerous international participants. This dataset might be a reference for 2D LiDAR-based behavior analysis with unbiased occurrences and multiple properties. Cross-validation datasets were included to enhance the diversity. A pre-trained ResNet model was used to train the dataset and classify individuals according to test occurrences. Impressive identification results ensure the credibility of the model.

Finally, our focus is on group recognition to understand the behavior of a group of people moving together. Various applications (i.e., museum-guided robots, tourist trackers, etc.) can use this system without disclosing individuals' identities. This dissertation introduced a novel approach to finding human behaviors using only 2D LiDAR sensors, which empirically established its integrity with insightful legacies.

**Keywords:** *2D LiDAR, People Tracking, Ankle level Tracking, Density-based Clustering, Person Property Estimation, People Behavior Understanding, Person Identification, Group Recognition, Deep Neural Network, ResNet.*

# Table of Contents

<b>Acknowledgements</b> .....	<b>III</b>
<b>Abstract</b> .....	<b>IV</b>
<b>List of Figures</b> .....	<b>X</b>
<b>List of Tables</b> .....	<b>XII</b>
<b>Chapter 1</b> .....	<b>1</b>
<b>Introduction</b> .....	<b>1</b>
1.1 Motivation.....	1
1.2 People Behavior Understanding .....	2
1.2.1 What is Person Tracking?.....	3
1.2.2 What is Person Property Estimation (PPE)?.....	3
1.2.3 Gait to Identify Individuals.....	3
1.2.4 Pedestrian Group Recognition.....	4
1.3 2D LiDAR.....	4
1.4 Challenges of LiDAR-based Behavior Understanding .....	5
1.5 Precise Objectives .....	7
1.6 Research Contribution .....	7
1.7 Thesis overview.....	9
<b>Chapter 2</b> .....	<b>11</b>
<b>LiDAR-based Detection, Tracking, and Property Estimation: A State-of-the-art Review</b> ....	<b>11</b>
2.1 Introduction .....	11
2.2 Overview .....	13
2.2.1 Video camera-based Initiations .....	14
2.2.2 Sensor-based Initiations.....	15
2.3 LiDAR-based Tracking and detection approaches .....	17
2.3.1 3D LiDAR-based approaches .....	17
2.3.2 2D LiDAR-based approaches .....	22
2.4 Modern Research trends along with LiDAR sensors.....	25
2.4.1 Initiatives based on 3D LiDAR .....	25
2.4.2 Initiatives based on 2D LiDAR .....	27
2.5 Conclusion.....	27
<b>Chapter 3</b> .....	<b>28</b>
<b>Tracking People using Ankle-Level 2D LiDAR for Gait Analysis</b> .....	<b>28</b>
3.1 Introduction .....	28
3.2 Proposed Method .....	29

3.3 Experimental Results .....	31
3.4 Conclusion.....	33
<b>Chapter 4.....</b>	<b>34</b>
<b>Person Tracking Using Ankle-Level LiDAR Based on Enhanced DBSCAN and OPTICS .....</b>	<b>34</b>
4.1 Introduction .....	34
4.2 Methodology.....	37
4.2.1 Sensing .....	39
4.2.2 Clustering .....	40
4.2.3 Tracking.....	54
4.3 Result and Discussion.....	55
4.4. Conclusion.....	60
<b>Chapter 5.....</b>	<b>61</b>
<b>Person Property Estimation based on 2D LiDAR Data using Deep Neural Network .....</b>	<b>61</b>
5.1 Introduction .....	61
5.2 Proposed Method .....	63
5.2.1 Dataset Preparation .....	63
5.2.2 Person Property Estimation .....	64
5.3 Results and Discussion .....	68
5.4 Conclusion.....	71
<b>Chapter 6.....</b>	<b>73</b>
<b>Person Identification by Evaluating Gait using 2D LiDAR and Deep Neural Network .....</b>	<b>73</b>
6.1 Introduction .....	73
6.2 Proposed Method .....	76
6.2.1 Integrated System Overview.....	76
6.2.2 Gait-based Person Identification .....	77
6.3 Experiments and Discussion .....	81
6.3.1: Gait-based person identification .....	81
6.3.2 Comparison of different data types.....	84
6.3.3 Comparison with contemporary studies .....	87
6.4: Conclusion.....	88
<b>Chapter 7.....</b>	<b>89</b>
<b>Conclusion and Future Work .....</b>	<b>89</b>
7.1 Conclusion.....	89
7.2 Future Work.....	90
<b>Publication List .....</b>	<b>91</b>
Published Journal:.....	91



Under Review:.....	91
<b>References.....</b>	<b>92</b>

# List of Figures

Fig. 1. 1 Concerning issues of video surveillance system.....	1
Fig. 1. 2 Automatic monitoring systems .....	3
Fig. 1. 3 People Behavior Estimation .....	4
Fig. 1. 4 HOKUYO 2D LiDAR and its scanning area.....	5
Fig. 1. 5 Showing challenges with LiDAR-based analysis .....	7
Fig. 1.6 Overview of the Research .....	9
Fig. 2. 1 Block diagram of Person Identification and Property Estimation Research trends [207].....	13
Fig. 2. 2 LiDAR-based person behavior tracking [207].....	19
Fig. 2. 3 Overhead sensor setup for person tracking [207].....	20
Fig. 2. 4 Multi-layer LiDAR setup for person sensing [207].....	23
Fig. 3. 1 (a) Lidar sensor placed on ankle level and getting data (b) Persons’ ankle movements, and standing positions are showed and (c) Tracking people with a handler marker sign ‘ ’ and shows his/her moving directions. [93].....	30
Fig. 3. 2 (a) Ankles are in different positions, and it goes in the overlapping position, (b) frame-by-frame detection [93].....	31
Fig. 3. 3 Kalman Filter-based movements in different circumstances; Upper Row: Ankle Positions on the frame; Lower Row: Direction of the movements [93].....	32
Fig. 4. 1 Block Diagram of LiDAR-based Person Tracking [94] .....	36
Fig. 4. 2 (a) Different distance values of ankles (b) LiDAR-based ankle positions on the frame (c) Considering close ankles as a person (d) Ankles in different orientations make a cluster. [94] .....	38
Fig. 4. 3 Image sensing using LiDAR sensor and preprocessing [94].....	39
Fig. 4. 4 DBSCAN mechanism .....	41
Fig. 4. 5 The uppermost images show an ankle position on the frame—lowermost images pointing persons with the handler [94] .....	43
Fig. 4. 6 Enhanced DBSCAN (EDBSCAN) mechanism [94] .....	44
Fig. 4. 7 Reduced Radius of EDBSCAN [94] .....	45
Fig. 4. 8 Anomaly in tracking without using maxPts [94].....	45
Fig. 4. 9 (a) Core Distance, (b) Reachability Distance.....	47
Fig. 4. 10 Principle of Traditional OPTICS.....	48
Fig. 4. 11 Sample run of EOPTICS with varied reachability plot.....	50

Fig. 4. 12 EOPTICS with LiDAR data; a) Ankle Positions in front of the sensor, b) Kalman filter-based tracking [94] .....	52
Fig. 4. 13 OPTICS; a) one frame later, ankle positions, b) Corresponding Kalman based tracking [94] .....	53
Fig. 4. 14 EOPTICS, a) ankle positions, b) Kalman filter-based tracking [94] .....	54
Fig. 4. 15 Flow diagram of UKF based person tracking .....	55
Fig. 4. 16 EDBSCAN based Group Peoples tracking with LiDAR sensor [94] .....	55
Fig. 4. 17 EOPTICS based individual tracking with LiDAR sensor [94] .....	56
Fig. 4. 18 Tracking and occlusion handling, a frame-by-frame analysis [94] .....	57
Fig. 5. 1 Block Diagram of the Proposed System [106] .....	63
Fig. 5. 2 (a) Motion history images (Color Image and grayscale) (b) Image Dataset [106].....	64
Fig. 5. 3 Experimental setup and person walking in front of LiDAR sensors [106] .....	65
Fig. 5. 4 DNN based height estimation through 2D LiDAR [106] .....	66
Fig. 5. 5 ResNet structure for the application [106] .....	67
Fig. 6. 1 A system overview of 2D Lidar-based Estimation .....	75
Fig. 6. 2 Creation of Motion History Image.....	76
Fig. 6. 3 LiDAR-based Person Tracking, Property Estimation, and Recognition.....	77
Fig. 6. 4 Experimental setup for Person Identification .....	78
Fig. 6. 5 KoLaSU, Two Persons' data: Upper one is MHI and lowers one is the corresponding pose.....	80
Fig. 6. 6 Gait based Person Identification .....	81
Fig. 6. 7 Cross Validation: Gait Performance test with Cross-Data.....	85
Fig. 6. 8 Combined dataset performance analysis.....	87
Fig. 6. 9 Contemporary studies with state-of-the-art technologies .....	88

# List of Tables

Table 2. 1 Performance analysis of different LiDAR sensors used for 3D navigation [207] ....	21
Table 2. 2 Comparative analysis of LiDAR-based object detection datasets [207] .....	22
Table 2. 3 Comparison with existing Approaches and Precisions [94] .....	25
Table 3. 1 Experimental data and its performance [93] .....	32
Table 4. 1 EDBSCAN Algorithm .....	46
Table 4. 2 EOPTICS Algorithm .....	51
Table 4. 3 Experimented results for EDBSCAN and EOPTICS [94].....	58
Table 4. 4 Comparison of EDBSCAN and EOPTICS with DBSCAN and OPTICS 94] .....	59
Table 4. 5 Comparison with existing Approaches and Precisions [94] .....	59
Table 5. 1 Confusion Matrix of resnet18 based height estimation [106] .....	69
Table 5. 2 Confusion matrix of three classes of height estimation and age estimation [106]	70
Table 5. 3 Overall system performance with ResNet18 network [106] .....	71
Table 6. 1 Gait-based person identification on different parameters.....	83
Table 6. 2 Performance test with different DNN model.....	85

# Chapter 1

## Introduction

### 1.1 Motivation

Many initiatives have been taken to address security and risks in recent times. Outside of personal protection, automated security is becoming increasingly popular. Camera-based surveillance is on the rise and has become an integral part of almost every organization. On the other hand, people's privacy is declining. Security surveillance cameras in every place are complicating the lives of ordinary people. Low and ultra-light systems reduce the camera's ability to take pictures. In addition, the installation and operation of security cameras are costly. In addition, various natural disasters, such as fog, smoke, etc., are significant obstacles to taking pictures. Fig. 1.1 shows conventional security, privacy, and obstruction in a video-based surveillance system.

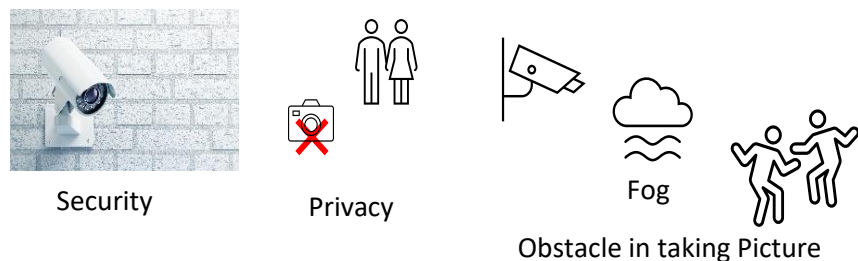


Fig. 1. 1 Concerning issues of video surveillance system

We were looking for an alternative system that would, at the same time, ensure human security and protect privacy. Many sensors have already started to be used as an alternative

to the camera. 3D LiDAR is one of them. Through this, the three-dimensional shape of an object emerges, which is currently being implemented through autonomous driving. But it is relatively expensive and does not fully protect privacy. In this study, we used a 2D LiDAR to guide the movement of people. Using a 2D LiDAR was a very challenging task. With the help of these, only two-dimensional coordinates of distance can be obtained. We created images from it, which we later used to track and identify the person. Its use ensures security as well as privacy. The LiDAR sensor can be used in any lighting or disaster also.

## **1.2 People Behavior Understanding**

People's behavior is an activity that occurs during a person's physical, mental, and social engagements in his life. The way of talking, moving, walking, acting, and approaching everything belongs to his behavior. This study generalized our domain into four comprehensive activities of a person as people behave. Firstly, People Tracking is the activity of determining meaningful and suspicious events. Secondly, People Property Estimation (PPE) is another behavior that facilitates more detail of a person being tracked. Thirdly, analyzing Gait properties to detect individuals in different locations within the scope implies People Identification. Finally, we considered Group Recognition through ankle-level 2D LiDAR data to determine the behavior of a group of people in an event. This behavior understanding decoratively supports many benefits over established video-enabled systems.

It is essential to understand people's behavior to ensure safety and security. It is necessary to keep up to date with the current information on the surroundings to enable a security system. Security hazards and safety measurements depend on people's behaviors also. In a nursing home or an elderly monitoring system, even physically challenged people or child monitoring, etc., are startling demands nowadays. Here neither individual recognition nor criminal identification is required. A LiDAR-based system fits the model in these applications only for understanding the observers' behaviors. Service robots in museums, rescue robots, tourist trackers, etc., applications can be excellent for LiDAR-based deployments. Fig. 1.2 shows some practical usage of the monitoring system.



Fig. 1. 2 Automatic monitoring systems

### 1.2.1 What is Person Tracking?

Person Tracking is a mechanism of recognizing people in a video and interpreting them as a set of paths with great accuracy. The term detection is relatively changed in continuous identifications in image sequences. People tracking refers to locating pedestrians in complex road scenes and their walking ways across the entire video. Fig. 1.3(a) shows machine-based tracking of individuals with an ID.

### 1.2.2 What is Person Property Estimation (PPE)?

Identifying individual measurements of a person from a video refers to PPE. Many properties could be estimated from a person's video. Height, age, gender, region, etc., properties are commonly measured properties from video-based monitoring systems. In our study, we never dealt with direct videos from cameras. We created our image set from ankle-level LiDAR data and focused on estimating height and age—Fig. 1.3(b) showing symbolic estimations of person properties.

### 1.2.3 Gait to Identify Individuals

Many biometric features are used to identify a person in different circumstances. Most biometric recognition systems are designed and require close contact between the sensor and observer. Gait is a pattern that performs by walking by the individual to identify him based on his way of walking. There is no necessity for close contact with the sensor in this model. Nowadays, Gait is known for the valuable evidence in court. A new modality with gait analysis is introduced: Ankle level 2D LiDAR-based person tracking and recognition. Fig. 1.3(c) shows the walking style of a pedestrian.

### 1.2.4 Pedestrian Group Recognition

People used to move in a group, especially when they visit. Some commonly visited places, i.e., museums, shopping malls, tourist spots, etc., sites provide some guided robots or understand an individual group's behaviors. It is primarily necessary to identify them in the video. Group people recognition is relatively challenging especially when only a 2D LiDAR sensor is used to get their moving trajectories. This study rigorously showed the way of recognition. Fig 1.3 (d) shows a group of people moving in front of the LiDARs sensor.

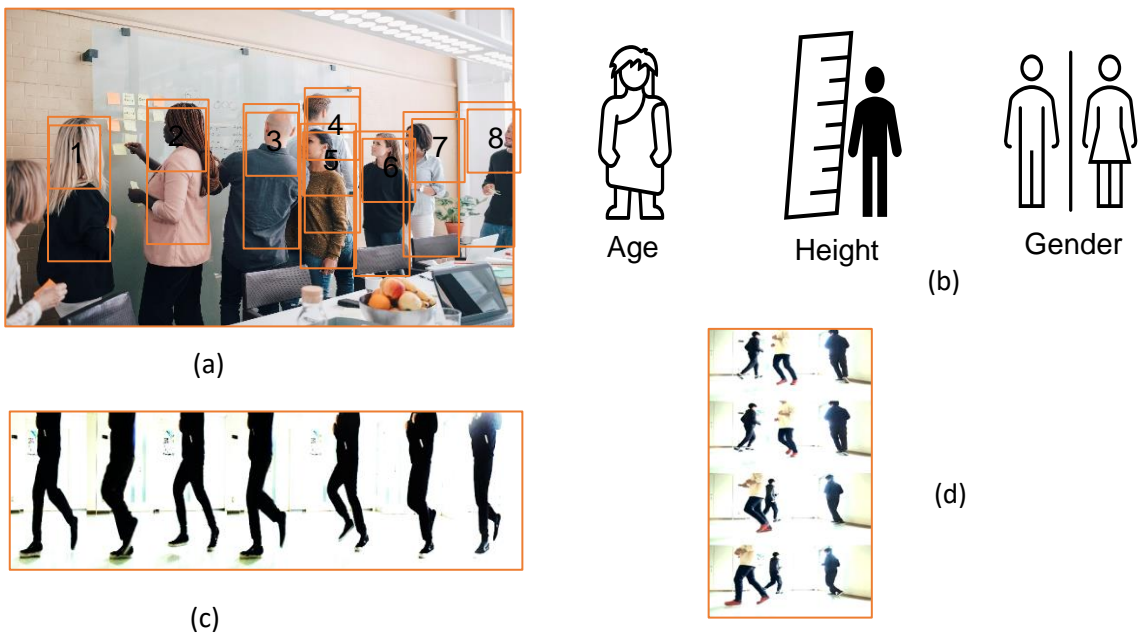


Fig. 1. 3 People Behavior Estimation

### 1.3 2D LiDAR

LiDAR refers to Light Detection and Ranging. An eye-safe technology emits laser light to create a demographic representation with distance values of the surveyed environment through machines. This technology is considered an unparallel innovation in finding the dimension and depth of objects and is impressively faster in capturing than cameras or even RADAR. Today's modern manufacturing, innovative infrastructure, industry, automotive, trucking, robotics, etc., widely implemented LiDAR-based applications to improve the



credibility of production or maintenance. Recent era, autonomous navigation depends on a LiDAR-based navigation system. There are two types of LiDAR sensors: three-dimensional (3D) and two-dimensional (2D). In this dissertation, we considered 2D LiDAR sensors for our experiments. It only collects data on the X and Y axis, which is cheap and easy to process. For our study, we used HOKUYO UTM-30LX 2D LiDAR, whose angular range is 270 degree and the distance coverage range are thirty meters. Fig. 1.4 shows our used LiDAR and its covering ranges.

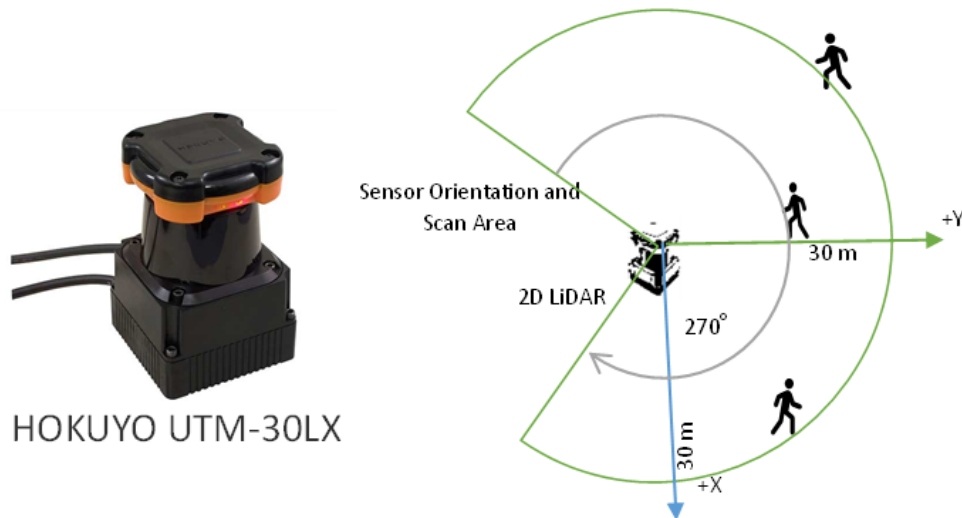


Fig. 1. 4 HOKUYO 2D LiDAR and its scanning area

## 1.4 Challenges of LiDAR-based Behavior Understanding

Object detection, Person Tracking, and Person Property Estimation (PPE) are identical innovation areas trying to improve their accuracy in different parameters to fit various real applications. For many years, so much research has been done in these fields. Many scientists also used many more techniques and algorithms. In general, challenges come to comply with the basic concerns: system accuracy, computational cost, sustainability, and robustness. Implying this novel idea of LiDAR-based monitoring, we were enthusiastically impressed by its precision—later chapters show these explicitly. There is not much installation cost with LiDAR sensors, and even 2D LiDARs are cheap to install, which keeps the cost minimal. LiDARs credentials are proven in terms of capacity and processing.

These features ensure sustainable surveillance with LiDAR only. Wide covering range and distance and easy and faster-capturing ability ensure the robustness of LiDAR sensors. Moreover, some technical challenges also arrive when 2D LiDAR is used in our model.

**LiDAR positioning:** Finding the appropriate position of the LiDAR sensor was a significant concern in this study. We experimented with different experiments, found many ideas, and experimented with results. Finally, we placed our LiDAR sensor in ankle positions, ensuring accurate data acquisition and best tracking through ankle movements.

**Determining person from LiDAR data:** It is easy to distinguish person and objects in a camera image, but LiDAR images have moderate information to decide. We applied various density-based clustering rather than conventional algorithms to ensure maximum accuracy.

**Multiple sensor combinations:** This research implemented people tracking through a single LiDAR-based architecture. But further studies for PPE, Identification or group recognition applications used multi-LiDAR design. Fusing multiple LiDARs data into a single frame and creating its Motion History Images (MHI) was complex. We meticulously crafted the datasets and performed the required studies there.

**Diversified dataset creation:** To heighten the diversity of the dataset, we invited different nationalities of people to our experiments. Twentynine participants participated, and we created our dataset KoLaSU (Kobayashi Laboratory of Saitama University).

Besides these, we faced many difficulties and challenges in this research but could solve those by discussion, experiments, or even trial and error methods. In Fig 1.5, a comparative analysis of camera-based and LiDAR-based tracking shows. Here, the challenges of LiDAR sensors are placed.

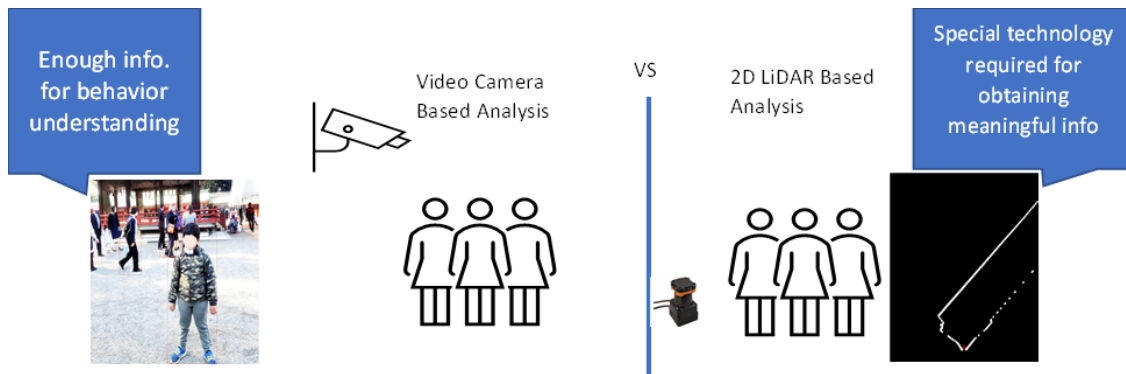


Fig. 1. 5 Showing challenges with LiDAR-based analysis

## 1.5 Precise Objectives

Recent advances in video camera autonomous recognition still face many technical and social issues. Obsessive surveillance systems have become unpopular in several societies. Eventually, an urgency for alternate inspection was raised. There should be some models that relieve privacy concerns but could maintain the surveillance issue. Other problems that a camera-based network should also consider are installation cost, real-time processing, capturing challenges, etc. This research focused on these grounds and empirically proposed a system that best fits the primary objective and effectively addresses similar issues. Our proposed 2D LiDAR-based person tracking using Gait is a novel approach to tracking and identifying persons without disclosing their identity. This LiDAR-based model is cheap, easy to process, has comprehensive range coverage, and does not affect any lighting conditions to get its data.

## 1.6 Research Contribution

This study contributed to four primary segments: Person Tracking, People Property Estimation, Gait Analysis for Person Identifications, and Group People Recognition. Person tracking was challenging for us to initiate 2D LiDAR sensors and create appropriate LiDAR images with meaningful information. Placing the sensor on the plane to capture

accurate data was a heuristic approach, and ankle-level positioning was best. We created Motion History Images (MHI), the time-series data to process in the whole study. A background subtraction method was applied to get only ankle positions on the frame, which enhanced the processability of the data. In the LiDAR images, identifying persons based on their ankle positions was another challenge. We applied density-based clustering to determine a person and Kalman filter for tracking the paths where they moved. Our enhanced study covered the improved performance of density-based clustering by modifying and adding new parameters on DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and OPTICS (Ordering Points To Identify the Clustering Structure), the two well-known algorithms of dense data processing. Our proposed enhanced DBSCAN (EDBSCAN) and enhanced OPTICS (EOPTICS) improve conventional algorithms' performance and solve overlapping and false clustering problems.

Person Property Estimation (PPE) is another approach where we tried to find some characteristics of observed people. Based on LiDAR data, we estimated a person's height and age into two categories. This study helps the monitoring system know much information about pedestrians moving on the premises. We fused multiple LiDAR data in the images to get much information to process accurately. A deep learning model was applied to classify the input data into desired classes. Another impressive study of this study is identifying persons by evaluating Gait and deep neural network (DNN). A comprehensive dataset KoLaSU was developed by considering fourteen combinations of twenty-nine participants from different nationalities. To our knowledge, this might be a new benchmark of 2D LiDAR-based person recognition as no such dataset was created earlier. We are working on group recognition which enables us to determine a group of people moving together. The overall scenario of the contribution of our research is shown in Fig. 1.6.

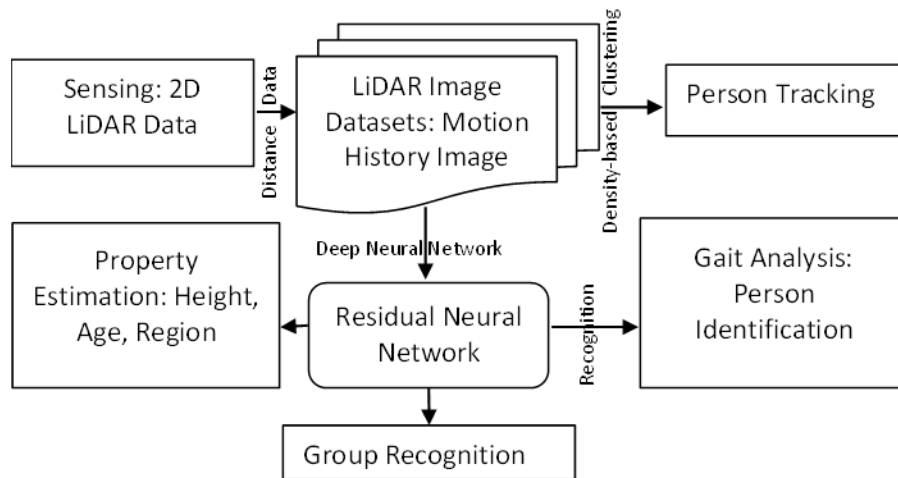


Fig. 1.6 Overview of the Research

## 1.7 Thesis overview

In **chapter 2**, we studied many articles and found enthusiastic outcomes with these sensor setups to understand contemporary technology and its efficacy. We categorized these research articles into video camera-based studies and other sensor-based studies. This chapter covered most of the recent possible sensor-based studies of detection, person tracking, and property estimation except camera (all, RGB, RGB-D, etc.) based learning.

**Chapter 3** discussed the **Person Tracking** system with a LiDAR sensor. This chapter shows experimental setups, data capturing techniques, clustering, and tracking mechanisms. Our fundamental goal of this research is to replace a video camera with a device (2D LiDAR) that significantly preserves the user's privacy, solves the issue of the narrow field of view, and makes the system functional simultaneously.

**Chapter 4** shows a LiDAR-based people tracking technique based on **modified clustering algorithms** to cope with the problems and enhance the performance of our previously developed model. This chapter describes modified density-based algorithms EDBSCAN and EOPTICS for clustering 2D LiDAR data to track people. The experimented results confirmed that our approach significantly improves the accuracy and robustness of people tracking through the experiments.

**Person Property Estimation** is another parameter of our human behavior analysis through the 2D LiDAR sensor, described in **chapter 5**. In this chapter, we figured out a way of estimating a person's property, i.e., height and age, etc., based on LiDAR data. Integrated sensing with multiple LiDARs in multi-angle and multilayer positions enables us to get more detail about a person in the LiDAR images. Applying DNN for property estimation helped us discover hidden patterns in the data. This chapter describes the ways.

A **gait** is a pattern that performs by walking the individual. All studies of gait-based **person identifications** are performed by RGB or RGB-D cameras. Very few studies were done by using 3D LiDAR data. This research conducted a comprehensive exhibition of 2D LiDAR data with a rigorous self-made dataset (**KoLaSU** – Kobayashi Laboratory of Saitama University) and customized residual neural network to identify an individual, clearly described in **chapter 6**.

In conclusion, person tracking, property estimation, and identification are always challenging and sometimes crucial in different circumstances. **Chapter 7** describes the overall summary of the research and its future opportunities.

## **Chapter 2**

# **LiDAR-based Detection, Tracking, and Property Estimation: A State-of-the-art Review**

### **2.1 Introduction**

Biometric measurements ensure superior authentication in practical applications [207]. Most of the analyses are very much dependent on individual sensor setup and modern algorithms. The emergence of cameras made this mature study day after day. Most biometric studies rely closely on video analysis. Computer vision research focuses on the issue very effectively. Innovative algorithms empowered the accuracy and practical usage here. With the vast research area, it is tough to get all comments to contribute above the standard. We found some comprehensive studies performed by video cameras [11], [178]. We tried to figure out other contemporary studies rather than cameras. Some very notable research we found there. Our analysis in this article is to find out most of these.

Early age research focused on people tracking based on video cameras. A broad study was done [1],[18]. They have shown some object tracking systems that are using the multi-camera system. An adaptable classification was done on it. Other initiations also happen during this period. Some sensors were used to do such a tracking system, especially thermal, infrared, etc. Some multimodal research was also done to improve the credibility of the tracking system. Combining RGB and infrared sensors, a person tracking system was also developed [2]. Here enhanced results showed its integrity, showing that sensor-based

tracking is fruitful enough. Some other applications used background modeling that combined visual and thermal data to detect pedestrians [3],[4],[5],[6],[7]. Sometimes research fails due to a stand-alone dataset. Caltech Pedestrian Dataset [8] is one of the larger datasets that provided different dimensional data to analyze a system working on pedestrian detection. Most of the previous studies were done offline, but some online object tracking [9],[10],[12],[181],[182],[183] is a new idea that helped prospective future exploration in this arena. Some studies considered statistical models [15],[16],[17] rather than sensor explorations. Imposing new algorithms. Applying new statistical models, i.e., kernel [13],[14], graph optimization [19], etc., they tried to enhance the tracking credibility that diverse others to develop a property estimation method.

Person detection by visual recognition system is always challenging, especially if rescue robots, supporting robots, or an Autobot system executes independently [125]. Here visual information can be low or hard to receive due to environmental inadequacies. A thermal imaging system can be an excellent alternative to uplift this shortcoming. But most thermal sensor-based studies were concurrently used with RGB cameras [23]. Only thermal sensor-based tracking or property estimation was rarely staged [20],[21]. Two-step-based pedestrian detection by thermal imaging [22] was performed to subdivide the images into many regions where the classification step distinguishes it from non-pedestrians. Thermal sensor-based scanning was a primary alternative sensor setup rather than video cameras for person tracking. Other sensors, i.e., LiDAR, RFID, Infrared, RGB-D, 3D Range sensor, Laser sensor, Radio sensor, etc., were also used for person tracking in different applications. We will consider these studies in this article to get a clear idea of contemporary studies on sensor-based applications.

It is inevitable and parallelly well established that video camera-based object detection, person tracking, or property estimation mechanisms are more effective and suitable. But, some natural calamities, lighting limitations, processing deficiencies, and environmental issues influenced people to think differently to find an alternative of cameras. People introduced parallel sensors set up with the camera to solve these constraints, even individual acquisition techniques rather than cameras. In this study, we will try to find these initiatives that were stand-alone performances besides video camera-based tracking and



property estimation phenomena—Fig. 2.1 shows a block diagram of the contemporary research hierarchy of person tracking and property estimations. We considered only data acquisition mechanisms in this study. Initially, these can be grouped into two classes. Our focus is on sensor-based approaches rather than video camera-based estimations. We categorized sensor-based estimations into four categories. Data acquisition using thermal sensors, LiDAR-based methods, depth sensor-based applications, and the other applications we kept into other groups. Our focus is on LiDAR-based estimations. This study can be subdivided into two classes: 3-dimensional (3D) LiDAR-based and another one is 2-dimensional (2D) LiDAR-based approaches. We will emphasize colored marked notations in this study.

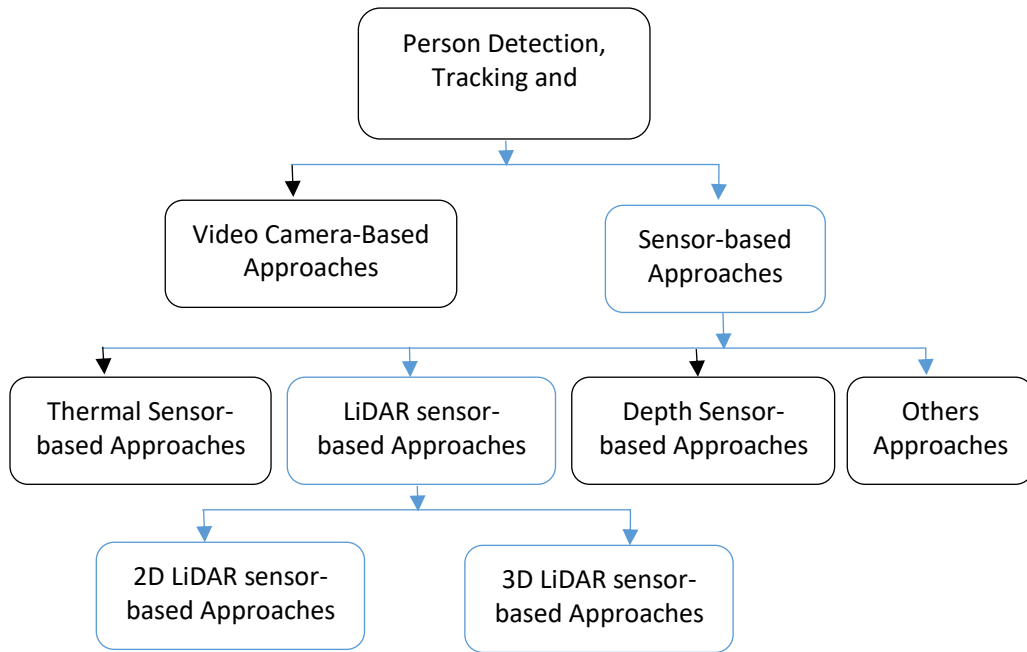


Fig. 2. 1 Block diagram of Person Identification and Property Estimation Research trends [207]

## 2.2 Overview

Machine-made Person identification and different property estimation is not a recent trend. It was started a long ago with the advert developments of the camera technology and applied artificial intelligence aggregated in an intelligent system that is profoundly capable

of functioning. Though this article will concentrate on modern sensor-based advancements, a few initial studies should describe here to recall the initial initiatives that drove us to this intelligent age.

### **2.2.1 Video camera-based Initiations**

A video surveillance system is very demanding nowadays. Military, defense, public security, private organizations, etc., widely install video surveillance systems for safety and security. The focus of modern research on intelligent video surveillance systems lies in object tracking and detection [24]. Here, person tracking is most challenging in video-based object tracking [184], [185], [186], [187], [188], [198], [199] because of different viewpoints in illuminations. Furthermore, this research is complex because of the intraclass's non-deformable shape, occlusion, pose, and size dissimilarity. Another drawback of video-based analysis is its limited field of view (FOV) that initiates the installation of multi-camera networks. Real-time person tracking is another issue that is a big concern of video-based tracking and property estimation. A multicamera-based network must deal with a lot of real-time data that is always tough to handle.

Moreover, some cutting-edge issues: occlusion, pose variations [180], non-rigid deformations, etc., degrade the performance of the real-time system. Typically, person tracking can be categorized into two classes. One is discriminative tracking, and another is generative tracking.

Person tracking algorithms [25] locate a person in every frame in discriminative tracking. Then track in every frame by target associative mechanisms (JPDAF, FNF, MHT, etc.) [26],[27],[28],[29],[30]. In the generative tracking, all persons' positions are calculated with iterative revising of location from preceding frames. Here, Kalman Filter (K.F.) [31],[32] is a very well-known tracking method. Kernel-based Tracking (KT) [33],[34], Particle Filtering (PF) [35],[36], etc. are other ways to do so. Most survey papers are based on object tracking using cameras [37],[38]. Their focus was on object tracking rather than a person. Hou et al. [24] contributed in a distinct way to find out the initiatives on person tracking over camera networks. They distinguished it into two modules: person tracking within a camera and tracking through a non-overlapping camera network. Occlusion is one of the major challenges in object tracking. It always caused misclassifications. There are

some research have done on it also. BY Lee et al. [39] performed a survey on it. They showed recent approaches on occlusion handling. To handle the occlusion overhead camera setup [40],[41],[42],[43] is one of the feasible solutions. Some wide-angle viewing facilities also can be ensured by overhead camera setup. Some analysis was done in this research also [44]. Person reidentification [159],[160],[163],[164],[165],[166],[167] is now very trendy research in computer vision. Imposing deep models and multiscale feature learnings make the system more authentic and exact to the original one. Furthermore, 3-D facial landmarks detection from video [161] is another interesting approach. A deep convolution network is applied for image segmentation [162],[168],[169],[173]. Sometimes extracting common objects from multiple images is essential. Intelligence surveillance is now a widespread area of research. Many more studies have been done in this area [170],[171],[172],[175], but we focused on sensor-based tracking systems except a camera. Thus, we move forward to the next step.

## **2.2.2 Sensor-based Initiations**

Besides video cameras, many sensor-based applications were made to track a person in the real-world environment. Here we discuss these approaches with their prior technology and contributions.

### **2.2.2.1 Infrared sensor-based approaches**

Infrared radiation was invented at the beginning of the eighteenth century by Sir FW Herschel, the inventor of planet Uranus. This radiation is applied for communication, distance measurements, etc. Now a sensor is developed to determine the properties foreground of it. Infrared sensor-based person tracking is new in concept and application. A Bayesian network-based person monitoring [45] by the infrared sensor was depicted in 2005. This application was robust to environmental changes, and remote tracking was possible in close security surveillance. Some researchers used low-resolution thermal infrared (I.R.) sensors [2] to track a person in multimodal sensor setups. A low-cost I.R. sensor made the system computationally and operationally inexpensive and provided high accuracy. This combined system provided enhanced accuracy compared with stand-alone RGB camera-based tracking systems.

An infrared sensor setup is used to track a person, but it could extract human behaviors [46] by collecting long-edge data. They proposed a framework to compute persons' habitual behaviors from infrared sensor data. It is evident that to observe proper human behavior, needed long-term data. It is tough to track all the targets accurately in a low-cost network when targeted persons are more than one. A binary infrared sensor network [47] was proposed by Tao s. et al. to overcome the issue. They developed a soft tracking system and an algorithm for tracking multiple people simultaneously. Soft authentication by the infrared sensor was also used in several studies [48]. However, it is slightly different from complex authentication, which is strictly designed for security and applicable for many people tracking. Though the hard authentication is more accurate, it is relatively expensive (Camera) to install, and a used concern is needed before applying. The thermal infrared sensor was used for person tracking in many applications. Nonparametric background modeling combining visual and thermal spectrum data [3] pedestrian was detected earlier. Some neural network-based person tracking was also performed with I.R. sensors [56]. Several neural network models were tested for the analysis, and 1D-CNN [179] was found most performing over other networks.

#### **2.2.2.2 Thermal camera-based approaches**

Thermal images are captured by measuring reflected, transmitted, and emitted radiations from an object in an area. Because many sources can emit radiation, creating a thermal image from the environment is challenging. Thermal imaging is beneficially associated with point-based imaging. It can detect temperature differences in an area; thus, it is sometimes more applicable during some high-temperature regions. Person tracking precise detection using a thermal camera [49],[50],[51],[52] are now well-established mechanisms. Change detection or background subtraction [53] is a challenge in the thermal imaging system. It always needed a training period which is not existing for all purposes. There are two main ideas [54] of thermal imaging tracking systems. One is to track only warm objects against an excellent background. Another one is tracking in thermal imaging is different than ordinary grayscale imaging. Some research uses low-resolution thermal array sensors in indoor environments [55]. Apart from video cameras, it is also competent to identify persons' heat emissions without light. These applications are commonly used in elderly people support [57] division, quickly identifying them in dark rooms and tracking their

movements. ThermoSense [58], a thermal array sensor network, was invented to find the occupancy in the building. This network can track person movement, an outstanding initiative alternate to cameras. Thermal sensors did some other indoor sensing research. Basu et al. [59] used a thermal array sensor to track a group of people using the background subtraction method. Though many other sensor-based applications were proposed from time to time, we will focus on LiDAR-based research. The following section describes modern LiDAR-based object detection and tracking initiations.

## **2.3 LiDAR-based Tracking and detection approaches**

To find the moving trajectory of objects over time is known as tracking. Tracking and detection are always meant as two separate problems where tracking depends on detection. Object detection can be done individually or even combinedly. Objects or people can be tracked in different ways. It can be short or long-term as well as single or multi-sensor-based tracking. Some systems can track single objects only while others can multiple objects parallelly. In this study, we will try to find all types of tracking initiatives that were performed by LiDAR sensors. Here, LiDAR sensor types are another issue. Some applications that used 3D LiDAR sensors for their research used very few 2D LiDAR sensors. Here we first describe some contemporary studies of 3D LiDAR-based object tracking systems.

### **2.3.1 3D LiDAR-based approaches**

Qian R. et al. [60] proposed an end-to-end framework for the pseudo-LiDAR system in a 3D object detection mechanism. They showed a higher accuracy level in multimodal adjustments of LiDAR sensors with cameras in autonomous driving. Hybrid Voxel Network (HVNet) [61] is another point cloud-based 3D object detection method developed for driving. Their proposed network is one stage 3D object network that uses a 3D LiDAR sensor for object detection. It was evident that most of the recent LiDAR-based research is done for autonomous driving. A 3D LiDAR viewing [62], [177] is parallelly used with the camera set up to make the system more accurate. Person detection and even tracking systems are essential there.

Much recent research has been done on supervised detection. But unsupervised object detection is one of the vital research fields. Tian H. et al. [63] presented an unsupervised object detection with LiDAR clues. LiDAR-Aug [64] is a rendering-based augmentation framework that can detect 3D objects. Data augmentation is a fundamental issue in training neural networks for the expensive labeling cost. Sometimes occlusion problems badly influence the augmentation. To solve these drawbacks, they proposed a 3D LiDAR-based object detection method. This method can be used as a plug-and-play module with other 3D object detection frameworks. Other LiDAR-based 3D detection modules, especially with autonomous driving research and Neural Network (N.N.) computational capabilities, are developed. LiDAR R-CNN [65] is a 3D object detector, a second-stage detector that can enhance present 3D detectors' performance. RSN: Range Sparse Net [66] is another 3D object detector based on a LiDAR sensor, which is light in weight and computationally efficient. It can handle 60 frames per second on a 150 \* 150 m range. To minimize the gap between image and LiDAR-based object detection, a model Pseudo-LiDAR from visual depth estimation [67] is proposed in the object detection mechanism. They artificially convert image-based depth maps into pseudo-LiDAR representations to create LiDAR signals. This overcomes the high expenses of 3D LiDAR sensors.

Cross-modality research is always performed to improve the system's performance. Some research was also done to detect 3D objects and 3D point cloud data of LiDAR sensors [68]. A multimodal fusion for 3D Object detection by taking binocular images and 3D LiDAR images as an input. Some results were shown with KITTI [69] datasets and found sophisticated outcomes there. 3D point cloud semantic segmentation [70] was also introduced in a multi-projection fusion framework. It achieved enhanced segmentation benefits over a single projection system.

Research focuses on person behavior analysis based on 3D LiDAR [71]. Keeping a person in the sensor view, they tried to follow the target person. This application is considered a practical application of caregiver facilities. It could automatically track older adults. Some applications were previously invented to detect a specific person, and some used 2D LiDAR and ESPAR antennas [75]. Using an antenna can see the targeted person smoothly even if they are occluded or out of the viewing angle of the LiDAR. Combining two

methods improves service robot efficiency—Fig. 2.2 shows an overview of person behavior tracking based on a 3D LiDAR sensor.

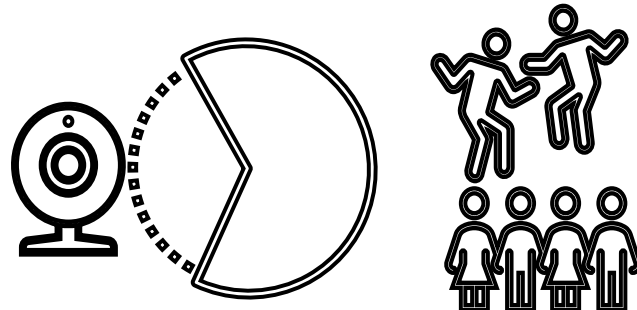


Fig. 2. 2 LiDAR-based person behavior tracking [207]

Some other studies considered multiple 3D range sensors for person tracking in public spaces [76]. Overhead mounted sensor setup to get maximum coverage, and minimum occlusion improved the system's performance in practical use. Structured Light Cameras (Microsoft Kinect1 and Asus Xtion PRO<sup>2</sup>), Time-of-Flight Cameras (Panasonic D-IMager EKL3105), and Multilayer Laser Scanners (Velodyne HDL-32E) sensors were used in this experiment. Fig. 2.3 shows a practical overview of the overhead multisensory person tracking system. These systems can be installed in exhibitions, shopping malls, and even museums to track single or groups of people visiting together. A study [77] was performed to investigate the performance of the PointPillars [78] network, which is designed to interact with LiDAR data. By increasing the LiDAR sweeps, its performance was increased significantly in the 3D object detection mechanism.



Fig. 2. 3 Overhead sensor setup for person tracking [207]

Autonomous driving is now a broad application of 3D LiDAR sensors. To model the actual map in 3 dimensions, LiDAR sensors, visual cameras, and depth sensors are widely used to model the accurate map in three dimensions. These modelings detect the images and track their livelihood trajectories in a 3D map. A comprehensive study was done on the performance of 10 contemporary 3D LiDAR sensors used for autonomous driving [79]. They analyzed the performance of different LiDAR sensors with detailed findings—Velodyne HDL-64, HDL-32, Sick LMS151, Hokuyo UTM-30LX, URG-04LX, Pandar64, Pandar40p, etc. LiDAR sensors were considered to check their performance in different ways—table 2.1 shows the brief comparison of LiDAR sensors' performance [110] in 3D navigation.



Table 2. 1 Performance analysis of different LiDAR sensors used for 3D navigation [207]

3D LiDAR	Visibility	Miss %	Average PPE	% Return	Avg. S.P. K.L. Div.
HDL 32	Weak	2.40	60056	83.4	$1.10 \cdot 10^4$
HDL 64	Weak	0.81	124296	86.3	$2.68 \cdot 10^2$
OS1-16	Very weak	0	14696	91.9	$2.00 \cdot 10^2$
OS1-64	Weak	0	57749	90.2	$1.88 \cdot 10^3$
Pandar40p	Usual	0	67687	94.0	$4.42 \cdot 10^4$
Pandar64	Usual	0	108556	94.2	$3.86 \cdot 10^3$
VLP 16	Usual	6.06	25551	88.7	$3.07 \cdot 10^4$
VLP 32	Usual	2.59	50372	87.4	$2.80 \cdot 10^4$
VLS 128	Excellent	5.93	220668	95.8	$7.10 \cdot 10^2$

3D object detection is the primary key of different intelligent systems [81]. We analyzed many approaches to these. There are so many reviews that have also been done in this field. But LiDAR-based review was minimal, and most of the up-to-date works were not covered in many surveys. Wu y. et al. [80] meticulously crafted the topic and showed very modern methodologies in their study. A complete survey with 3D LiDAR-based object detection and profound neural network-based approaches was covered thoroughly. This review emphasizes the hardware and software revolutions over the periods. Some well-known datasets were analyzed in their study. There are several public datasets available for autonomous driving research. In table 2.2, we have shown some of these. Sometimes wearing acceleration sensors [82] was used for person detection in a natural environment. Multiple depth sensors (3D LiDAR, etc.) were used for position tracking. Finally, calculating acceleration from two separate sources identifies the sensor holder. Some embedded systems were developed with 3D LiDAR sensors for a person or object detection [83]. An object tracker (IMM-UKF-JPDAF based) was designed to detect and track with 3D LiDAR point cloud classification. Service robots are performing simultaneously with the scanning. So online learning framework is badly essential there. Yan z. et al. [84] proposed online learning with 3D LiDAR-based tracking for person detection. An online retaining classifier iteratively learns by robots' provided data over time.

Table 2. 2 Comparative analysis of LiDAR-based object detection datasets [207]

Dataset Name	Released Year	Release Authority	Frames	Classes
A2D2 [111]	2020	Audi	40,000	38
ApolloScape [112]	2020	Apollo	100K	6
Argoverse [113]	2019	ARGO AI	324,000	113
Berkeley [114]	2020	UCB	100M	10
CityScapes [115]	2016	CityScapes	25000	30
Comma2k19 [116]	2019	comma.ai	280h	
Google-Landmarks [117]	2018	Google	2M	30K
Google-Landmarks-v2 [117]	2019	Google	5M	200K
KITTI [69]	2012	KIT	4x83,000	19
LeddarTech PixSet [118]	2021	LeddarTech	29K	97
Level 5 Open Data [119]	2019	Lyft	55K	4
nuScenes [120]	2019	Motional	1.44M	23
Oxford Radar RobotCar [121]	2016	OxfordUni		100
PandaSet [122]	2021	Hesai and Scale AI	48K	28
Waymo Open [123]	2019	Waymo	200000	1000
H3D Honda 3D [124]	2019	Honda	27K	8

There is much more research has been conducted on object detection using 3D LiDAR sensors [126],[127],[128],[129]. This research obviously could detect pedestrians as it could detect other objects. In the meantime, 2D LiDAR-based person detection was successfully developed in mobile robot applications [130], [131], [132], [133], [134], [135], [176]. In the subsequent section, we will discuss 2D LiDAR-based approaches.

### 2.3.2 2D LiDAR-based approaches

We found very few works on 2D LiDAR base person tracking. Because 2D scanning is less informative and it is tough to handle the fundamental data for advanced applications where ultimate accuracy is highly needed, 2D LiDAR-based applications are minimal. Oishi, S. et al. [72] developed a robot that can adaptively assist a precise person based on their behavior. A Finite State Machine (FSM) recognizes a person's actions. Here they used sets of 2D LiDAR sensors for scanning in multiple layers. Fig. 2.4 shows a schematic

diagram of a multi-layer LiDAR setup. It uses 2D-LiDAR sensors in its different layers to get the actual view of the target person in front of it. Some research fused multisensory data and demonstrated their outcome with enhanced accuracy. A video camera and a 2D LiDAR sensor were used parallelly [73] to detect a human and track mobile service robots. A multi-functional algorithm using the sequential Kalman filter implementation to track a person in a realistic environment, a sensor-based leg and camera-based face detection approach was fused. Besides humans, some research was performed to detect only moving objects and track them in a furnished living room [74]. Using only a laser range detector (LiDAR), they tried to detect an object. This system was computationally inexpensive, and power consumption was only 2% of the total.

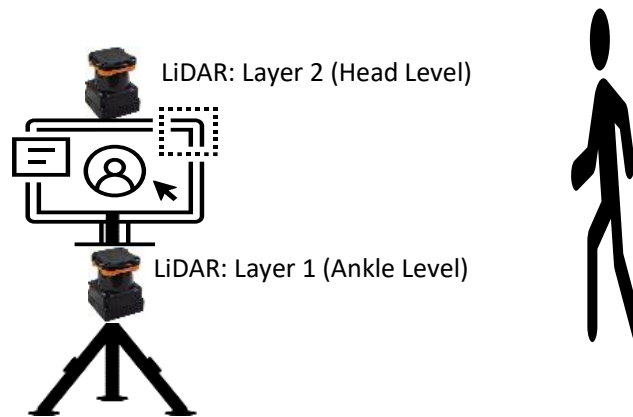


Fig. 2. 4 Multi-layer LiDAR setup for person sensing [207]

There are many onboard LiDAR-based people tracking systems. Tracking only moving legs by a 2D LiDAR sensor is always challenging due to occlusion. On the other hand, 2D sensors are cheap in cost and installation, and their processing is more effortless. PeTra [85], a convolutional neural network (CNN) based on person tracking, was invented with 2D LiDAR sensors only. Its performance is quite good than Leg-Detector (L.D.) [86] in the robot operating system (ROS) [87] environment. Some research also proposed two sensor fusion architectures [88] for pedestrian detection. A 2D LiDAR sensor and a camera were used to get the signals and blended by a trainable fusion method. This combined method also provided better results. Some model-free applications [89] were developed to detect and track moving objects in autonomous driving applications. A 2D laser scanner was used to get the data where most of the applications were 3D scanning related. Many

robotics applications deal with detecting and tracking moving objects (DATMO) [90] problems. A core algorithm was developed to handle 2D range finding and 3D range-finding data. This research was a milestone in object detection and tracking. It could take around 100 objects simultaneously. Other analyses were performed to solve practical issues like traffic injuries and fatality. An obstacle motion tracking [91] by 2D LiDAR scanning module was developed there. To ensure the mobility of robotic navigation, it is always needed robust software which can detect, track, and follow a person automatically. Leigh A. et al. [92] developed a 2D LiDAR and camera-based people following indoor and outdoor mechanisms.

Most contemporary research emphasized LiDAR-based detection, but a parallel camera implementation was performed in most cases. Even though these fusions enhance the system performance, its computational complexities can be ignored. Hasan M. et al. [93] developed a simple 2D LiDAR-based ankle-level person tracking system that used only LiDAR data. They empirically crafted the motion history images (MHI) from numeric lidar values and enhanced their study for gait analysis. A density-based clustering algorithm was used to determine the ankle and the person moving on the plane. A Kalman filter was used for tracking the persons' movements. Here, LiDAR sensor was placed in ankle position, and data are plotted on a blank image with 40 frames per second (fps) rate. A boundary-removing technique was applied to remove unwanted data. A density-based algorithm [174]: DBSCAN was used for clustering the data as an ankle on the image. The distance threshold between ankles determined the ankles as a person.

Finally, a Kalman filter tracking was applied to monitor the person moving on the plane. They enhanced their study in their subsequent analysis, where they developed a tracking mechanism by proposing two new algorithms: EDBSCAN and EOPTICS [94]. A more robust person tracking was presented in this updated study. They modified two well-known density-based algorithms, DBSCAN and OPTICS, with a modified parameter for LiDAR images. Here, clustering could be performed in more sophisticated ways to handle close anomalies and occlusion problems efficiently. Here the overall system performance increased remarkably compared with the previous one. In table 2.3, the most recent clustering approaches are described with recent precessions [94].

Table 2. 3 Comparison with existing Approaches and Precisions [94]

Approach	Cluster Threshold	Precisions
Euclidean Clustering [107]	0.5 m	64.5%
DBSCAN [93]	Adaptive	93.7%
Depth Clustering [108]	10°	39.2%
Run Clustering [109]	Params <sub>SSLR</sub>	51.7%
Online Learning [102]	Adaptive	89.8%
EDBSCAN [94]	Adaptive	93.9%
EOPTICS [94]	Adaptive	96.9%

## 2.4 Modern Research trends along with LiDAR sensors

### 2.4.1 Initiatives based on 3D LiDAR

Much research addressed gender and age classification from video data. But the performance of those research was not adequate. To process much data are challenging, especially on the internet. Deep CNN network-based age and gender classification [95] methods have been introduced in recent initiatives to deal with this problem. A shallow network is proposed to prevent the data overfitting problem and tested on the Adience benchmark, which outperformed. Most of the latest research was a video camera-based approach. Most cameras are installed in top positions in the video surveillance system to get the best viewing angle. By calibrating three parameters: camera height, focal length, and tilting angle, human size can be easily measured by a nonlinear regression model [96]. Human height also could be estimated from a single image [97]. With the empirical developments of deep learning schemes, some research was done to find a persons' height from his simple photos.

Similarly, some research was also initiated to determine a person's size from their video [98]. The motion footage was given as an input with the external force of gravity. Color deep learning and 3D depth conversion [99] was also recently applied to estimate human height. Some research was done with straightforward calculations. Using hand-span

measurements [100] even joint ratio of fingers [101], a persons' height can be estimated approximately.

Nowadays, online learning is more popular and appropriate for detecting people. Yan Z. et al. proposed a 3D LiDAR-based detection method was proposed by Yan Z. et al. [102]. Some limitations prohibit RGB or RGB-D cameras from being outdoors, where LiDAR must be a good alternative. On the other hand, gait recognition is a recent research field that enabled us to recognize a person without close contact with the sensor. Yamada H. et al. [103] thought differently with 3D LiDAR sensor. LiDAR data with the long short-term memory-based network was one of the fundamental research topics. To the best of our knowledge, Benedek C. and his team [104] were the first researchers who initiated gait analysis with LiDAR data. They demonstrated 3D LiDAR-based gait analysis and activity identification in the surveillance system. Tu J. et al. [105] introduced a physically realizable adversarial example for LiDAR-based object detection. The system accuracy was also impressive compared with other initiatives.

#### **2.4.1.1 Deep Learning approaches**

With the emergence development of deep learning algorithms and sophisticated outcomes with these learning techniques, today's research achieves a new milestone. 3D LiDAR-based initiatives based on deep learning can be categorized into two classes, largely: a one-step detector and another one is two-step detector. Based on point cloud or voxels, one-step detector [136],[137],[138],[139] uses backbone network to process. On the other hand, two-step detectors [140], [141], [142] create another bounding-box step to generate enhanced predictions. The deep learning-based simultaneous segmentation and detection method (SSADNet) [143] was another new autonomous driving approach that could recognize driving areas and obstacles. Based on a single neural network, it could perform segmentation and detection simultaneously. In some prior studies, U-Net [145] or FCN (Fully Convolutional Network) [144] was used for semantic segmentation. Eventually, RetinaNet [147], Faster-RCNN [146], or even YOLOv3[148] networks were frequently used for object detection. With the drastic advancements in DNN research PIXOR [149], RangeNet++ [150] even PointSeg [151] networks were used in the point cloud of LiDAR in object detection scenarios. Multi-Object Tracking [190], [194], [195], [196], [197], [200],

[201], [202] is another recent trend. Bilinear LSTM [188], VOT2014 [189], Byte Track [193], Siamese CNN [191],[192], etc. deep learning approaches are frequently using in tracking.

### **2.4.2 Initiatives based on 2D LiDAR**

All the initiatives mentioned above were based on cameras or 3D LiDAR sensors. But Hasan M. et al. [106] estimated person property based on a 2D LiDAR setup. They used a deep neural network (ResNet) to train their data and measured person height, age, etc., from ankle-level LiDAR data. Besides this, there are a few initiatives that were done there in the very early stage [152], [153], [154]. These were manually operated tracking performed sequential scans. Some contemporary works [155], [156] replaced manual operations and improved detection quality in mass. Recent advances [157], [158] rebuilt DNN methods on range data by one-dimensional CNNs. To improve the detection performance of LiDAR scan, DR-SPAAM [134] detector was introduced. All these initiatives focused on more sustainable developments in LiDAR-based detections. The next era is multimodal sensor-based research, where LiDAR will play an enormous role.

## **2.5 Conclusion**

There are so many surveys that have been done on contemporary issues of person tracking and detection. In this exponential growth of the research world, everything ought to become accessible and accurate. LiDAR plays a significant role there. We realized a lack of thorough study with LiDAR sensors. We provided a comprehensive survey on LiDAR-based person detection, tracking precise identification. We focused on LiDAR-based initiatives but considered some recent camera-based detection and tracking studies. We discussed some deep learning studies also. We expect this survey will provide a graceful insight into LiDAR-based object detection, tracking even in recognition.

## **Chapter 3**

# **Tracking People using Ankle-Level 2D LiDAR for Gait Analysis**

### **3.1 Introduction**

Person Tracking (PT) with a machine is a salient field in Human-Computer Interaction (HCI) [93]. Research on person tracking reached utmost approximations in recent years. It involves surface mapping, pointing persons' positions and consequent movements, differentiating these with other properties, and finally projecting these on the desired surface. Time series of individual position data enables us to analyze trajectory for many purposes (e.g., marketing). PT with 2D and 3D cameras play a significant role in practical applications. Real-time PT from a live video makes it more robust and usable in different scenarios. Some statistical models and their efficacies make the PT well accepted by all. Here video camera plays the role of data acquisition, and some of these can perform enhancement of that captured data. Recently, with the development of deep learning-based image processing, people detection and tracking performance using cameras are dramatically improved. However, when we consider using cameras everywhere in daily life, privacy issues cannot be ignored. In addition, some phenomena make it challenging



to use cameras in a particular situation, such as smoke, fog, etc. Furthermore, though low-cost cameras (not only RGB cameras but also RGB-D cameras) are widely used, the computational cost of image processing is not the least. Besides, it may go to the apex using deep learning techniques, especially for many cameras for broad area surveillance.

This research focuses on using a sensor that will not compromise privacy but enhance tracking efficiency. The main issue of tracking people in this sensor setup is how to discriminate individuals from the isolated observation of multiple ankles of multiple persons. To cope with these problems, we propose a new people tracking technique using 2D LiDAR. Cost and real-time computational facility played the key influence behind this. To minimize the occlusions of pedestrians each other, we put 2D LiDAR at an ankle level and range the target area horizontally. Here, we proposed an avant-garde method to classify individuals using the time series data. Individual ankle trajectories were considered for movement detection. The well-known Euclidean nearest neighbor technique calculated distances between ankles. This approach helps determine the cluster of every pair of ankles as a person. We identified walking and running paths even if it goes very fast. This approach calculates the number of frames of every moving object and finally creates a video based on those frames. These videos can be further used for surveillance or any other use. Our method provides accurate and robust tracking when the target is walking or running. Experimental result shows the effectiveness of our proposed method.

### **3.2 Proposed Method**

We introduce a tracking method based on the LiDAR sensor. Our approach works in different environments. We have used a 2D lidar sensor for its lower price and computational effectiveness. We used the HOKUYO UMT30LX Lidar sensor in a plane ground surface for our experiment. We placed our LiDAR at ankle level and collected data in 270-degree directions. When people walk within the LiDAR sensor's range, it collects the actual position of the moving object and its distance from its position. We plotted the ankle positions of persons on the plane and tracked their movements. As shown in Fig. 3.1, a LiDAR is placed at the ankle level of a person and provides visual information to the

corresponding computer. In the first frame, white lines indicate the boundaries of the surface, i.e., walls. We then removed the frame's borders by background subtraction and showed only the ankle position on the surface. The green marked lines indicate the ankle position of a person. If more than one person appears before the sensor, it identifies the person. Distances between pixels of an ankle and between two ankles are measured and used for deciding the number of people moving in front of a sensor. An experimented threshold is used for determining a person in different circumstances. Here we used the Euclidian distance measurement approach for tracking persons.

$$dist((x_a, y_b), (x_b, y_b)) = \sqrt{((x_a - x_b)^2 + (y_a - y_b)^2)} \quad (3.1)$$

Where  $dist$  is a function for calculating the distance between pixel  $(x_a, y_a)$  and  $(x_b, y_b)$  of an ankle's position then between all appeared ankles on the plane.

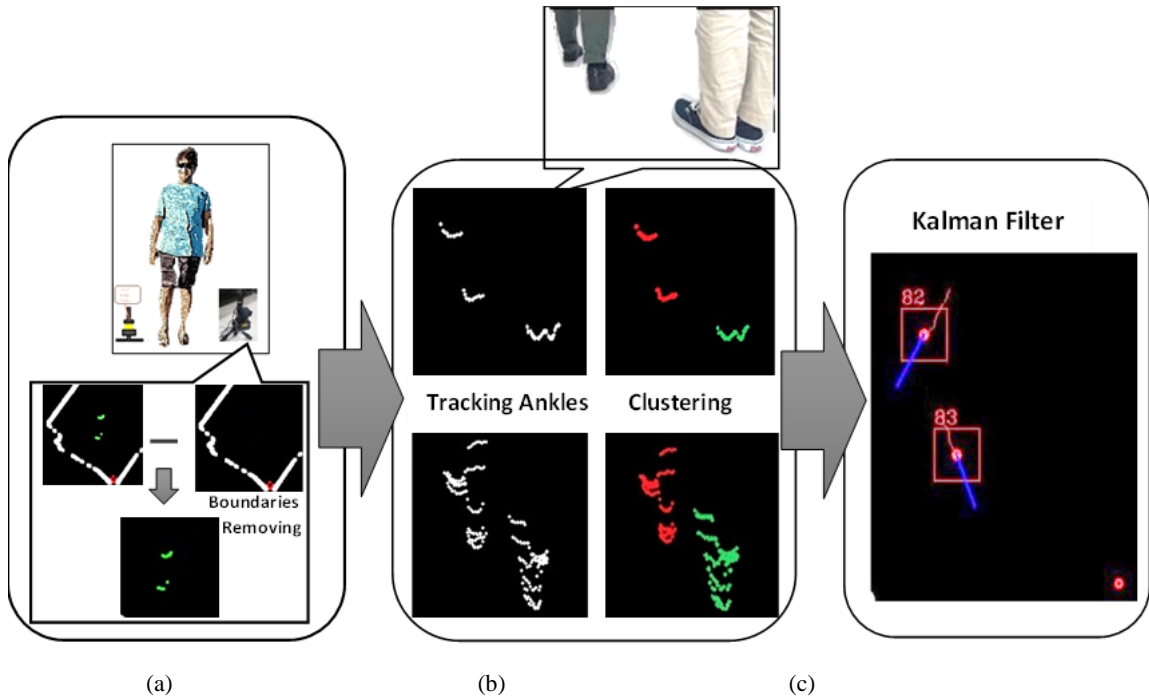


Fig. 3. 1 (a) Lidar sensor placed on ankle level and getting data (b) Persons' ankle movements, and standing positions are showed and (c) Tracking people with a handler marker sign '|' and shows his/her moving directions. [93]

Depending on the positional relationship with the sensor, sometimes one ankle may occlude another one. But this situation does not influence the decision. Using cluster-based

techniques does not assume that both feet are visible. The system only calculates the distance between ankles that appeared in the LiDAR field of view. If no other ankle is found, the system identifies that only one person is walking on the floor. If the distance is larger than the threshold, our system tracks the ankles as a separate person even though one ankle position is found. Thus, the system overcomes the obscure disappearance of feet on the floor. Here Fig. 3.2 shows that our system identifies the ankles with the disappearance. In two frames before, it calculated the ankle distance and found that it is more than the maximum threshold, and there are two persons here. But in one frame, before the distance became smaller, it counted that this was within the range, and only one person was there. Finally, in the current frame, it finds that another occludes one ankle, and without misclassification, it shows that there is one person in the frame.

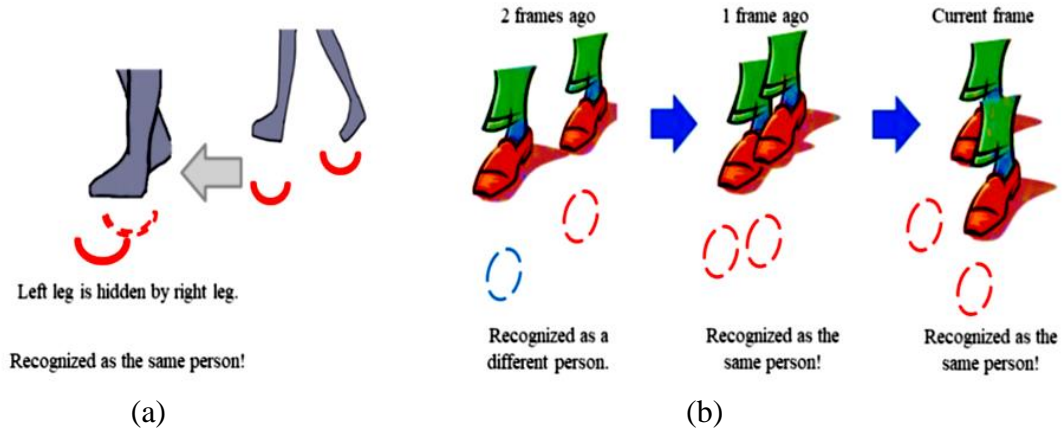


Fig. 3. 2 (a) Ankles are in different positions, and it goes in the overlapping position, (b) frame-by-frame detection [93]

### 3.3 Experimental Results

There are no well-known data sets for 2D LiDAR-based person tracking and gait analysis. We prepared our own data sets for the experiments and appraised our method on this benchmark, with 35 samples with 27 female and eight male participants. The standard contains two scenarios: normal and highly crowded. We ensure that we evaluate the accomplishment of our proposals on the validation data set. We used the Kalman filter to predict for tracking individuals. Here our system can track a person even if only one ankle has appeared on the frame.

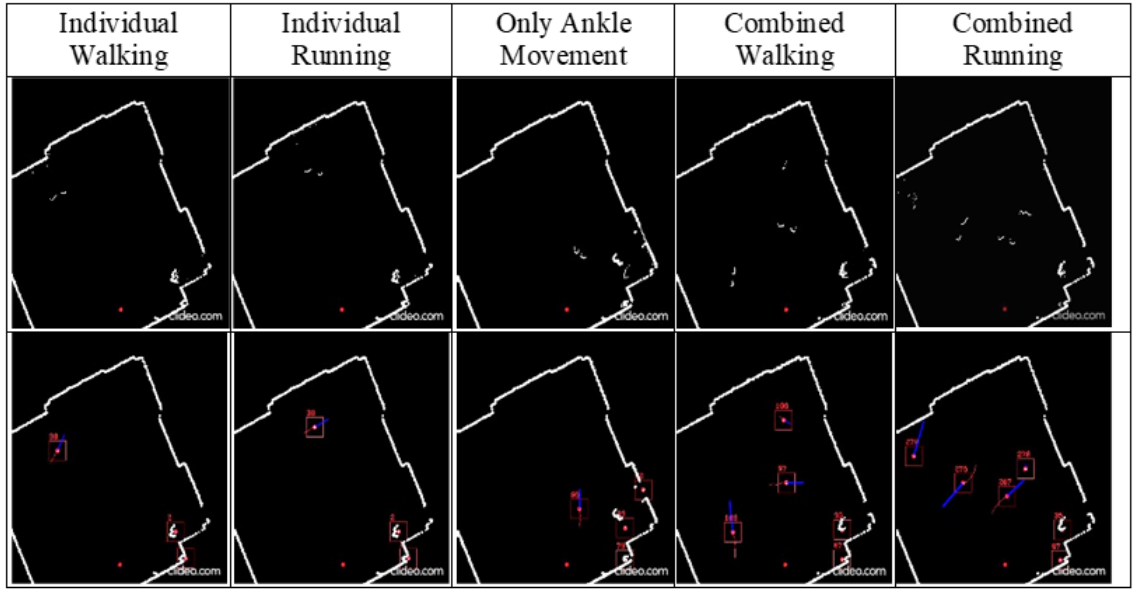


Fig. 3. 3 Kalman Filter-based movements in different circumstances; Upper Row: Ankle Positions on the frame; Lower Row: Direction of the movements [93]

In Fig. 3.3, we see that our system can accurately predict the movements of different persons under different criteria. Individual walking is shown in the first upper frame, and the corresponding lower frame shows that person is going far from LiDAR. In the 2<sup>nd</sup> frame, one person is running in different directions. The 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> frames show other scenarios, and corresponding lower frames show their positions tracked by LiDAR using the Kalman filter technique, respectively.

Table 3. 1 Experimental data and its performance [93]

	Individual Walking		Individual Running		Only Ankle movement		Combined Walking				Combined Running							
Persons/ Frames	4	48 Frms	4	51 Frms	4	43 Frms	2	42 Frms	3	47 Frms	4	43 Frms	2	47 Frms	3	47 Frms	4	42 Frms
Gesture Correctly Identified	4	48 Frms	4	50 Frms	4	40 Frms	2	41 Frms	3	44 Frms	4	39 Frms	2	36 Frms	3	33 Frms	3	29 Frms
Percentage	100 %		98.04 %		93.02 %		97.62 %		93.62 %		90.70 %		76.60 %		70.21 %		69.04 %	

Table 3.1 shows different persons and their recorded videos for our experimental evaluation with different gestures. We categorized our experiments in five different ways. Here individual walking, running, only ankle movement, and combined walking can be tracked with absolute confidence. We have a few observations on a combined running scenario. We considered four people and their captured LiDAR videos for performance evaluation. For validation, we considered about 4 seconds of videos of every person in different gestures, where from each video, we evaluated 42-51 frames. This system performs relatively flat in different running situations from the above table. But compared with other camera-based systems, the performance is impressive. A gait analysis and person height estimation based on ankle movements is also performed on the dataset, and we interestingly found some consequences with walking and running patterns.

### **3.4 Conclusion**

In the cyber world, a person is being tracked every time, everywhere. But when the question is privacy, people want space from the eyes around them, even if it is in a real or virtual world. On the other hand, questions about surveillance cannot be ignored. Here LiDAR plays as a kernel of these issues. Without disclosing a person's identity, our approach tracks a person and identifies their movements. Now our system is ready for commercial use. We will enhance our study to estimate the property of human activities using LiDAR. We are walking on Gait analysis using Ankle level 2D LiDAR. We will integrate the tracking and analysis into one system in the future.

## **Chapter 4**

# **Person Tracking Using Ankle-Level LiDAR Based on Enhanced DBSCAN and OPTICS**

### **4.1 Introduction**

Nowadays, people are under surveillance in their offices, on the roads even at home [94]. To realize a safe and secure society protecting civilians from different offenses, we must keep everything under supervision. On the other hand, people claim their privacy and secrecy because surveillance videos are possibly leaked. Services in public places like airports, stations, museums, and super-shops require people trajectories to provide higher quality services. Our 2D LiDAR-based tracking system can track individuals robustly and accurately. When individual identification is not a concern but only tracking a person is necessary, our system plays a vital role. A wide range of views and low computational costs are also the benefits of 2D LiDAR-based tracking. Using video cameras, 3D LiDAR, and deep neural networks (DNN) makes the system pricey in cost and time. This application is highly applicable for emergencies, especially rescue systems in natural disasters. The 2D LiDAR sensor may be more worthy over RGB/RGB-D cameras in terms of the fast computation and robustness against occlusion.

Video-based person tracking is a widespread technology accepted broadly in different applications. It is necessary to discriminate somehow data points belonging to one person

to track individuals. We replaced cameras with 2D LiDAR and tried to find an alternative. We put our 2D LiDAR at the lower leg (ankle) level and gathered distance information in 270-degree headings. We plot the distance data on the 2D image plane like a top view (we call this image a "LiDAR image"). Using the background subtraction technique, the system can eliminate structural information such as walls or stable objects. Finally, we get only the position data of moving objects drawn on the LiDAR image. Generally, the method takes two steps: the data points are clustered for each ankle, and two ankles are paired for one person. In other words, we measured the distance between the pixels to create clusters to decide on an ankle. The distance between two ankles is then calculated to discriminate between individuals. In this approach, there are several difficulties. Firstly, two-step clustering is a time-consuming process. Processing speed is a significant factor for tracking; Secondly, finding multiple persons' ankles makes it challenging to recognize which two ankles belong to one person in the multiple-person context. In constant, our proposed method can identify individuals using only one-step clustering. Usually, two ankles belonging to one person pass each other closely while walking. Therefore, we make motion history images from 2D LiDAR data (time series of distance data plotted on one image), including the frame closely passing two ankles.

Then we apply one-step clustering to discriminate against individuals. For this data clustering, we previously used traditional clustering algorithms. Here, we have used density-based algorithms. Density-based clustering tremendously influences finding clusters with random contours and easy noise cancellation techniques. Another significant advantage of the density-based algorithm is not to set several clusters previously. It makes clusters based on data density. We modified our system's two well-known density-based algorithms, DBSCAN and OPTICS, by considering the maxPts parameter. This modification makes these algorithms more robust for person tracking. Interestingly, these enhanced algorithms' time complexities are as previous, and performance increased with a considerable amount of time.

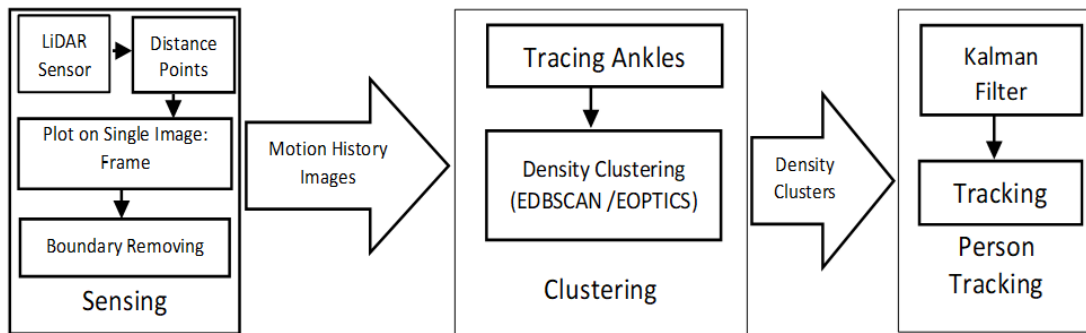


Fig. 4. 1 Block Diagram of LiDAR-based Person Tracking [94]

Figure 4.1 depicts the block diagram of our LiDAR-based person tracking. Three individual sections work together to track a person. The first section is sensing. A 2D LiDAR sensor is used for getting distance data. These distance data of a particular time are plotted on an image (we call this a "frame"). We applied a boundary-removing technique to remove unwanted data from the sensing image. Ankles are traced from motion history images. After getting ankle positions in the frame, we applied our modified clustering algorithm to track a person accurately. Here, the Kalman filter is used for tracking the person. Our primary focus in this study is to find a sophisticated person tracking system using 2D LiDAR data. We included two density-based algorithms separately to increase the system accuracy and showed that enhanced algorithms provide more accurate tracking results. Either EDBSCAN or EOPTICS can be used for our clustering purpose. Both algorithms have a better effect than others, though EOPTICS performs better here. This study will also help find clustering algorithms in time series 2D LiDAR data processing.



## 4.2 Methodology

Tracking a person is always a difficult task. How to track and where to track is a vital question. Tracking based on only LiDAR sensors is more complex than using video cameras. But the mechanism behind LiDAR is relatively easy. Glow a little light from a laser beam at a surface and calculate its total travel time to return to the sender. Hence the distance calculation can be

$$D = \frac{T * V}{2}, \text{ Where } T = \text{Total time of Travel}, V = \text{Velocity of Light} \quad (4.1)$$

Tracking a person is a critical decision about where to place the sensor. Hence, establishing a LiDAR sensor in an ankle position can be the right solution. Because LiDAR can track every ankle, which will help to determine walking paths. Here, it can avoid people's body occlusion and other applications. The mobile robot usually puts LiDAR at ankle level because of obstacle detection.

It is challenging to track a person independently. When multiple people walk on a surface, close ankle positions may misclassify. Figure 4.2(a) shows an apparent description of 2D LiDAR sensing. It provides horizontal distances of moving ankles in X and Y coordinates relative to their position. These distance values are plotted on a frame, and we can get deals, as shown in figure 4.2(b). An individual clustering approach of these distance values is performed to decide on an ankle, and two pairwise ankles make a person chronologically. Figure 4.2 (b) shows that some people are walking on a plane. In figure 4.2 (c), red-marked circles indicate a person where two closed ankles are acted like a person. The fourth figure calculates the cluster differently. This is an anomaly of clustering and occurs misclassification.

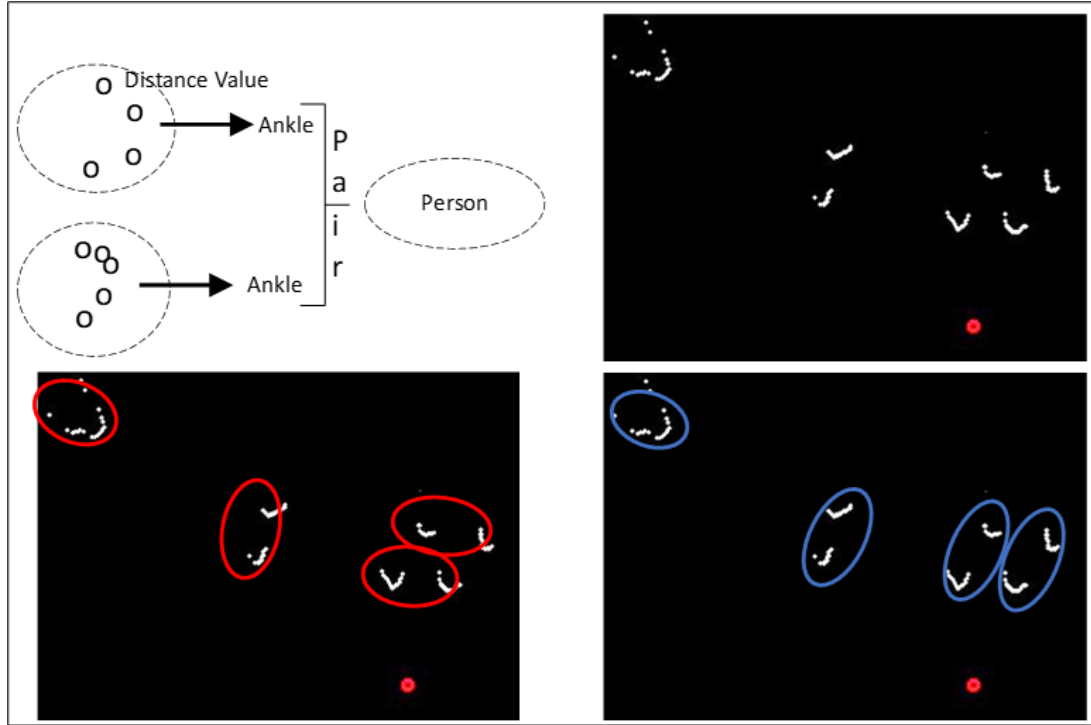


Fig. 4. 2 (a) Different distance values of ankles (b) LiDAR-based ankle positions on the frame (c) Considering close ankles as a person (d) Ankles in different orientations make a cluster. [94]

Sensing data from LiDAR and its better classifications was our primary objective. We repeatedly placed the sensor in different positions from 10 to 200 cm. Usually, human leg lengths vary from 0 to 100 cm. We put our LiDAR sensor at 20 cm height to get these data and significant influences. We see that this height provides the most accurate data for tracking. We faced a downfall in a group movement with fantastic accuracy over single pedestrian movements. Which ankle belongs to which circle? Taking this decision was troublesome indeed. We repeatedly changed cluster thresholds of traditional clustering approaches, but problems persisted. We introduced a state-of-the-art tracking method based on density-based clustering for better approximation and accuracy. Here we can solve wrong clustering scenarios.

## 4.2.1 Sensing

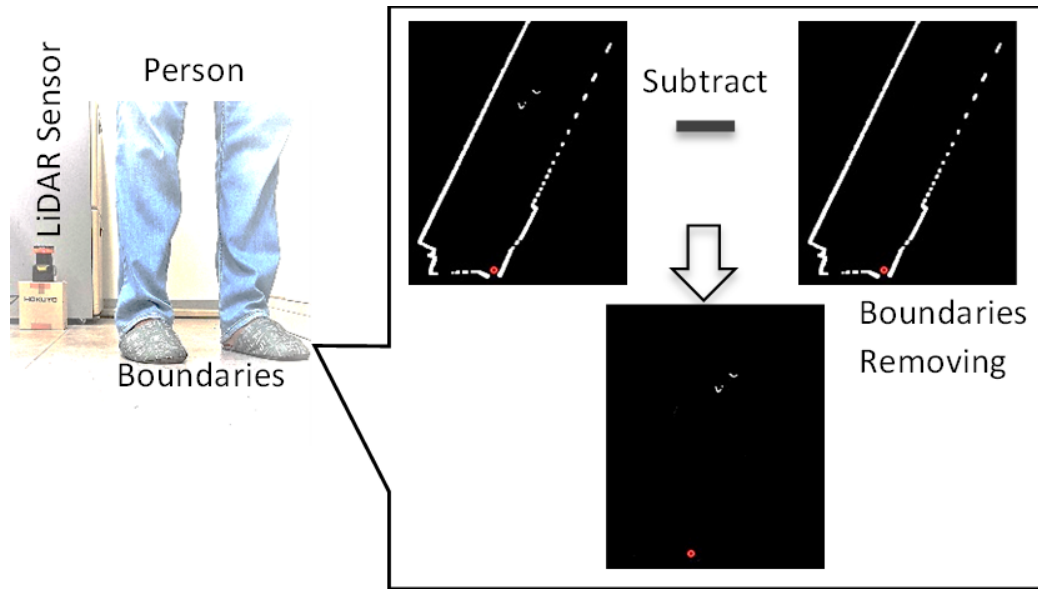


Fig. 4. 3 Image sensing using LiDAR sensor and preprocessing [94]

We used the HOKUYO UMT30LX LiDAR sensor for data acquisition. We keep our sensor static for our experiments and consider objects moving. Here objects refer to the pedestrian. In the above picture (Fig. 4.3), we held our sensor in an ankle position and captured time-series video data using LiDAR. These data were being stored for further surveillance. We applied some preprocessing techniques to make the data more robust and accurate. In the second part of the picture, a LiDAR-based image is placed with a white marked surface boundary. Here centered white marked dots show the ankle positions of a person in a specific frame. We tried to remove the white background from captured images. It helped us handle only the frame's ankle positions without unwanted noise. We applied the background subtraction method to eliminate boundaries. After that, we got only the ankles marked frames of every particular moment. Sequences of those frames created a LiDAR-based video.

Feature extraction from preprocessed data is another big concern in this research. We considered sequential frames for feature selection. Every chronological data represents every ankle position on the frame. Movements of an ankle and two ankles of a person were considered as a person. But critical situations always appear in data sensing. Occlusion, very fast-moving ankles, and omnidirectional movements of a pedestrian all these

challenges make this study more complicated and troublesome. To cope with these shortcomings and make the system more capable of time and calculation, we thought about using a second LiDAR sensor in another direction on the plane. Using another sensor makes the system inefficient and causes a higher installation cost. We changed our experimental architecture and focused on a single sensor-based diagram. Here we emphasized making a vigorous development of algorithms that can handle all these deficiencies.

## **4.2.2 Clustering**

The next section of this research is to cluster the data to handle all critical scenarios. We consider density-based algorithms. Density-based clustering can find random-shape clusters and also can manage noises. In density-based clustering approaches, clusters are created according to dense areas distinct from sparse regions. These algorithms work on data density. Here no previous assumption is needed. Clusters are made based on data density. DBSCAN is one of the popular density-based clustering algorithms we used in our research. We also applied the OPTICS algorithm in this research. Both algorithms performed very well on person tracking. Despite the many practical benefits of these two algorithms, still, some constraints must be fixed. We tried to solve these issues and proposed modifying these two algorithms. Enhanced-DBSCAN and Enhanced-OPTICS algorithms are the transformation of traditional algorithms but can perform better on LiDAR-based tracking.

### **4.2.2.1 Enhanced (E) DBSCAN**

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is devised to locate non-sphere-shaped clusters. Two specific parameters: first, radius  $\epsilon$ , a precise distance that defines the neighborhood, and the second one is the minimum number of points (minPts) that are considered as neighbor points, determine the dense neighborhood of DBSCAN. Figure 4.4 shows DBSCAN mechanisms.

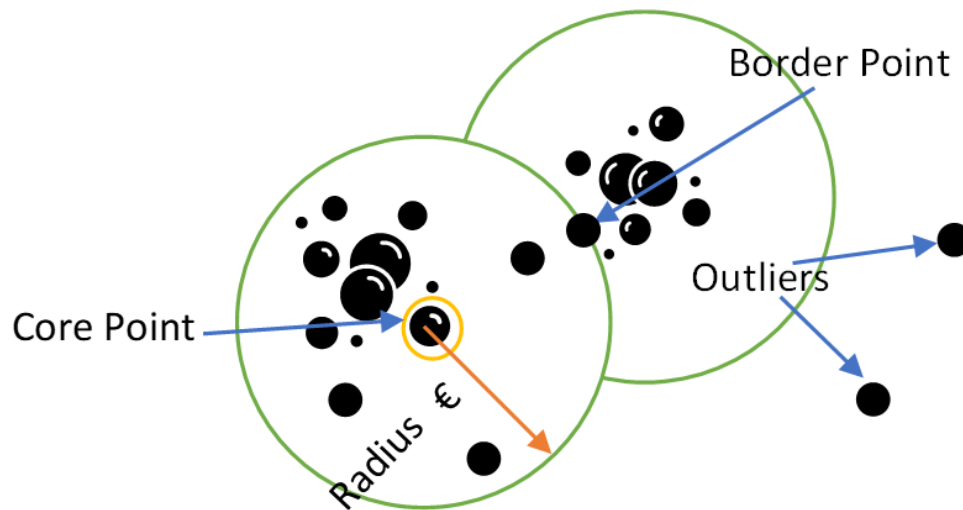


Fig. 4. 4 DBSCAN mechanism

**Core point:** Any data point is specified as a core point if this has minimum data points ( $\text{minPts}$ ) within its radius  $\epsilon$ .

**Border point:** Any point is specified as a border point if there are fewer data points than the minimum number but has at least one core point within the range  $\epsilon$ .

**Outliers:** If a data point is neither a core point nor a border point will be considered an outlier. Sometimes outliers are called noise.

**DBSCAN working procedures:** a) Categorize all data points as Core data points, border points, and outliers. b) A connectivity graph is formed according to core data points. Here, consecutive core points are linked if they are within the radius  $\epsilon$  of another. c). Select connected elements that are associated with one another within the range. d) Designate all border points to their corresponding components, where it fits best. e) Ignore all uncorrelated outliers. f) Get the Desired cluster.

DBSCAN is an exciting development to make a cluster based on LiDAR data. But DBSCAN still faces some limitations that degrade tracking performance to make a robust tracking system. Selecting initial parameters ( $\epsilon$ ,  $\text{minPts}$ ) is one drawback of this algorithm. But it has a high computational cost of  $O(n^2)$ , which affects real-time computations. Compared with other shortcomings of ordinary hierarchical algorithms, DBSCAN is a

better-performing algorithm, especially in LiDAR data processing. Here to track an ankle, we successfully set its initial parameters that work without disappointment in most cases. Furthermore, we used a LiDAR sensor and applied background subtraction, reducing data congestion and processing. So, our system was rated and functional with real-time pedestrian tracking.

We faced another difficulty in using DBSCAN. When multiple people are walking on the surface, their ankles come nearer. As we described, DBSCAN creates clusters based on  $\text{minPts}$  and  $\epsilon$ , ankles of different persons come so closer that the difference between two ankles of other people seems like a single person's ankle, where the difference value is lower than the radius. Here, DBSCAN fails. It creates a cluster with all closer ankles within its range, even if it belongs to different persons. Changing the thresholds of the radius may solve the problem. But in a practical application, the radius cannot be altered based on crowd size and movements. So, we must refrain from the structure of DBSCAN that best fits our application and solve this drawback of data overfitting.

From the images (Fig. 4.5), we can get a clear idea about the shortcomings of DBSCAN. The left topmost image shows different ankle positions on the frame. The corresponding lowermost image creates a cluster based on DBSCAN parameter  $\text{minPts}$  and radius  $\epsilon$ . Here we set our parameters as  $\text{minPts} = 20$  and  $\epsilon = 900$ , which means DBSCAN will create a cluster if it gets at least 20 points within the radius  $\epsilon$  and determined as a person. We see four persons walking on the plane from the first image, but the corresponding lowermost image detects it 5. Here fifth person's two ankles distance is more significant than our threshold  $\epsilon$ , but every ankle satisfies the condition of being a cluster that makes two clusters and shows as two persons. Two frames later, in the second figure of the top row, it recalculates the distances, creates four clusters, and shows the corresponding lowermost image. In the third frame of the top row, persons come closer two frames later. Here their ankles are occluded by one another. In this situation, the ankle distance between two or more persons comes lower than  $\epsilon$ , which creates a single cluster, an example of data overfitting. The last image shows two persons' results in the second row. This is a drawback of the straight use of DBSCAN.

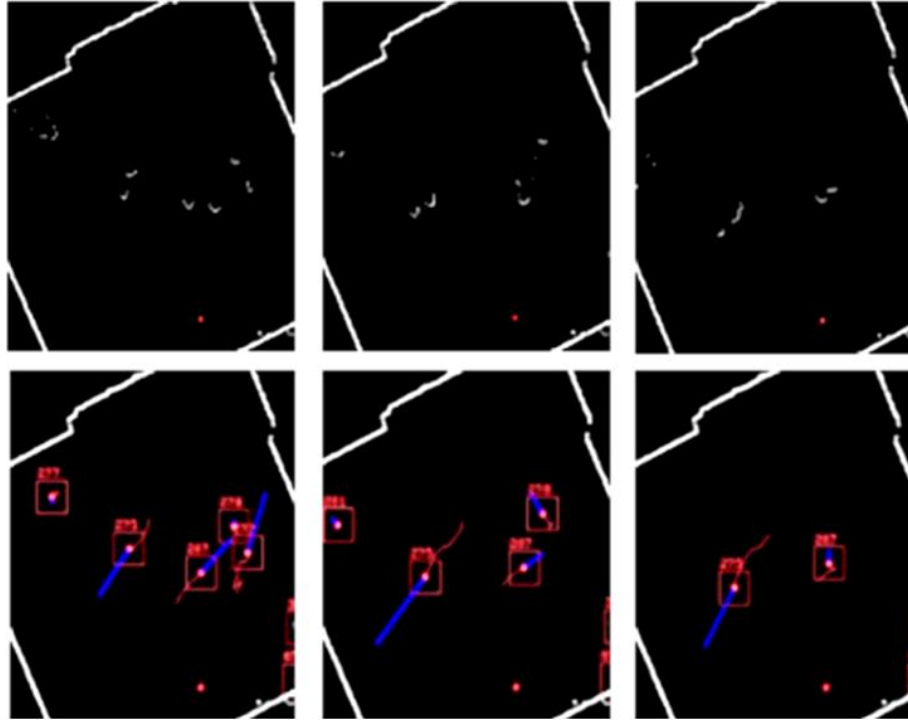


Fig. 4. 5 The uppermost images show an ankle position on the frame—lowermost images pointing persons with the handler [94]

We set the third parameter for the DBSCAN algorithm to solve these issues. We consider maximum points ( $\text{maxPts}$ ) and two previous parameters ( $\text{minPts}$  and  $\epsilon$ ). Setting this parameter can quickly reduce data overfitting and fixed radius problems. Suppose 10 points can create an ankle; then, at least 20 points will be needed to develop a cluster to be a person. Then  $\text{minPts}$  of DBSCAN can be 20. Again, we can set the distance  $\epsilon$  between two points using Euclidean distance as 900 units. If four ankles of two people come closer and their points distance lies within 900 units, then previous DBSCAN considers these as a single cluster. DBSCAN thinks about minimum data by which it can create a cluster. Here there is no obligation for maximum data. But LiDAR-based tracking fails in this concept. If we set  $\text{maxPts}$ , another parameter, like 50, this DBSCAN can create a cluster of data points within 20 to 50. As less than 20 points will not make a cluster parallelly, it will also discard the cluster if it goes a maximum of 50. Adopting this parameter can solve the data overfitting problem. If more than two ankles come closer and their combined number of points exceeds  $\text{maxPts}$ , Enhanced DBSCAN (EDBSCAN) will not accept this cluster.

EDBSCAN discards this cluster immediately. But there is still another problem: creating clusters again with its overfitted data. Here EDBSCAN changes its radius  $\epsilon$  immediately. It decreases its radius significantly and finds clusters with discarded points, not in the whole dataset. This step will continue until the overfitting problem is solved. Here minPts and maxPts values will remain the same; the only radius decreases.

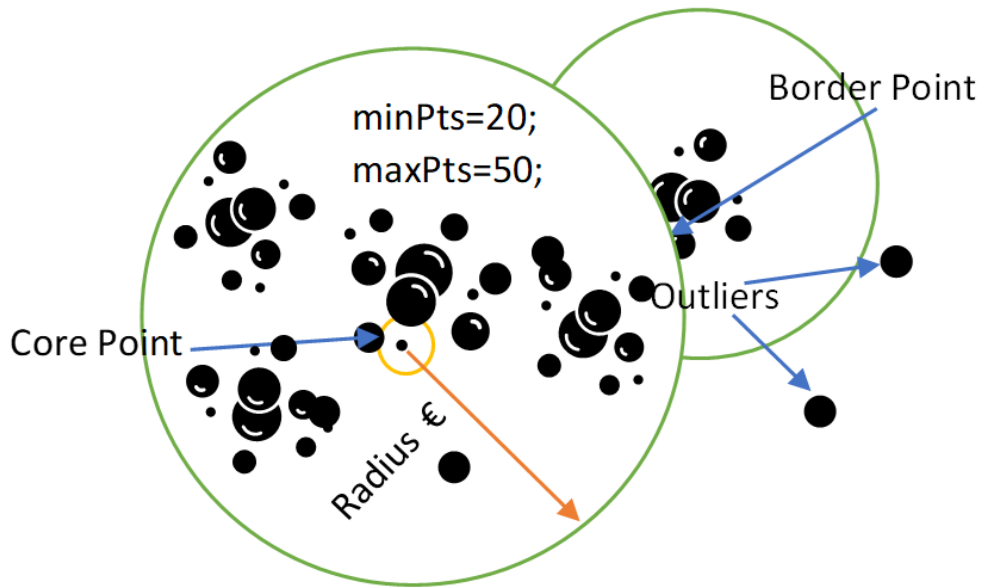


Fig. 4. 6 Enhanced DBSCAN (EDBSCAN) mechanism [94]

In figure 4.6, within the radius  $\epsilon$ , there are so many points within the circle. EDBSCAN will discard this cluster as it has more data points than maxPts. This method justifies the concept that more than one person walks in front of the LiDAR sensor whose ankles come closer. Here the bottleneck problem of traditional DBSCAN can be solved by our proposed Enhanced DBSCAN. Now another proposed portion of EDBSCAN is varied radius selection. As the initial radius selection creates a data overfitting problem, we can reduce the radius with a significant amount where minPts and maxPts parameters will be the same as previously. A person can be identified again in a crowded situation. In figure 4.7, radius  $\epsilon$  is reduced to  $\epsilon'$ ; now, it can find a new cluster with initial minPts 20 and maxPts 50, which will see those clusters that can satisfy the condition. Here though, the ankles of different persons are so close that there is a chance of misclassification, but a reduced radius can quickly solve the issue. Only very close points will make a cluster.



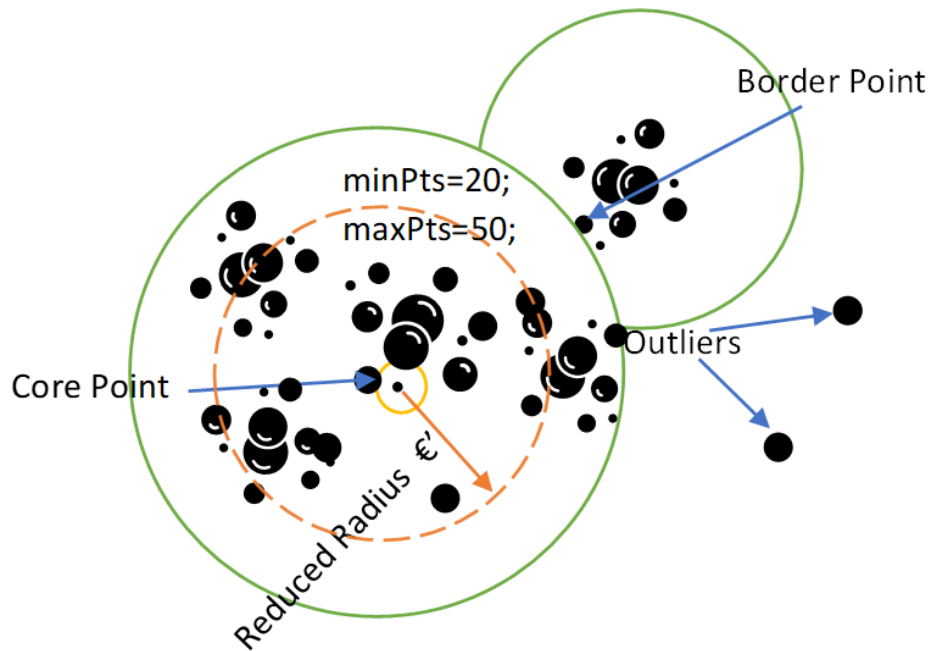


Fig. 4. 7 Reduced Radius of EDBSCAN [94]

We can also handle false clustering problems by setting max point (maxPts) parameters in DBSCAN. Suppose there is a moving object on the plane rather than an ankle. If it creates with required minPts and  $\epsilon$ , traditional DBSCAN will create a cluster as a person moves on the plane. Our proposed maxPts parameter can handle it easily. It produces clusters based on motion history images. It will discard all false clusters if it exceeds maxPts and does not appear on consecutive images.

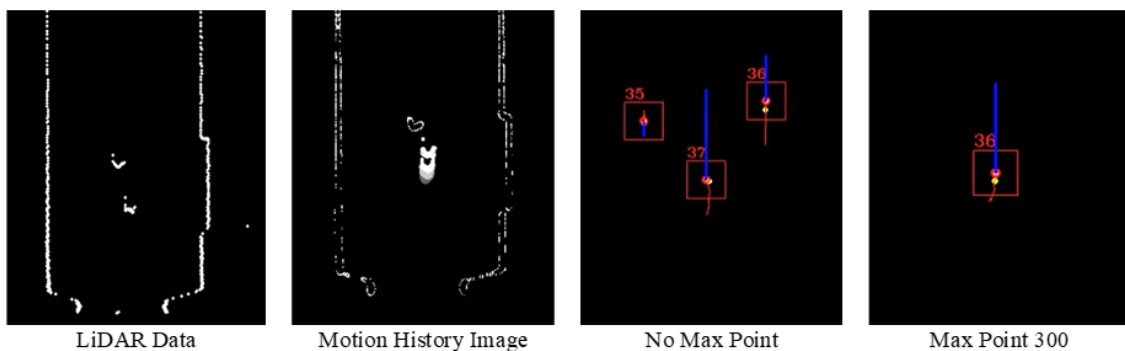


Fig. 4. 8 Anomaly in tracking without using maxPts [94]

Suppose in the above figure 4.8; there are four frames of a moment of a person walking in front of the LiDAR sensor. In the first frame, this is LiDAR data captured by the sensor. The second frame shows a motion history image of that moment. We see that only one person is walking on the frame from the previous frames. When conventional DBSCAN was used to track a person, an anomaly in tracking was shown in the third frame. But due to garbage data and false moving of objects, DBSCAN detects another two persons on the frame. When using the maxPts parameter with value 300, our EDBSCAN removed erroneous classification and showed only one person in the frame as original data. Table 4.1 shows the modified algorithm of EDBSCAN.

Table 4. 1 EDBSCAN Algorithm

<p>Input: Motion History Image <b>I</b>, Radius <math>\epsilon</math>, Minimum Points <b>minPts</b>, Maximum Points <b>maxPts</b>  Output: Cluster <b>C<sub>L</sub></b></p> <ol style="list-style-type: none"> <li>1 Set <b>C<sub>L</sub></b>:= <b>0</b> and queue <b>Q</b>:= <b>0</b> for the points that need to be checked</li> <li>2 <b>∀</b> points <b>Pt</b> on the image <b>I</b>, where <b>Pt</b> <math>\in</math> <b>I</b>, and <math>i=1,2,\dots,n</math>; <b>do</b></li> <li>3   label all points as core border and outliers.</li> <li>4   eliminate all outliers,</li> <li>5     <b>if</b> <b>Pt<sub>i</sub></b> is treated <b>then</b></li> <li>6        continue</li> <li>7     <b>end</b></li> <li>8     <b>∀</b> core points <b>Pt<sub>c</sub></b>, those are not yet treated</li> <li>9        set a new temporary cluster <b>Q<sub>i</sub></b> with point <b>Pt<sub>c</sub></b></li> <li>10      measure neighboring distance <b>D<sub>i</sub></b> <math>((x_a, y_a), (x_b, y_b)) = \sqrt{((x_a - x_b)^2 + (y_a - y_b)^2)}</math>  from point <b>Pt<sub>c</sub></b></li> <li>11        <b>if</b> <b>D<sub>i</sub></b> <math>\leq \epsilon</math> <b>then</b></li> <li>12          add all neighboring points <b>Pt<sub>i</sub></b> to <b>Q<sub>i</sub></b></li> <li>13          <b>n</b>:= count <b>Pt<sub>i</sub></b></li> <li>14        <b>end</b></li> <li>15     <b>end</b></li> <li>16     <b>if</b> <b>n</b> <math>\geq</math> <b>minPts</b> and <b>n</b> <math>\leq</math> <b>maxPts</b> <b>then</b></li> <li>17        add temp: <b>Q<sub>i</sub></b> to clust: <b>C<sub>L<sub>i</sub></sub></b></li> <li>18        <b>∀</b> <b>Pt<sub>i</sub></b> <math>\in</math> <b>Q<sub>i</sub></b> <b>do</b></li> <li>19          assign <b>Pt<sub>i</sub></b> := treated</li> <li>20        <b>end</b></li> <li>21        set <b>Q</b>:= <b>0</b></li> <li>22     <b>end</b></li> <li>23     <b>if</b> <b>n</b> <math>&gt;</math> <b>maxPts</b> <b>then</b></li> <li>24        reduce Radius <math>\epsilon</math> by unit amount</li> <li>25        repeat step10 to 22</li> <li>26     <b>end</b></li> <li>27 <b>end</b></li> <li>28 <b>Return</b> <b>C<sub>L</sub></b></li> </ol>
--

#### 4.2.2.2 Enhanced (E) OPTICS

OPTICS (Ordering Points to Identify Cluster Structure) is close to DBSCAN but refers to another procedure. This new approach is based on data density to make a cluster on spatial data. When data density varies, DBSCAN has a weak performance on data for making clustering. OPTICS addresses this drawback and can create profound clusters with differing densities. It works like the previous one in that all points in the input data are ordered as core points, border points, and outliers. OPTICS also works with two input parameters: maximum distance to be considered, radius  $\epsilon$ , and a minimum number of points to create a cluster, minPts. Some definitions can be used for making clustering as:

**Core Distance:** A core distance is the minimum value of radius  $\epsilon$ , which determines whether a point is a core point. The core distance cannot be calculated if a point is not a core point. Radius  $\epsilon$  and core distance are the same if any core point has exactly minPts data. But core distance may be less than radius  $\epsilon$  if any core point has more than minPts values as its neighbor and within the radius. Suppose from figure 4.9(a); a point Q needs a minimum of 6 minPts within its radius of  $\epsilon$  7mm to become a core point. But within a 4mm distance point, Q has six minimum points as its neighbor. So, point Q is a core point with a core distance of 4mm and a radius of 7 mm.

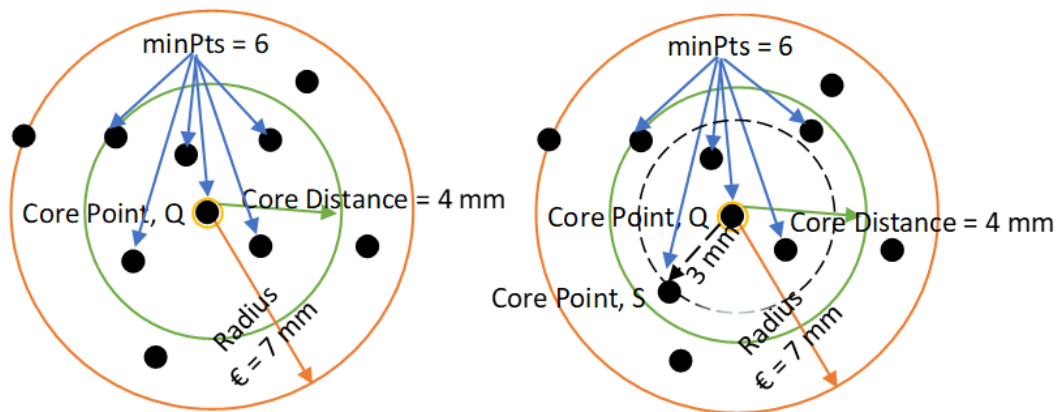


Fig. 4. 9 (a) Core Distance, (b) Reachability Distance

**Reachability Distance:** A distance  $R_d$  is said to be reachability distance if it calculates a distance between two core points Q and S. Here, the maximum  $\epsilon$  value between a core

distance of point Q and the straight distance between Q and S will be considered as a reachability distance. Suppose in figure 4.9(b) a point S has a plane distance of 3mm from core point Q. Here core distance of point Q is 4 mm. Now reachability distance  $R_d$  between core points Q and S will be the maximum value between a plain distance of these points and core distance. So, it is 4mm. The reachability distance  $R_d$  between core points Q and S can be measured as:

$$R_d = \max\{Dist(Q,S), Core Dist(Q)\} \quad (4.2)$$

The OPTICS algorithm aims to group data into significant subclasses as part of an experimental process to understand data's inner relation. Furthermore, this approach can also be applied as a preprocessing phase of other algorithms—figure 4.10 shows the traditional OPTICS mechanism principle.

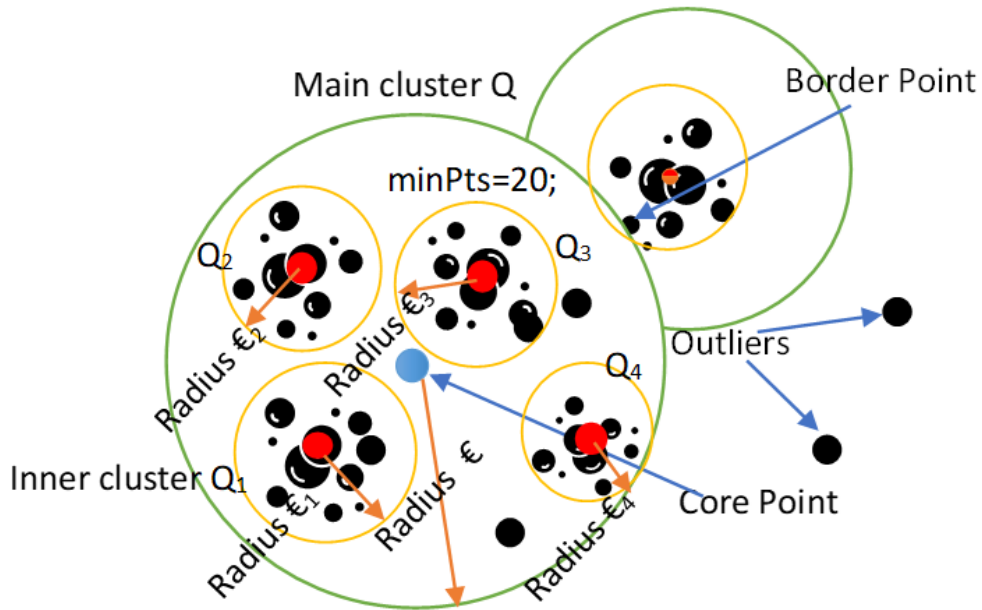


Fig. 4. 10 Principle of Traditional OPTICS

OPTICS does not create a data set clustering. It provides cluster ordering. This is a linear directory of items and denotes density-based grouping of data. Here it does not provoke users to give a density threshold. OPTICS follows a precise ordering of objects or points for concurrent clustering. Based on preserving orders, OPTICS selects density-reachable

objects related to the lowest radius to finish the clusters with higher density (lower radius). The database preserves the appropriate reachability-distance and core distance for every item. To produce the ordered output OPTICS, maintain a list named OrderSeeds. All items in OrderSeeds are preserved based on the shortest reachability distance as of their corresponding nearest core objects.

OPTICS works on the same cluster, which can find sub-clusters within the cluster. This approach can determine the inner trends of clusters within the central cluster. We have proposed the development of OPTICS as Enhanced OPTICS named EOPTICS, where another parameter is applied as maximum data points named maxPts along with previous minPts and Radius  $\epsilon$ . The idea is exciting and aligned with our practical application, LiDAR-based tracking. To make a cluster data point should be within the minPts, maxPts, and radius  $\epsilon$ . Our approach is to create a density-based cluster that does not exceed maxPts with its data point. The remaining process will be the same as OPTICS. Here it will check cluster density. This approach is a clear development and significant modification of OPTICS. Here highly dense areas where more than one people's ankles come closer and traditional OPTICS cluster creates a single cluster with all data points that can be ignored. Now enhancement of OPTICS has another significant benefit. If several points exceed maxPts, our algorithm discards the outer cluster, which is the main cluster. The inner clusters based on core distances and reachability distance will be calculated here. These internal clusters will help to determine the ankle's positions on the plane. Our EOPTICS then reduces radius  $\epsilon$  significantly and computes clusters again. Inner clusters will also be calculated to find the associativity of data within the main cluster. This approach will help to find correlated subclusters that imply a person on the plane.

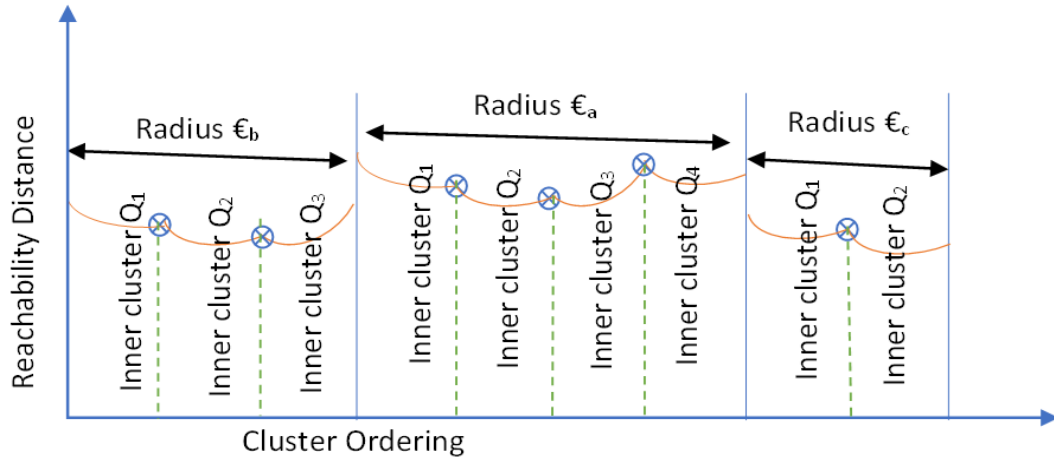


Fig. 4. 11 Sample run of EOPTICS with varied reachability plot

In this figure 4.11, we see varied radius and clusters with different sub-clusters. Suppose initially Radius size was  $\epsilon_a$ . It creates a density-based cluster with four inner sub-clusters, which means four core points have a minimum number of minPts that satisfy the EOPTICS criteria. If this cluster exceeds the maximum bound of the maxPts parameter, then EOPTICS reduces its parameter and assigns a value  $\epsilon_b$ . With this new radius, EOPTICS tries to form a cluster with minPts and maxPts parameters. Concurrently it finds reachability order with inner clusters. If it fail it reduces its radius again as  $\epsilon_c$  and calculates again. This process will continue until we get desired density cluster with minPts and maxPts. Internal subclusters show data affinity towards the cluster. This is an extensive modification over DBSCAN. In our application, misclassification can be handled with this concept. When many peoples are walking together, we can handle misfit data easily. Our imposed maxPts limit the abnormal cluster size. Then reducing the radius will allow finding a cluster again and again until we get the desired cluster. Here Inner subcluster helps to track ankle positions and their associativity corresponding to a person. Ordinary DBSCAN and OPTICS create a cluster based on radius and minPts parameters. Here varied density of data cannot be handled accurately. Highly dense data and loosely dense data cannot be tracked in the same way. EOPTICS nicely manages this issue. There is no pressure to create a single cluster with different dense data here. Our modifications give a sophisticated solution over these two well-known algorithms—table 4.2 shows Enhanced OPTICS (EOPTICS) algorithm.

Table 4. 2 EOPTICS Algorithm

```

Input: Motion History Image I, Radius  $\epsilon$ , Minimum Points minPts, Maximum Points maxPts
Output: Cluster CL, Mi
1 Set CL:= 0 and queue Q:= 0 for the points that need to be checked
2 for all points Pt on the image I, where Pti  $\in$  I, and  $i=1,2,\dots,n$ ; do
3     Pt.Reachability_dist:= Unspecified
4     for all unprocessed points Pt on the image I, where Pti  $\in$  I, and  $i=1,2,\dots,n$ ; do
5         label all points as core, border, and outliers.
6         N:=count_neighbors(Pt,  $\epsilon$ )
7         treat Pt as processed
8         if Pti is treated then
9             continue
10        end
11        if core_distance(Pt,  $\epsilon$ , minPts, maxPts) != undefined then
12            OrderSeeds = empty Queue
13            Update (N, Pt, OrderSeeds,  $\epsilon$ , minPts, maxPts)
14        end
15        for all core points Ptc, those are not yet treated
16            set a new temporary cluster Qi with point Ptc
17            measure neighboring distance from point Ptc
18            if Di  $\leq$   $\epsilon$  then
19                add all neighboring points Pti to Qi
20                N:= count Pti
21                Mi:= count Ptc
22            end
23        end
24        if N  $\geq$  minPts and N  $\leq$  maxPts then
25            add temp: Qi to clust: CLi
26            for all Pti  $\in$  Qi do
27                assign Pti := treated
28            end
29            set Q:= 0
30        end
31    end
32    if CL:= 1 then
33        if Mi>1 then
34            remove CL
35            Set CL:=Mi
36        end
37    end
38    if N>maxPts then
39        reduce Radius  $\epsilon$  by unit amount
40        repeat step1 to 42
41    end
42 end
43 Return CL, Mi

```

In all cases, cluster parameters will be calculated based on some critical criteria-

- a) If the cluster radius creates only one cluster, it will calculate the core distance and find subclusters. If there is more than one sub-cluster, it will discard the outer main subcluster and consider inner sub-clusters as the main cluster. If only a few clusters were created with very few minPts, it will increase its radius  $\epsilon$  and recalculate clusters again.
- b) If for  $n$  minPts EOPTICS creates  $m$  cluster where  $m \geq \frac{D}{2}$  then discard the cluster and recalculate minPts as  $minPts = 2 * minPts < maxPts$ . The process will continue until  $m < \frac{D}{2}$  Where  $D$  is data points.
- c) If there is no cluster found with the radius  $\epsilon$ , EOPTICS increases  $\epsilon$  and tries to calculate the cluster. If more than one cluster is found, then it will keep it; discard otherwise.

In figure 4.12, we can see the traditional approaches of the EOPTICS algorithm. In the left picture, we see a certain frame where four people are moving in front of the LiDAR sensor. This is close to OPTICS. It gets ankle positions from the sensor and makes these clusters. Kalman filter-based tracking tracks all these persons based on data. Parameter  $maxPts$  setting is a heuristic process. It depends on crowd size and frequency on the surface. For our experiments, we varied this from 50 to 300 points.

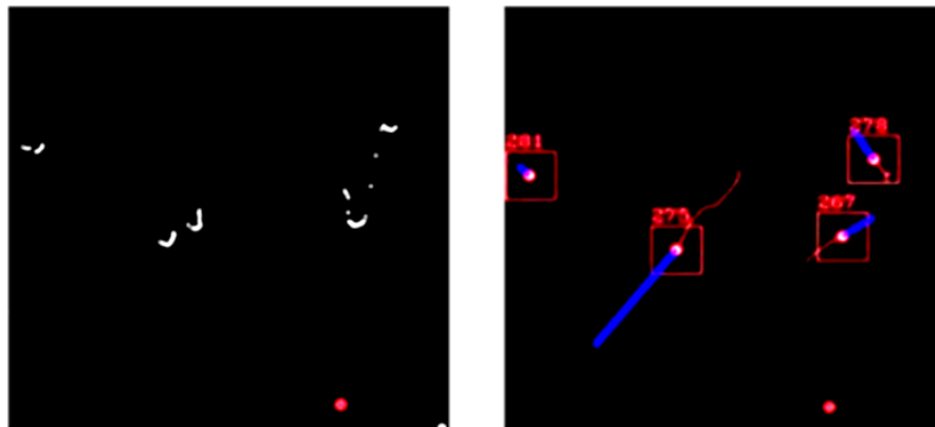


Fig. 4. 12 EOPTICS with LiDAR data; a) Ankle Positions in front of the sensor, b) Kalman filter-based tracking [94]



A moving direction is also shown in which direction they are moving. One frame later, one ankle of a person goes out of range of the sensor, and two people come closer than their ankle's distance lies within the scope of our radius. Ordinary OPTICS creates only two clusters here. One cluster is omitted because of a shortage of minPts parameters. Here only one ankle cannot fulfill minPts requirements. The second cluster was created with no anomalies that show ID 275 in figure 4.13 (b). The critical condition appeared in the next section. Here two persons come closer. In the previous frame, they have identified accurately with ID 267 and 278. But ordinary OPTICS created only one cluster 267 and showed it as a single person. This is the shortcoming of OPTICS.

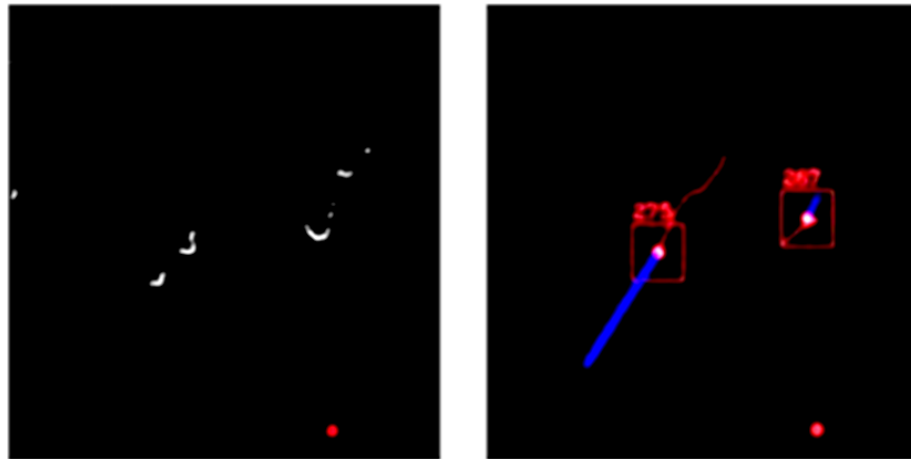


Fig. 4. 13 OPTICS; a) one frame later, ankle positions, b)  
Corresponding Kalman based tracking [94]

Our proposed EOPTICS plays a very significant role here with this inadequacy. In figure 4.14, we see a very substantial enhancement over OPTICS; EOPTICS gave 3 clusters where the Kalman filter provides three persons on the frame. The second cluster was created as previously. For the same reason, OPTICS rejected one cluster here; EOPTICS rejected it for the same that the data points are less than one parameter minPts.

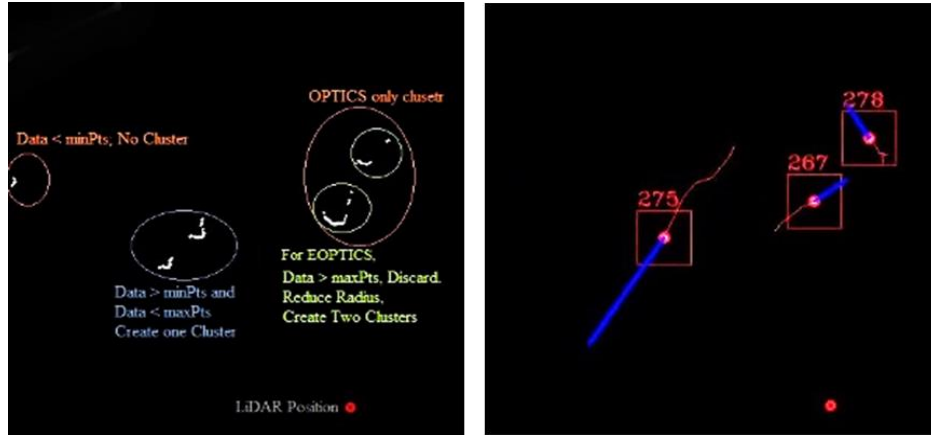


Fig. 4. 14 EOPTICS, a) ankle positions, b) Kalman filter-based tracking [94]

Here the data points lie within our range greater than and less than minPts and maxPts, respectively. EOPTICS revised the third cluster than OPTICS. When OPTICS created only a single cluster with two person's ankle points, EOPTICS denied making this a single cluster. Here the idea is, it counts minPts as well as maxPts. In this situation, data points satisfy the minPts condition but have more data points than maxPts. As conditions were dissatisfied, EOPTICS stopped creating a cluster. It works with its next phase. It reduced its radius  $\epsilon$  and tried to develop clusters again only with these sorted data. After reducing  $\epsilon$ , only relevant ankle points satisfy all the conditions and create two separate clusters. In the right-side figure, handlers 267 and 278 show these.

### 4.2.3 Tracking

The next step of our research is tracking. The Kalman filter tracks the gravity point of the cluster. A forecast equation and a revised equation are the base of the Kalman Filter (KF).

$$\begin{aligned} \text{Forecast, } \mathbf{F}(n) &= \mathbf{X}\mathbf{F}(n-1) + \mathbf{S}(n-1) \\ \text{Revise, } \mathbf{R}(n) &= \mathbf{Y}\mathbf{F}(n) + \mathbf{M}(n) \end{aligned} \quad (4.3)$$

Where  $\mathbf{F}(n)$  and  $\mathbf{R}(n)$  are forecasting estimation and revised measurement variables,  $\mathbf{X}$  and  $\mathbf{Y}$  are State-Transition-Matrix and Measurement matrices, respectively.  $\mathbf{S}$  represents system noise, and  $\mathbf{M}$  represents measurement noise. Input motion history images are taken as input, and KF tries to forecast it. If it can detect, then forward it to the revised section. If any revision is needed, it again feedbacks it to the forecast section. This process will continue until the desired tracking is found. Finally, a person is being tracked. A handler is used to show a person's walking path in figure 4.15.

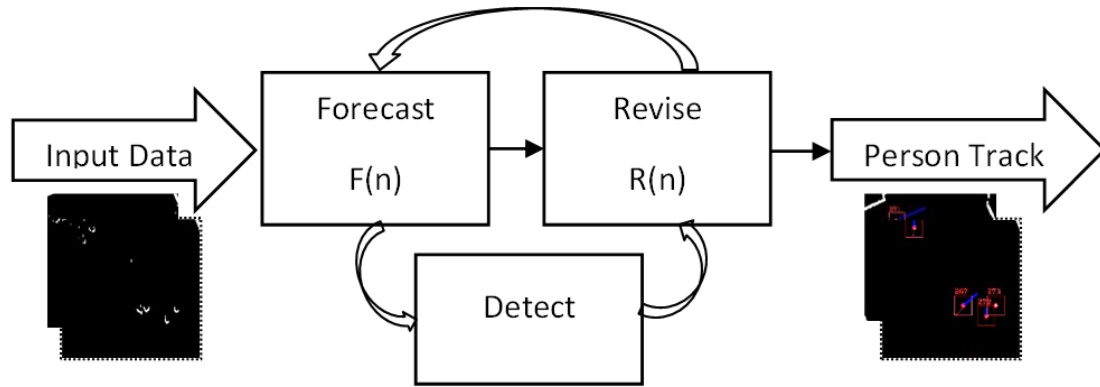


Fig. 4. 15 Flow diagram of UKF based person tracking

### 4.3 Result and Discussion

As there were no well-established datasets on LiDAR-based person tracking, we prepared our datasets for experiments. LiDAR-based tracking always suffers from data misfits and low accuracy problems. We used a LiDAR sensor that suffers a lot in this situation. Initially, we used recorded videos for our experiments. We found significant results there. Further, we enhanced our system for real-time tracking. Interestingly we get comparable improvements with the video-based person tracking system.

Different Gestures (EDBSCAN)	Ankle Positions on the frame	Direction of the Movements
Combined Walking		
Combined Running		

Fig. 4. 16 EDBSCAN based Group Peoples tracking with LiDAR sensor [94]

We performed our experiments with different male and female participants. We considered another group of people for our experiments. We chose five substantial categories for this analysis: Individual Walking, Individual Running, Ankle Movements Only, Combined Walking, and Combined Running. We applied our developed EDBSCAN to individuals and groups of people tracking. Here in figure 4.16, combined walking and running are focused. In the first frame, we see three people walking in front of the LiDAR sensor. Our progressive approach tracks these people and shows them in the rightmost picture. We see four people running in the second picture, and our progressive approach identifies all these people. Here EDBSCAN plays a very significant development of traditional DBSCAN. When more ankles come closer, it can prevent misclassification. maxPts parameter discards these clusters and recalculates again.

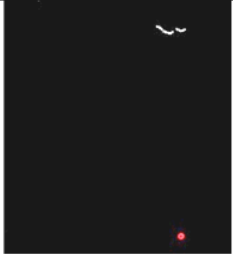
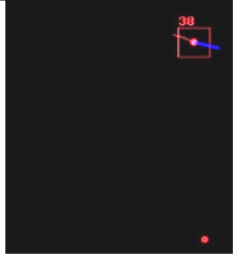

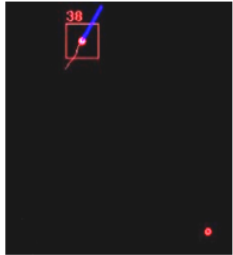

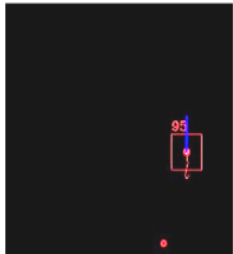
Different Gestures (EOPTICS)	Ankle Positions on the frame	Direction of the Movements
Individual Walking		
Individual Running		
Only Ankle movement		

Fig. 4. 17 EOPTICS based individual tracking with LiDAR sensor [94]

We have also examined EOPTICS for our experiments. The development of OPTICS as EOPTICS shows significant results in our system. We applied this algorithm to our same data sets and found EOPTICS most effective in the frame-by-frame analysis. In figure 4.17, we placed three gestures together and found EOPTICS more functional than previous DBSCAN and OPTICS. Here, most pictures show the ankle positions in the frame, and corresponding rightmost pictures show their movements and tracking based on KF. If any misclassification occurs in an individual frame, EOPTICS solves this with its parameter update techniques in the next frames. This same approach also applied to the combined walking and running of different people.

We solved occlusion problems also. It occurs when one ankle comes in front of another, and due to obstacles, LiDAR cannot get a signal from the next one. In a person tracking system, this occlusion problem creates a problem, and sometimes it is impossible to track a person accurately for this scenario. Our proposed LiDAR-based tracking system can handle this problem. In figure 4.18, suppose two frames ago, our proposed algorithms calculated the distance between ankles and found this more than the acceptance threshold. This will decide it as two persons. In one frame back, both frames come closer, and their distance becomes less than our radius, and here it is updated as one person. This is the continuous up-gradation of tracking. If any anomalies occurred in any frame, they could be modified in consecutive frames. These tracking anomalies can be accepted in an individual frame or very few frames for a person tracking.

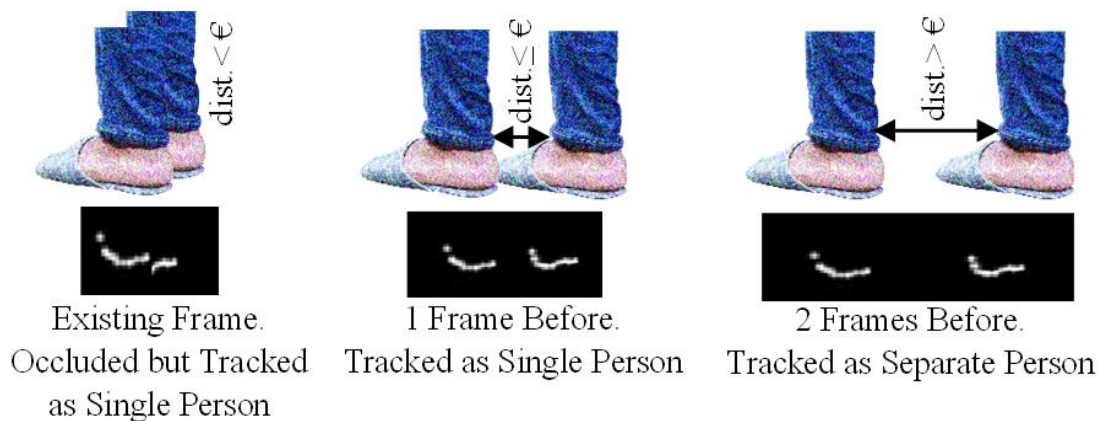


Fig. 4. 18 Tracking and occlusion handling, a frame-by-frame analysis [94]

Similarly, in the current frame, one ankle is occluded by another, and ankle distance cannot be calculated. It happened because of a lower gap between ankles or occlusion. There is no problem if the radius is less than the threshold. It satisfies the condition. If data points are less than the maxPts parameter, it will create a cluster and identify it as a person. If multiple people come so close that their ankle distance cannot be measured and the number of points goes over the maxPts, it will discard the cluster, recalculate the radius, and form a cluster again.

In Table 4.3, we have shown some experimental results of different gestures. Here five motions were considered. Another group of people participated in our experiments for a couple of minutes. We took random frames here for validation and verification. Suppose in the table below for individual walking and running 3 and 4 people have participated. We took 44 and 49 consecutive frames randomly. We got almost correct recognition based on our developed EDBSCAN and EOPTICS algorithms.

Similarly, we got considerably better results for only ankle movements with 52 and 53 frames accurate recognitions out of 56. For combined walking and running, we took nine people in our experiments. In the case of the group walking here, EOPTICS performs better than EDBSCAN. But compared with previously developed algorithms, it is also impressive development. We have a good scope to improve our system for a combined running situation. If a group of people runs fast to and fro, it is not easy to trace with utmost accuracy. We considered 136 frames for our experiment where EOPTICS recognized 106 frames and EDBSCAN could 101 frames accurately with a total elapsed time of the algorithms are 32.41 seconds and 27.63 seconds, respectively. Our used LiDAR sensor (UTM-30LX) took 25 milliseconds per scan.

Table 4. 3 Experimented results for EDBSCAN and EOPTICS [94]

Gestures	Persons	Frames	Correctly Identified by EDBSCAN	Correctly Identified by EOPTICS
Individual Walking	3	44	44	44
Individual Running	4	49	48	48
Ankle Movement	5	56	52	53
Combined Walking	9	131	123	127
Combined Running	9	136	101	106

We compared our improved algorithms with traditional DBSCAN and OPTICS. In Table 4.4, we have shown these performances and accuracy for better understanding. For individual walking, all four conventional and improved algorithms performed well. This scenario is the same for individuals running also. For only ankle movements, EOPTICS plays sizable improvements here. It can handle some misclassifications more than others. These developments also persist in group walking and running. Almost in all cases, EOPTICS plays better results. EDBSCAN also performs well compared with previously applied density-based algorithms.

Table 4. 4 Comparison of EDBSCAN and EOPTICS with DBSCAN and OPTICS [94]

Gestures	Accuracy with DBSCAN	Accuracy with OPTICS	Accuracy with EDBSCAN	Accuracy with EOPTICS
Individual Walking	99%	99.5%	100%	100%
Individual Running	98.01%	97.86%	97.96%	97.96%
Ankle Movement	92.07%	92.02%	92.86%	94.64%
Combined Walking	93.71%	93.81%	93.89%	96.95%
Combined Running	71.87%	72.11%	74.26%	77.94%

It is admissible that typical hardcore clustering algorithms are not well suited to LiDAR-based person tracking. We previously performed and have shown here some experiments regarding this issue. Both EDBSCAN and EOPTICS apply to these types of data sets. Furthermore, improvements in DBSCAN and OPTICS do not heighten algorithm complexity. The worst-case complexity persists in  $O(n^2)$ ; without matrix completion, it requires only  $O(n)$  memory.

Table 4. 5 Comparison with existing Approaches and Precisions [94]

Approach	Cluster Threshold	Precisions
Euclidean Clustering [107]	0.5 m	64.5%
DBSCAN [93]	Adaptive	93.7%
Depth Clustering [108]	10°	39.2%
Run Clustering [109]	Params <sub>SLR</sub>	51.7%
Online Learning [102]	Adaptive	89.8%
Our (EDBSCAN)	Adaptive	93.9%
Our (EOPTICS)	Adaptive	96.9%

We compared our approaches with present advanced precisions (Table 4.5). Different people used different clustering approaches and datasets for person tracking, but we wanted to show our improved EDBSCAN and EOPTICS precisions. In 2011, Rusu et al. [107]

stated a strategy for tracking with Euclidean Clustering. They got 64.5% accuracy with a cluster threshold of 0.5m. In 2016, Depth Clustering [18] approach was applied for the same and found incredible 39.2% precision with a  $10^\circ$  entry. Another technique Run Clustering [19], was considered in 2017 and found 51.7% accuracy in 3D tracking. Online Learning [20] proposed an adaptive threshold selection method in 2020 and found 89.8% accuracy. Our previous approach used only DBSCAN and got 93.7% accuracy in in-person tracking during walking. This research proposed two novel techniques, EDBSCAN and EOPTICS, with an adaptive threshold selection mechanism and found a maximum of 93.3% and 96.9% accuracy, respectively, with the same datasets.

#### **4.4. Conclusion**

This research introduced an augmented comprehension for person tracking with a LiDAR sensor through the exclusive introduction of density-based clustering with its prior modifications. We were applying new parameters and its collaborative expansions made these algorithms sophisticated, especially for LiDAR data. While video and multi-sensor-based human tracking were improved in the time being, they have a pre-cautious consideration of privacy. Eventually, these improved tracking systems need higher computations and costs to be implemented. We considered all these constraints and compared our method with other state-of-the-art technologies. Low installation and computational costs, runtime processing, and adaptive parameter selection of algorithms made our system robust. A well-prepared own dataset made our system more accessible to others.

We intend to apply this to a mobile robot that can detect a person independently in real-time. Some commercial applications can also be made with this approach and the caregiving facility. We intend to develop a LiDAR-based gait recognition system. In the future, a deep neural network architecture can be applied to overcome the shortcomings of many samples' learning and training.



## **Chapter 5**

# **Person Property Estimation based on 2D LiDAR Data using Deep Neural Network**

### **5.1 Introduction**

From the beginning of artificial intelligence, accurately identifying persons and their different properties was always challenging [106]. How a human think, memorizes and makes decisions, these techniques were tried to imitate a machine. Over the period, developments of new designs, machine learning algorithms, and increased computational capabilities help us find some sustainable and reliable models for different property estimations of a person. All these developments come to real momentum after the innovation of deep neural networks. Massive data can be processed with a brain-like model using deep learning. Our study also finds a breakthrough after using DNN models.

Video-based surveillance and analysis played a vital role in human property estimation. Introducing new video cameras made the analysis more accurate and sophisticated day by day. Depth, night vision, RGB, and wide-angle architectures helped us capture high-definition images and videos in different circumstances. Using machine learning algorithms and DNN architectures, it is relatively easier to process these data for any further enhancements. Above all these benefits, video cameras are not error-free and well accepted in all conditions. Some natural and privacy issues arise while using surveillance cameras in all situations. Some regulations and confidentiality policies also prevent us from using video cameras in all areas. Our study aims to focus on all these issues and find a well-fit alternative of cameras to identify a person's property accurately. Here we emphasize height and age for recognition. A very well-suited application of this property estimation

is museum guided robot. Analyzing a person's walking data can predict their height and age appropriately. Adults and a child can also be accurately identified and guided. Though walking speed and patterns vary during watching and working, our proposed system can handle these issues very well. It continuously scans the person and can update its previous data simultaneously. Some other applications, i.e., childcare agents, elderly support robots, aircraft assistance, etc., can easily support their clients with this application.

This research has used a 2D LiDAR sensor as an alternative to a video camera for data acquisition. As LiDAR sensors provide distance data of a moving object's territory, we collected these data, placed them at a particular time on a frame, and created motion history images. We considered human walking places' general situation and tried to gather their data from different analysis angles. We used our network's images of different persons as training and test data. We further validated the model with these data, which increased the model's efficiency. For our research, we used Residual Network (ResNet) as the backbone of our network. We used twenty-nine-person data for our analysis and got a significant performance. We categorized our study differently to find the person's various properties with multiple considerations. Our study covered the analysis of predicting a person's height in two and three categories. We also extended our research to indicate a person's age with these data. Even though LiDAR-based applications have a bottleneck of accuracy, we expanded our study with different neural network models and tried to find an optimal model with various disjoint data.

In Fig. 5.1, we can see a block diagram of our proposed system. We collected LiDAR data by sensing the sensors continuously. For our experiments, we placed LiDAR sensors at the ankle level. We evaluated four different sensor data for our investigations and showed significant results with all the data. We processed these LiDAR data and created motion history images with varying frames per second (FPS). These images are used as input of the deep neural network (DNN) for analysis. A pre-trained ResNet is used as modeled network and analyzed for two specific person properties: height and age, as a research domain. Our study shows very substantial results in different conditions. We varied our

datasets in various patterns and performed our experiments for fault tolerance and the model's sustainability.

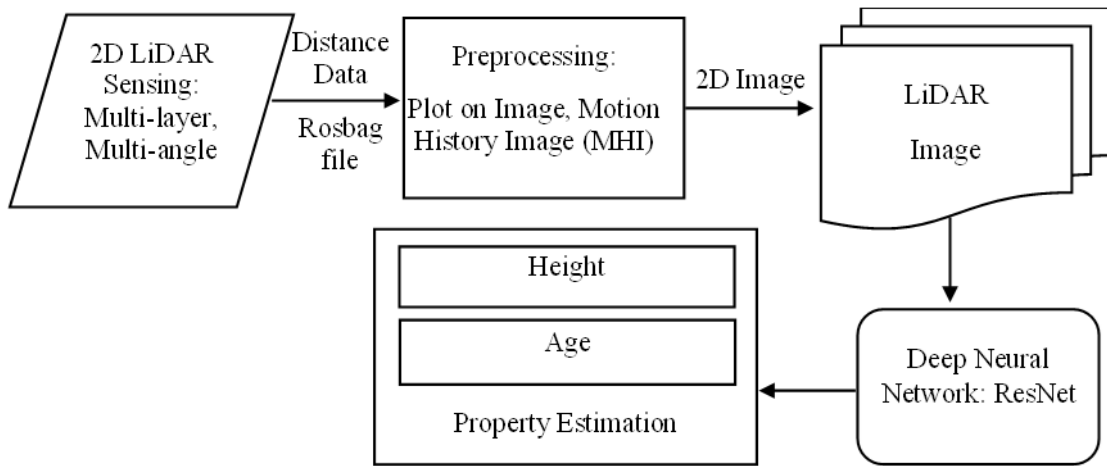


Fig. 5. 1 Block Diagram of the Proposed System [106]

This study emphasizes the use of a 2D LiDAR sensor as an alternative to video cameras. We enhanced our research to find the persons' different properties using LiDAR sensor as cameras are less performing in some vulnerable situations. Interestingly our model most accurately identified our test set data with disjoint train and validation data.

## 5.2 Proposed Method

### 5.2.1 Dataset Preparation

We prepared our dataset for this research. We considered 29 users in this experiment of different ages and gender. In our study, various geographical peoples also participated. Most of the people wore shoes, but few of them wore sandals. In our research, people of different heights and ages also attended. Sixteen persons were below 170 cm in height, and others were above the threshold. The age limit was between 22 to 36 years old. Thirteen persons were greater or equivalent to 30 years old, and others were below 30 years. We considered four different LiDAR sensors at different altitudes and angles. Then captured all these LiDAR data through ROS (Robot Operating System) environment to a '.bag file.' Individual LiDAR data of every person was stored in a separate bag file.

All bag files generate LiDAR motion history images (MHI). These images were used as an input to our proposed method. Fig. 5.2(a) shows motion history images of LiDAR data. Here different colors (i.e., red, green, and yellow) indicate different LiDAR data captured from different layers and/or different angles. The corresponding grayscale plot of these data shows the same data. All these lines are for a specific moment of a single person's ankle movement. We accumulated 0.5sec (20frames) of data from all LiDARs for generating MHI. These images are being used as input for our system. Figure 5.2(b) shows the image dataset of different participants. We categorize these datasets for height estimation and age measurements. Based on height, there are two classes of data we prepared: tall and short. Sometimes we incorporated another type as the medium. For age estimation, we grouped these data as young and elder. These categories help us accurately identify a person's class based on their walking data.

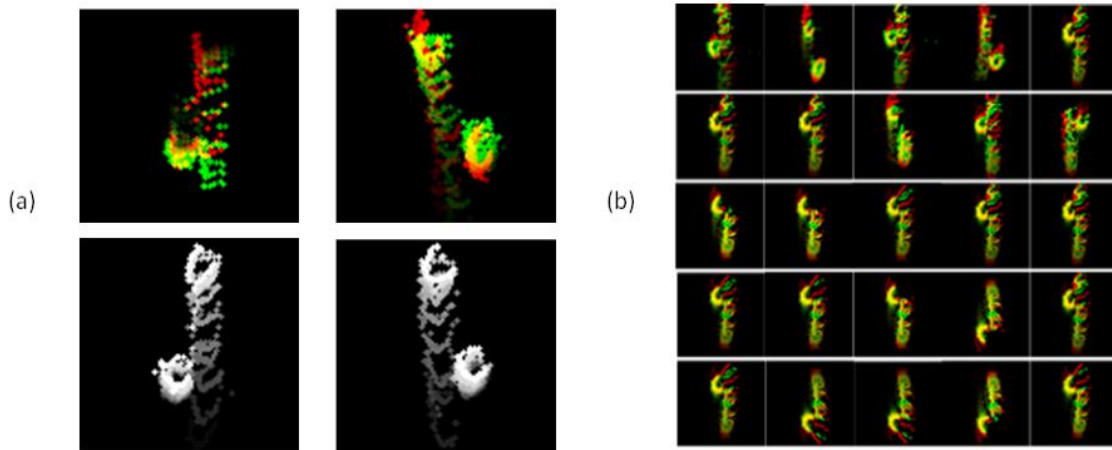


Fig. 5. 2 (a) Motion history images (Color Image and grayscale) (b) Image Dataset [106]

### 5.2.2 Person Property Estimation

For a human, the different properties can be estimated for various purposes. Here we considered height and age in this circumstance. Estimating a person's height is a vital property in other applications. Clothing, defense recruitment, safety and security agencies, rescue measurements, live programs, event management, etc., applications are closely related to human height. Some cases are restricted from disclosing a person's identity. Here no video cameras are allowed. Our LiDAR-based property estimation technique is an excellent alternative to these applications.

Moreover, LiDAR data are independent of bias by light, motion, and natural calamities. We placed our LiDARs in four positions. Two LiDARs are at the same angle but at different heights, and another two sensors were 2 meters apart from the first and at different heights and angles. These positionings are described here as multi-layer and multi-angle in this study. In Fig. 5.3, our experimental setup is shown. Four LiDAR sensors are placed in multi-layer and multi-angle positions. Persons are walking in front of these sensors, and they collect data. We used different LiDAR images in our experiments for training, testing, and validation.

For person property estimation, we placed LiDAR sensors at ankle level height. Peoples come in front of these sensors, and it collects their data. A rosbag package is used for hardware interface with LiDAR sensor and computer. Raw data of LiDAR sensors were stored as a bag file. We considered a batch program to make a motion history image (LiDAR image) from a bag file by combining different LiDAR data. These images are used as input for our application. We considered PyTorchLightning to train these images. With the developments of deep learning architectures, possibly deep Residual Network is the most revolutionary invention in Computer Vision and deep neural networks peoples in the last couple of years. ResNet enabled us to train top layers, even thousands, with encouraging performances and precise constraints. We used ResNet 18 and ResNet 50 in different conditions in our application and found a significant improvement in accuracy with all disjoint datasets.

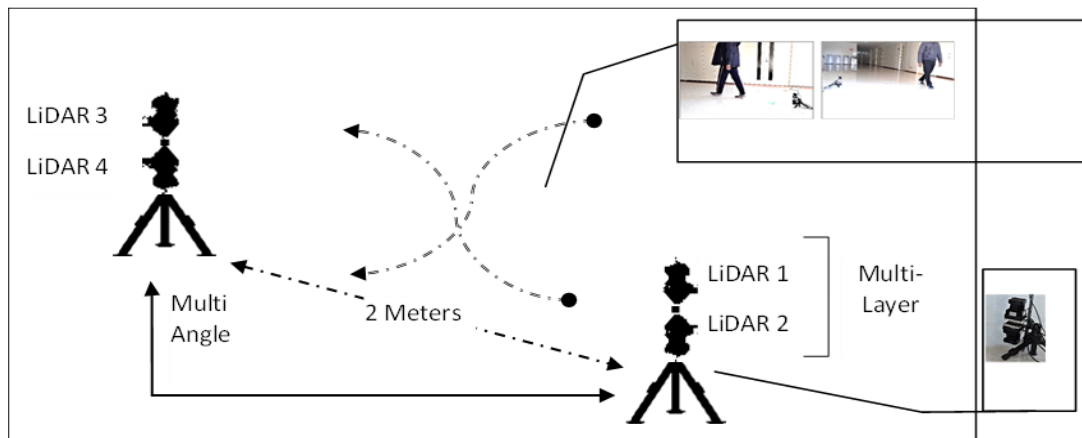


Fig. 5. 3 Experimental setup and person walking in front of LiDAR sensors [106]

### 5.2.2.1 Height Estimation

In this research, we considered two properties of a person to estimate by LiDAR data. The first one is height. Accurate height estimation is essential in different geometric estimations and scientific research. Some applications are susceptible to accurate height estimation. Even though very well-established research has been done on the topic, this is still a thrust sector. Some recent studies proposed human height estimation based on depth and color information [99]. The human body and head were extracted from color images and predicted their height based on depth information. Mask R-CNN [203] was used for extracting data from individual frames. Here, height estimation through LiDAR data makes this invention eventually exceeded by all. It convinced most of the shortcomings of traditional RGB and RBD-D-based applications efficiently. In Fig. 5.4, multi-angle ankle-level LiDAR sensors are used for data acquisition. A motion history image based on the LiDAR sensor's distance data is used as an input to our system. We resized all the images as our application that its processing goes unique. A pre-trained ResNet18 model was used for training our model especially binary classification of LiDAR images. It requires a  $224 \times 224 \times 3$  size input image, and 71 deep layers were used for analysis. For cross-validation, we used resnet34 and resnet50 also.

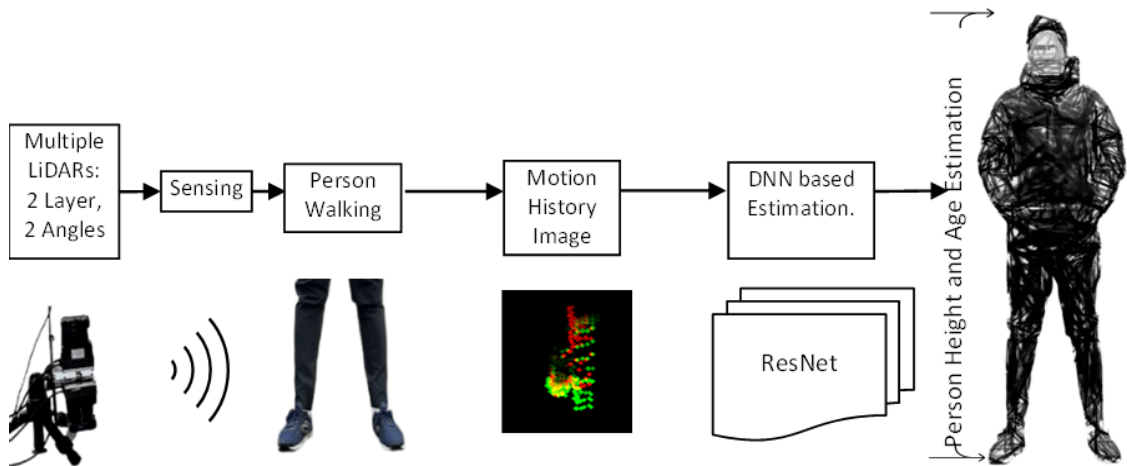


Fig. 5. 4 DNN based height estimation through 2D LiDAR [106]

A prevalent query about using the residual network is what the benefits are. Significantly faster convergence, easy optimization, and significant precision improvements over increased depth make ResNet well accepted by all computer vision researchers. ResNet18 is the best dealing model compared to performance among all other models. We discuss this model in detail here. In Fig. 5.5, a detailed explanation of ResNet18 architecture is explained. We first resize the input images gathered from the rosbag file. A very well-known image size ( $224 \times 224 \times 3$ ) is produced from given images. There are different convolutional layers responsible for filtering the input image. The first convolutional layer (Conv: 1) is accountable for providing low-level features, i.e., edge, gradient, color, etc.

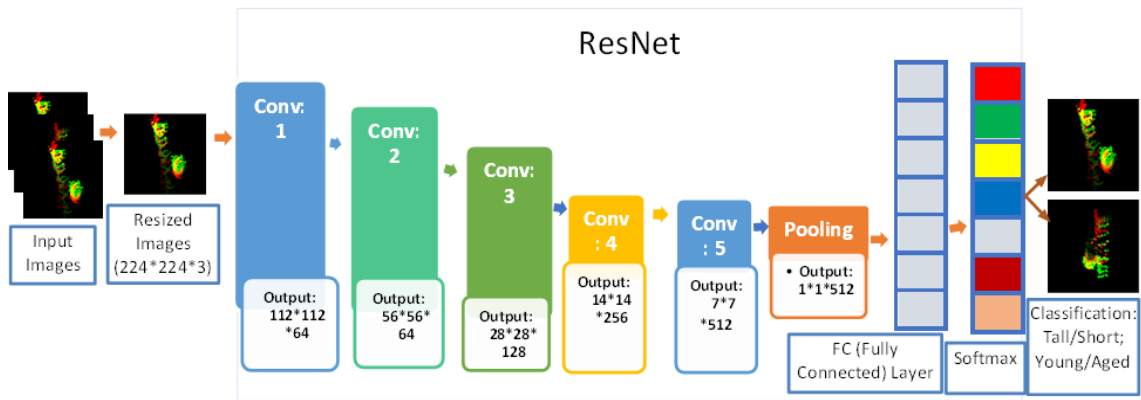


Fig. 5. 5 ResNet structure for the application [106]

The deeper layer provides relatively high-level features. Finally, a feature map is created by convolutional layers to predict the class probability for all gained feature maps. The pooling layer lessens the spatial volume of the convolution elements. The influential details, i.e., rotational, and positional invariants, are also found in the pooling layer. The fully Connected Layer (FC Layer) receives flattened output from the pooling layer and acts as a feedforward network. The SoftMax layer is responsible for a binary outcome. It limits the range within  $[0,1]$ . Here in our application, the results will remain tall and short or aged and young. When we changed binary output into three categories (tall, medium, and short), we changed the SoftMax layer. The promising efficacy of ResNet to incorporate the property of skipping connection to add the output of the previous layer to the next ensures proficiency from beginning to end in deep layers.

### **5.2.2.2 Age Estimation**

A standalone application to estimate a person's age from LiDAR images is unique and might be interesting in different applications. A caregiver company, various public dealings organizations, and transport agencies may need to know the person's age before offering any amenities to its customers. Video-based applications always suffer false identification due to facial expressions and makeup. Our proposed LiDAR-based application can handle these issues effectively. The process of age estimation is relative to height estimation. We prepared our datasets as per the parameter of age. Here we categorize all 29 users into two classes. Age below 30 is considered as young and over, and equal to 30 is elder. These values have been taken heuristically. Other's thresholds also could be taken, and similar results could be found. The same disjoint datasets are prepared for training, testing, and validation. Individual testing sets make the application more robust and accurate. We also performed cross-validation to check the system performance with all data types.

## **5.3 Results and Discussion**

We implemented our experiments on PyTorch Lightning environments where NVIDIA GeForce GTX 1060 GPU system. Other system requirements include an Intel Core i7 processor with 3.8 GHz clock speed, 6GB graphics memory, 64-bit ubuntu 18.04 release, 16GB RAM, and 1TB SSD.

Our dataset consists of different sizes. We used four various LiDAR sensors to collect our data. We plotted different LiDAR data on the image and prepared our motion history image. We were considering different scenarios: two LiDARs data with the same height, two LiDARs data with two layers, etc. In the two-layer dataset, we considered 1,76,678 LiDAR images. We categorized all images into two main subgroups with a threshold of height of 170 cm. Some other subgroups and entries could be taken, and here we categorize them this way. Persons above the size are considered tall, and equal or below are short. In this dataset, we have 89304 images for short people and 87374 images for tall people.

We measured almost 80% of the total images for training our model as a total of 141309 images. Here 71411 images are for short people, and 69898 images are for tall people. We



kept almost 10% of the total images as 17681 for testing our model. Here 8964 images were from short people, and 8717 images were from tall people. Around 10% of disjoint images were considered as validation datasets. Here 8929 images were from short people and 8759 images from tall people. With rigorous training with the ResNet18 model, we found very impressive accuracy there. In Table 5.1, a confusion matrix shows the detail of the system with all disjoint datasets. This matrix is formatted from test data and significantly found its efficacy here. Among all 8717 images of testing-tall data, our system accurately identified 8658 images as tall, and only 59 were misclassified. The same scenario was found for the short dataset also. Here only 62 images were falsely identified out of 8964 images. The accuracy of 99% relieved the cumbersome of previous estimations.

Again, we considered two LiDAR sensors at the same height where the only angle is different. In this condition, we considered 226241 images as a total dataset, with 121244 images for short persons and 104997 images from tall persons. The same 80 percent (181016) of total images were kept as test data. The remaining 20 percent (22632 and 22593) data were almost considered test and validation data equally. All data was disjoint here also, where the same persons were assessed in all three groups (train, test, and validation). We considered the same residual network, ResNet18, for analysis. The right side of Table 1 shows the confusion matrix of test data. Among 10498 images of tall people, our model can accurately identify 10411 images, and for 12134 images of short persons, this can detect 12027 images ideally. The system accuracy is near about 99 percent in total. Both Multi-Layer and Multi-Angle show almost the same precision but combining these two will increase the overall system performance in a complex environment.

Table 5. 1 Confusion Matrix of resnet18 based height estimation [106]

Confusion Matrix (Test)			Confusion Matrix (Test)		
Tall	8658	62	Tall	10411	107
Short	59	8902	Short	87	12027
		Tall      Short			Tall      Short
Height Estimation (Multi Layer)			Height Estimation (Multi-Angle)		

Further, we enhanced our study of height estimation for three categories. Here another class, 'medium,' was introduced. People over 175 cm are considered tall. Between 170 cm to 175 cm was medium, and below 170 cm was short. In our study, we assessed 216474 images of two-layer LiDAR data. 80% of the total (173180) images were considered training data, and the remaining (43294) images were equally split into the test and validation set. We segmentize all images as tall, medium, and short with 59327, 99656, and 57491 images, respectively. The overall accuracy was reduced by a little 4 percent in total by adding a new category. In Table 5.2, We placed all test data in this confusion matrix. Among 5932 images of tall persons, our system accurately identified 5676 images. Some photos were misclassified as tall, but there is no short classification here. The same scenario is in short-type classification. Only 268 images were grouped as medium but not tall. Two types of data bias are the middle class. Among 9966 images, 9523 were accurately categorized, but 330 were identified as tall, and 113 were short.

Table 5. 2 Confusion matrix of three classes of height estimation and age estimation [106]

Confusion Matrix (Test)				Confusion Matrix (Test)		
Tall	5676	330	0	Young	11535	412
Medium	256	9523	268	Elder	483	9217
Short	0	113	5481		Young	Elder
	Tall	Medium	Short	Age Estimation		
	Height Estimation					

Another property of humans is the age that we estimated through 2D LiDAR data. For this purpose, we considered 216474 images as an 'Age Dataset.' In this research, people over or equal to 30 years were considered elder aged, and below 30 years were deemed to be young—a total of 96295 images for older people and remaining images for young people. As the previous ratio total of 80 percent of images was set for training, the remaining 20 percent were equally distributed as training and validation. On the right side of Table 2. we see that among 12018 images of young people in the test dataset, 11535 images could be correctly identified by our model. For older people, 9217 images have been correctly

identified, while only 412 were misclassified. The overall precision of this model is approximately 95 percent.

Table 5.3 describes the overall properties of the experiments that we have performed in this research. We considered different batch sizes and ten epochs for all cases for our investigation. Here train, test, and validate accuracies showing system performances. On the other hand, different loss functions show system integrity and robustness with the datasets. This research emphasizes the best use of 2D LiDAR data for person property estimation, which could be an excellent alternative to RGB and RGB-D cameras.

Table 5. 3 Overall system performance with ResNet18 network [106]

Data	Experiments Type	Batch Size	Epoch	GPU	Model	Train Accuracy	Train Loss	Test Accuracy	Test Loss	Validation Accuracy	Validation Loss
New data 2 Layer LiDAR	Height 3 Category	24	10	Yes	Resnet18	0.963	0.0921	0.9599	0.1063	0.9593	0.1028
New Data 2 Layer LiDAR	Age 2 Category	34	10	Yes	Resnet18	0.959	0.0971	0.9575	0.1032	0.9546	0.1107
New Data 2 Layer LiDAR	Height 2 Category, Complete Random	34	10	Yes	Resnet18	0.9939	0.0173	0.9937	0.0178	0.99382	0.01776
New Data 2 Angle LiDAR	Height 2 Category, Complete Random	34	10	Yes	Resnet18	0.997	0.00976	0.9958	0.0125	0.99552	0.01187

## 5.4 Conclusion

Person property estimation is always challenging and sometimes crucial in different circumstances. Many efforts have been initiated on this topic from the beginning of computer vision. Simultaneously some applications were developed with the help of the LiDAR sensor. We concentrated on the amalgamation of these two phenomena. To find a suitable alternative of a video camera and improve the property estimation accuracy bring we out to do so. Our LiDAR-based person property estimation, especially introducing a

deep residual network, gives a well-accepted benchmark. Preparing a new dataset and finding its individual properties makes this study abundant. This dataset can be used for other 2D LiDAR-based research. In the future, we will try to enhance our study to find more properties of a person that he/she can be accurately traced. We will combine our previous tracking system with this property estimation technique to develop a LiDAR-based autonomous system. Our impending focus is on group recognition and predicting their behavior without compromising individual identity.

## Chapter 6

# Person Identification by Evaluating Gait using 2D LiDAR and Deep Neural Network

### 6.1 Introduction

Person identification is a vast and ancient research field. Various techniques have been invented in this arena. Diversified biometric [204,205,206] characteristics enhanced the accuracy of this vital space. Camera-based applications eventually led to this research. Sophisticated innovations, modernized features, and computational capabilities make this exact day by day. But privacy, Lighting issues of a camera, disasters, etc., are a big concern for video-based processing. The emergence of new biometric features meticulously crafted the credibility of human recognition in different delicate applications. Analogously, the urgency of close contact with the devices down worth the performance of biometric identifications in some cases. Hence gait recognition is an apt alternative to person identification where subjects are not supposed to be nearer to devices. In all these innovations, video cameras [205] were used as key identifiers to demonstrate individuals' states and distinctiveness.

Gait is a way of walking, and gait recognition refers to identifying a person based on their walking style. Some crucial circumstances and feasible benefits over traditional camera-based applications make Gait more popular nowadays. A very plausible gain 'remote access' over close camera contact is well accepted. Furthermore, without the cooperation of the subject, it can be recognized. Even if some biometric features (i.e., face, fingerprint, iris,

etc.) are absent or cannot be identified, Gait plays a vital role in identification. It is tough to impersonate the gait features, making it essential in crime analysis. However, video-based recognition suffers from low-resolution capturing, disasters, computational complexities, etc. In disasters, typical video-based applications fail to capture images due to darkness, smoke, fog, or obstacles. LiDAR sensors are free from these difficulties. This research proposes a new modality: ankle level 2D LiDAR-based person identification with gait data analysis. Recent enhancements in computational capabilities and a deep learning approach make this research more precise than ever.

Our previous study demonstrated a person tracking [93] system using 2D LiDAR at ankle level for gait analysis. Determining moving objects in front of the LiDAR sensor as a person was a challenging job. We fastidiously made the way of tracking by imposing density-based clustering over conventional approaches. This research started using multivariate density-based algorithms to fit the model best. A visualized tracking based on LiDAR data was undoubtedly challenging, and we prudently did the same. Developing density-based algorithms to augment the tracking performance was another challenge in this experiment. Our proposed two new algorithms [94]: Enhanced Density Based Scan (EDBSCAN) and Enhanced Ordering Points to Identify the Clustering Structure (EOPTICS) to create the best cluster to determine an individuals' ankle positions and identify their way of walking. This approach enhanced the performance of our previous tracking system and showed a new person tracking system based on only 2D LiDAR data. This method also helped us find an alternative to surveillance cameras for which people were worried about their privacy and secrecy. Here, the necessity of tracking and confidentiality intent meets a unique solution. Hence, some influential features (i.e., age, height, etc.) can be measured by only ankle movement data, especially tracking data. We have broadened our study with person property estimation by a 2D LiDAR sensor [106]. Here a deep neural network was used for training and testing the model. A detailed dataset of the experiments with different ethnicities, gender, and height were prepared to conduct all experiments. A parametric formulation was performed to do the experiments, and the results were distinctly identified. The outcomes of the trials were remarkable and trustworthy compared with the actual ones.

Some practical instincts suffer from RGB/RGB-D cameras for illusion, illumination, smoky or foggy conditions, and even real-time computational inaccuracies. LiDAR-based person identification is our new research goal that comprehensively deals with all inadequacies of visual imaging. Placing multiple 2D LiDAR sensors at the ankle level to acquire data enhances data detail to analyze studies further. LiDARs emanate pulsed light on the surrounding objects and compute the distance it travelled to receive the sensor again. An experimental model was set, where different LiDAR sensors were placed in the grounded tripods, and persons were let to walk in front of the sensors. A Robot Operating System (ROS) was used to capture the time series data in a bag file. Those distance data were plotted in images with a specific rate to create motion history images (MHI). These MHI were key inputs of our DNN model to identify an individual in a video. Continuous ankle movements on a surface develop a path of walking, which determines the tracking system efficiently. As all persons have distinct gestures and ways of walking, especially activities of ankles are significantly unique, diverse us to develop a gait-based person recognition system.

In figure 6.1, a block diagram is shown that clarifies the system overview in brief. A sensing module collects data from the LiDAR sensor and passes these distance data to create Motion History Images. These are our input datasets. These Image datasets were used to estimate the person property [106] through a residual neural network. The key foundation of this research was a previously developed tracking system [93] based on modified density-based clustering approaches EDBSCAN and EOPTICS [94] and a property estimation technique. Now it is the step-up of the recognition system to identify a person through gait analysis.

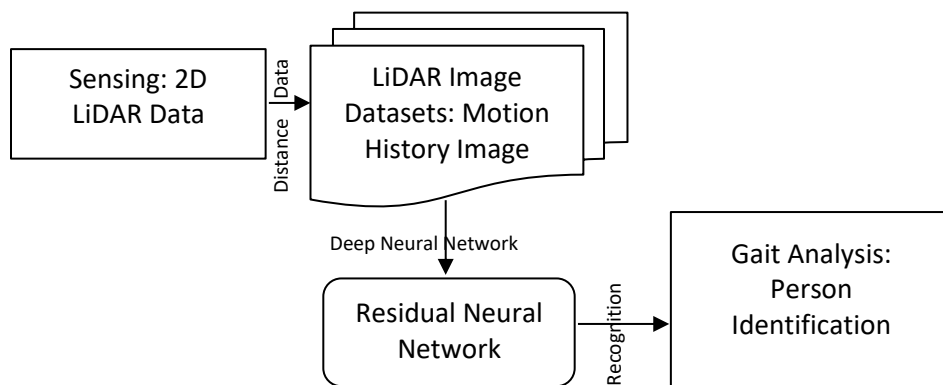


Fig. 6. 1 A system overview of 2D Lidar-based Estimation

## 6.2 Proposed Method

The fundamental concept of this research is to introduce 2D LiDAR sensors as an alternative to 3D Lidar sensors that comprehensively minimize the intrinsic and computational costs and enhance the system integrity on a big scale. Our previous proposed methods were a step-by-step enhancement of LiDAR-based individual following and property estimation. We were sensible to use video cameras, often liable for compromising privacy issues. Moreover, some natural and environmental shortcomings depleted the performance of RGB/RGB-D cameras.

### 6.2.1 Integrated System Overview

In figure 6.3, an overall system diagram is presented here. 2D LiDAR sensor is placed at ankle level to acquire the data. All time-series data are plotted on blank images at 40 frames per second rate, named 'Motion History Image (MHI).' The main benefit of MHI is to encode a range of time data in a single frame. Thus, human gestures and movements can be represented by MHI spans [33]. An update function  $\mu(x, y, t_i)$  could calculate the MHI  $M_{\mathcal{E}}(x, y, t_i)$ :

$$M_{\mathcal{E}}(x, y, t_i) = \begin{cases} \mathcal{E}, & \text{if } \mu(x, y, t_i) = 1 \\ \max(0, M_{\mathcal{E}}(x, y, t_{i-1}) - \varphi), & \text{otherwise} \end{cases} \quad (6.1)$$

Here,  $(x, y)$  is the position and  $t_i$  is the time;  $\mu(x, y, t_i)$  shows ankle position or motion in the current frame. The temporal extent of the movement is decided by the duration  $\mathcal{E}$ , where  $\varphi$  indicates decay in the images. Leaving past images as afterimages makes it easy to understand the time series data, thus creating the Motion History Image shown in figure 6.2.

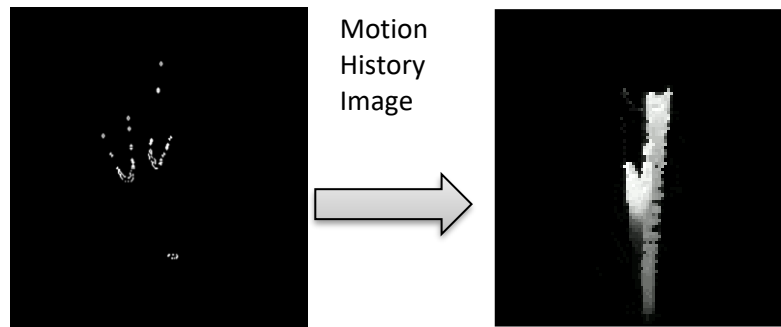


Fig. 6. 2 Creation of Motion History Image



These MHI are considered as input of our system. Besides traditional clustering approaches, we used modified density-based clustering techniques to determine an ankle of a person. Similarly, the same clustering approaches were used to determine a person based on closely moving two ankles on the plane. This heuristic approach was analyzed repeatedly until it came to tolerable accuracy. This study describes a gait-based identification based on 2D LiDAR only. Identifying a person using only a 2D LiDAR sensor is a novel approach and can be effectively applied to privacy issues.

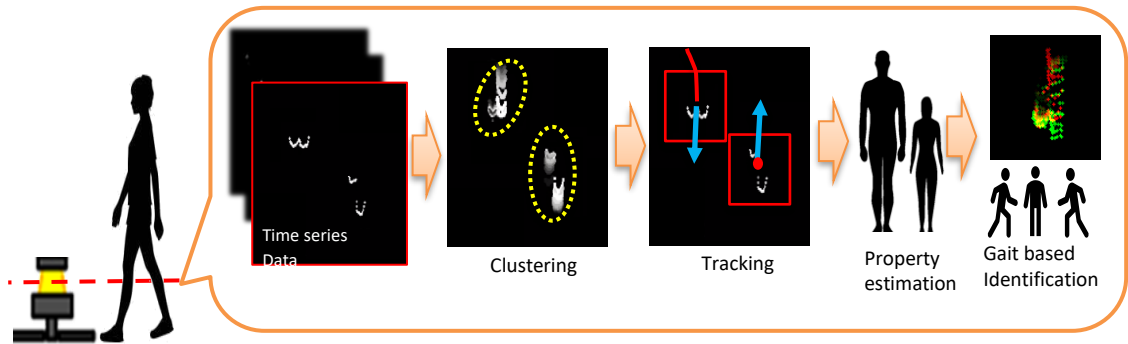


Fig. 6. 3 LiDAR-based Person Tracking, Property Estimation, and Recognition

## 6.2.2 Gait-based Person Identification

Person identification deals with much research over so many years. Other biometric features-based identification made this research more authentic and accurate. Various types of cameras and sensors were used to make this perfect. With the invention of the deep neural network, this research got a new dimension. Our focus on this research is to contribute differently to the sense of sensor and calculation.

### 6.2.2.1 Experimental Setup

Figure 6.4 shows the experimental setup for the research. We put four LiDAR sensors in the two stands at different angles and heights so everyone's data could be easily collected. Two LiDAR stands are placed at a two-meter distance, and the angled gap is 90 degrees. The LiDARs position was ankle-level, where one was six-inch up from the ground, and another was ten-inch up. We named two LiDARs in the same tripod as a multi-layer and LiDARs in a separate tripod named multi-angle. Participants were allowed to walk freely, with their typical style and motion. All the experiments were conducted indoors and in the

same season. Participants walked back and forth, round, and 0.5 to 25 meters range from the LiDARs. We considered LiDARs positions stationery, and participants are moving. Individual walking and group walking were considered in the experiments. For the analysis, we used different sensors' data in different ways. For this experiment, we used HOKUYO UTM-30LX 2D LiDAR sensors. It has a 30-meter and 270-degree scanning range. This sensor is lightweight and very suitable to use outdoors. We varied the room's lighting conditions and allowed participants to walk to and fro at different speeds. On average, 10 minutes of all pedestrians were asked to walk, move, or run to make the data more informative and detailed to analyze.

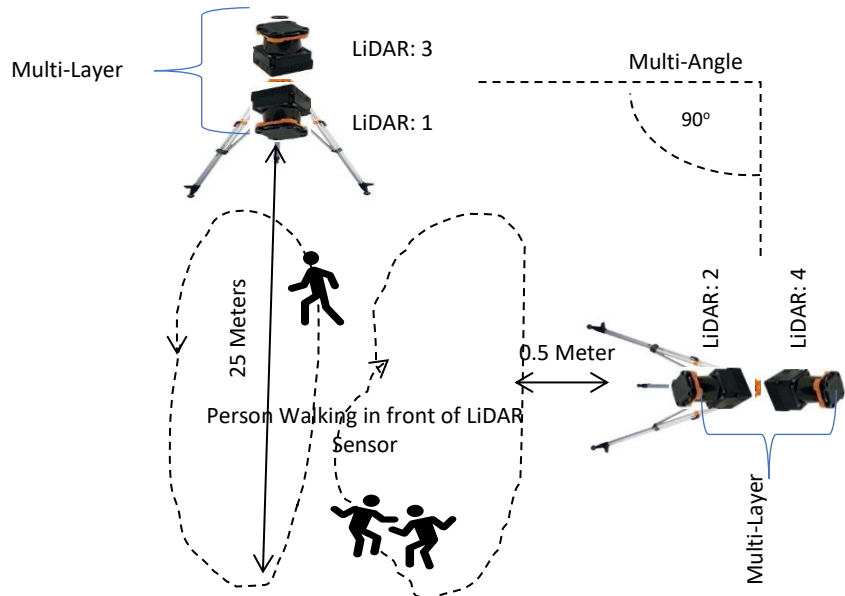


Fig. 6. 4 Experimental setup for Person Identification

### 6.2.2.2 Dataset Preparation

In this research, we used our developed dataset. No 2D LiDAR-based identification systems were previously found, so no public dataset is available. 2D LiDAR sensors only provide distance values of objects in front of it. Thus, making our image dataset challenging. We created different datasets based on enormous parameters and named these KoLaSU (Kobayashi Laboratory of Saitama University). Twenty-nine independent observers participated in our study, and all were unbiasedly allowed to walk in front of the

LiDAR sensors. Further, this well-structured dataset considers different international audiences perfectly diversified in height, age, gender, and ethnicity. In this study, 18 Bangladeshi, 10 Japanese, and 1 Filipino participated, where seven were female and 19 were male. Of 29 participants, three walked in sandals, and the rest of the participants with shoes. To make the data uniform, we skipped sandals data. For the cross-validation of the performance, those data are also used in rigorous analysis. We created one aggregated dataset considering all four LiDAR sensors even with individual four sensors' data. After that, we created a multi-layer and multi-angle dataset considering all possible conditions, i.e., multi-layer-13, multi-layer-24, multi-angle-12, multi-angle-14, etc. Here 1,2,3, and 4 indicate LiDAR position as per figure 4.

Among 26 participants, 12 people were above or equal to 170 centimetres, and the rest were below 170 cm. 17 people were below or equivalent to 30 years, and others were above 30 years. Figure 6.5 shows a brief description of our prepared dataset. Here seven chronological movements of two participants are shown with our created images. Different colours in the MHI indicate various LiDAR sensors' data in the images. Different colours lines in the images show ankle positions captured by different LiDAR sensors. In this datatype, three LiDARs data were fused into a single LiDAR image. All individuals' records were created at an average walking speed at a 40-fps rate in the MHI. But we considered a 100 FPS rate for running and fast movement pedestrians, which applies to coping with various data. A common query may arise of using multiple sensor setups and a combination of data. It is always challenging to get much information using 2D LiDAR only. When a single LiDAR is used to scan the person, it may not be well suited to occlusion or fast-moving.

Furthermore, it isn't easy to distinguish individuals from a single LiDAR view. If the multi LiDARs data were plated in a single image, data variety increased, providing much accuracy in the system. We try to fuse different combinations in our study to examine the system's best performance, and literally, it enhanced the system's performance significantly.

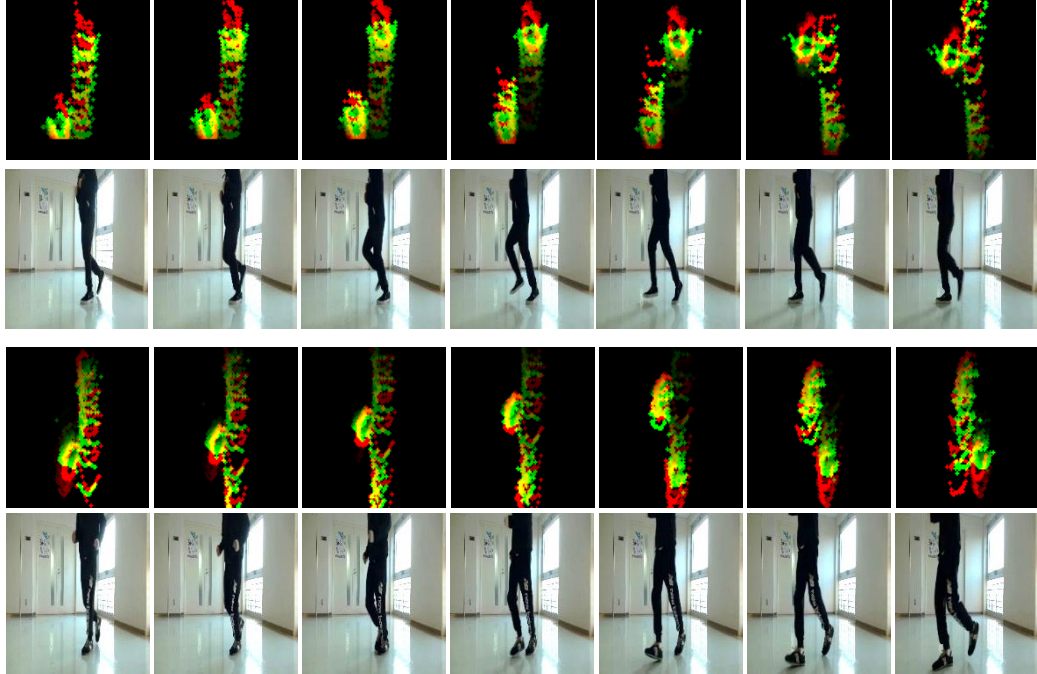


Fig. 6. 5 KoLaSU, Two Persons' data: Upper one is MHI and lowers one is the corresponding pose

### 6.2.2.3: ResNet based identification

Figure 6.6 shows detail of gait-based identification. As mentioned, motion history images (MHI) were created from LiDAR data as a neural network input. We categorized different ways and trained our model for the experiments. A pre-trained residual neural network (ResNet-18) was applied to validate and test the system. A pre-trained version of the network was loaded to train this KoLaSU dataset that was previously trained with more than one million images of ImageNet datasets. The benefit of using a pre-trained network is that it learned much feature representation with a huge set of diversified images. Identity mapping vectors and residual learning extract convolutional features by training a ResNet. In a ResNet architecture, residual can be defined by

$$Y = f(x) + x$$

Here,  $x$  is the input vector,  $Y$  is the output vector, and  $f(x)$  is the residual mapping function. In this article, we used 18 layers and 50-layer ResNet to extract the convolutional features of KoLaSU datasets. The second layer of the network is Max pooling comes after the first convolutional layer is used to minimize the data overfitting problem. A Fully Connected (FC) layer combined with Average Pooling and SoftMax layers provides features of human detection based on gait data. Without raising the training error rate, a substantial number

of layers can be trained by ResNet easily. ResNet is also competent in nullifying the vanishing gradient problem, giving it an extra advantage over other traditional neural networks. Finally, a consistent gait-based classification was done with meaningful accuracy.

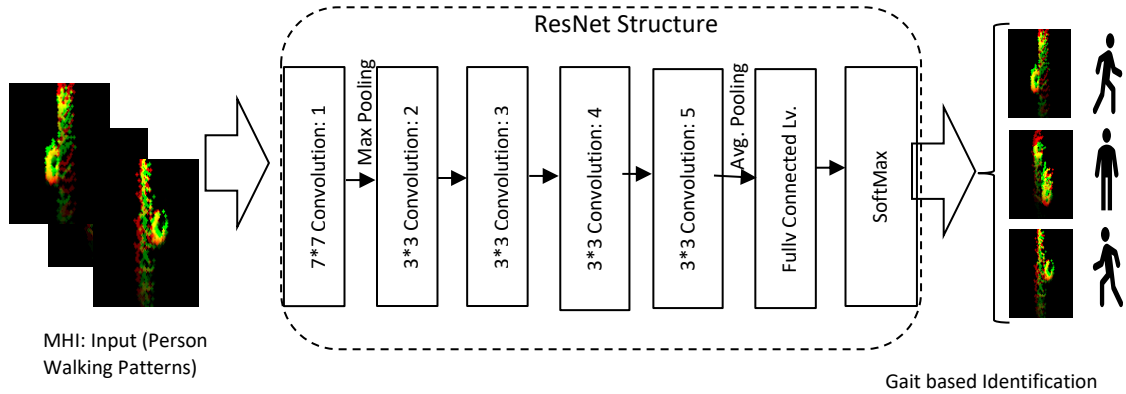


Fig. 6. 6 Gait based Person Identification

## 6.3 Experiments and Discussion

### 6.3.1: Gait-based person identification

We thought of the different data combinations during acquisition to make the application-wide in range and reliability. A homogeneous and heterogeneous LiDAR setup was considered for the collection of data. Though a single LiDAR is sufficient to get a pedestrian's desired data, critically distributed and overlapped data always dilute system performance. Ankle-level LiDAR setup was our primary focus to track and identify a pedestrian. We substantially found numerous walking styles based on their properties. Plotting LiDAR data on an image was our primary challenge, but we cautiously did the experiments and found excellent accuracy there. Our proposed KoLaSU dataset consists of fourteen outdoor sequences where twenty-nine participants attended with five to ten minutes of walking with the experiment setup. We considered a standard forty fps rate to write the data. For cross-validation, we considered the 100-fps rate also. In all our experiments, we kept our datasets into three groups. Sixty percent of the data were considered train data, while the remaining 40 percent were split equally into test and

validation sets. Different phases of data segmentations also were experimented. In another experiment, 80 percent of the total data was kept in the training group, where the remaining twenty percent were tested and validated equally. We cross-validated the data in different phases and augmented the train data to enhance the system's cumulative performance. The experiments in Figure 6.7 show the results mentioned above.

Table 6.1 shows an overall experimental result of gait-based recognition. There are 14 different conditions data in the KoLaSU person tracking dataset; nine were considered here. We split out all four LiDAR data individually as we kept four top rows in table 1. Similarly, in the table, we merged different LiDAR data as the number assigned (i.e., LiDAR 12, merged LiDAR 1 and LiDAR 2's data in MHI with 40 fps). It is highly needed to make a dataset versatile to combine multifaceted data. In our previous studies, we used only single LiDAR for data acquisition. But dealing with occluded data and handling group people is not easy to get accurate outcomes from there. Furthermore, data from different angles and layers enhance the dataset's credibility. It covers all wide angles, and every individual can be traced accurately. It also ensures that it does not necessarily deploy all these four LiDAR sensors in a practical application. To make an original dataset, we focused on credibility rather than cost. We used the GIGABYTE BRIX GPU machine for this research to analyze our data. The batch size was considered 38, and the number of epochs periodically varied from 25 to 50. We used a deep neural network to train our model. Here a pre-trained ResNet 18 network was used to train the dataset, especially for binary classification of LiDAR images. It requires a  $224*224*3$  size input image, and 71 deep layers were used for analysis.

Here, ResNet 50 was also used for cross-checking the system performance. Usually, the deeper model performs better but computational time is an issue that cannot be ignored. We used different-sized ResNet models to justify the system's credibility in the cross-validation section. We placed some results of these in the next section. We randomly selected our train, test, and validation datasets among all machine-generated data, where all segments were utterly disjoint. Initially, we kept every person's data in three parts: train, test, and validation group. Further, we enhanced our study for unknown test data sets. Without prior information, man or machine cannot identify a new individual; thus, the system reacts. From table 6.1, we see that accuracy in three different segments is very

impressive and near about 99 percent correctly identified. Though some data are not accurately captured due to congestion problems (i.e., KoLaSU LiDAR 3 and KoLaSU LiDAR 4), their performance in the test case did not go below 93 percent.

Here, height is the congestion problem that affects performance. Because this system performs well at a 6-inch height, the data detail decreases with an increase in the LiDAR height. Moreover, their combined dataset (KoLaSU LiDAR 34) performed significantly well with 99 percent precision. This is the initial step of introducing only 2D LiDAR for person identification. This study emphasized accuracy as performance, though previous studies applied other precisions. We used a pre-trained model (ImageNet) to classify our model with the best precisions. The accuracy above 90 percent was considered accurate in our analysis.

Table 6. 1 Gait-based person identification on different parameters

<b>Data</b>	<b>Experiments Type</b>	<b>Train Accuracy</b>	<b>Test \Accuracy</b>	<b>Validation Accuracy</b>
KoLaSU LiDAR 1	26 Persons Individual (60%,20%,20%)	0.99421	0.9843	0.9851
KoLaSU LiDAR 2	26 Persons Individual (60%,20%,20%)	0.99324	0.9846	0.98502
KoLaSU LiDAR 3	26 Persons Individual (60%,20%,20%)	0.97721	0.9336	0.93478
KoLaSU LiDAR 4	26 Persons Individual (60%,20%,20%)	0.98038	0.9479	0.94711
KoLaSU LiDAR 13	26 Persons Individual (60%,20%,20%)	0.9982	0.9962	0.996
KoLaSU LiDAR 24	26 Persons Individual (60%,20%,20%)	0.99831	0.99622	0.99611
KoLaSU LiDAR 12	26 Persons Individual (60%,20%,20%)	0.99719	0.9934	0.99305
KoLaSU LiDAR 34	26 Persons Individual (60%,20%,20%)	0.99821	0.99343	0.9942
KoLaSU LiDAR 1234	26 Persons Individual (60%,20%,20%)	0.99869	0.99658	0.99713

### 6.3.2 Comparison of different data types

To reduce the overfitting, we validated the network, and here its accuracy is impressive, and none of the datasets goes below 93 percent. The test accuracy also goes through with validation accuracy and follows its footsteps. Accuracies and losses are inversely proportional in a system. Our system is also showing this trend. This designed network and its performances instinctively develop a logical ground for using two-dimensional LiDAR sensors for individual identification in broad.

We performed rigorous testing with different datasets to test the system's performance. All fourteen datasets were considered for this cross-testing. We analyzed the results and found an essential symmetry in the performance analysis phase, and those wholly aligned with our theoretical expectations—figure 6.7 shows these results in detail. Suppose the top four datasets in the figure are KoLaSU LiDAR 24 and 13, respectively. We trained and validated the system with the same dataset but changed the test data only in four cases. LiDAR 24 and LiDAR 13 are created by merging sensors 2 and 4 and 1 and 3, respectively. For testing, we used only LiDAR 4 and 2 and LiDAR 3 and LiDAR 1 separately. Though training and validation accuracy is nearly absolute, the bar chart shows that test accuracy goes below 20 percent. We considered unbiased disjoint data for all cases. As the figure shows, the system performance will degrade if a person is accurately trained by the neural network and tested differently. The same performance persists for all the cases except the combined dataset LiDAR 1234 is tried with test data 24 and 13. Here the performance reached up to 38 percent but was not impressive. So here, we can conclude that to achieve the system's best performance of the system; it should be trained and tested with the same types of data; any other bias is not necessary there.



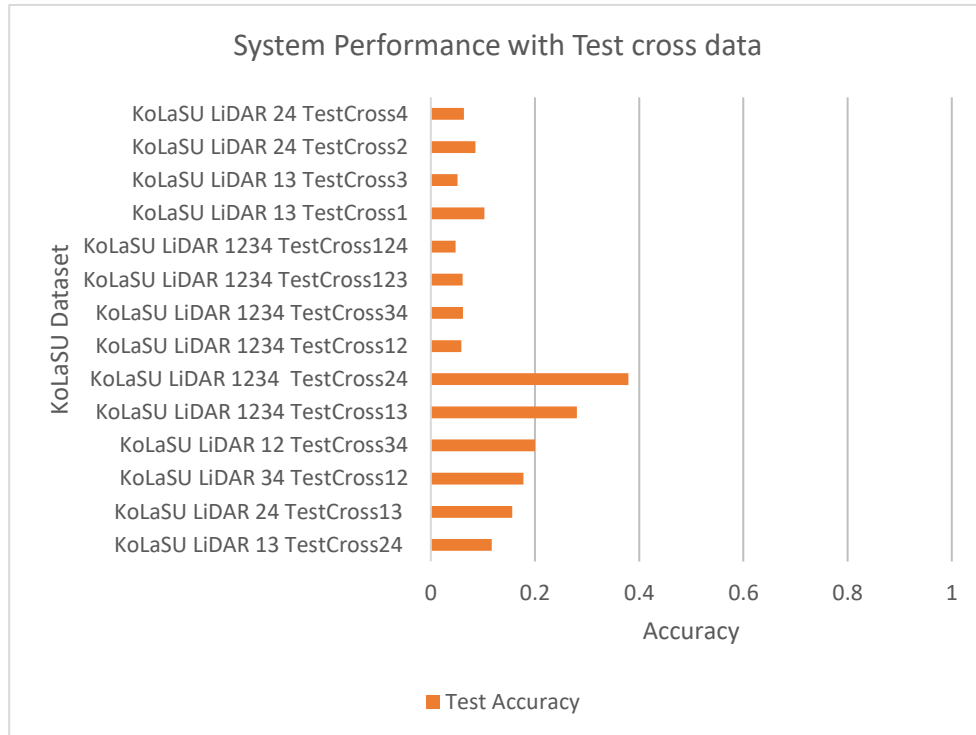


Fig. 6. 7 Cross Validation: Gait Performance test with Cross-Data

Table 6. 2 Performance test with different DNN model

Data	KoLaSU LiDAR 1234 TestCross24	KoLaSU LiDAR 1234 TestCross24
Experiments Type	26 Persons Individual (60%,20%,20%)	26 Persons Individual (60%,20%,20%)
Batch Size	38	38
Epoch	25	40
GPU	Yes	Yes
Model	ResNet 18	ResNet 50_2
Train Accuracy	0.99864	0.99999
Train Loss	0.00589	0.000354
Test Accuracy	0.379	0.4007
Test Loss	4.2741	3.5852
Validation Accuracy	0.99721	0.99956
Validation Loss	0.00589	0.001578

Besides ResNet 18, we analyzed different neural networks to test our data's system performance and effectiveness. We show one such type of analysis in table 6.2. For the same dataset KoLaSU LiDAR 1234, we trained and validated it by ResNet 18 and ResNet 50. The rest of the parameters remained the same except for epoch size. We tested the system with different datasets KoLaSU LiDAR 24 and tried to analyze the performance of two separate networks. ResNet 18 gave almost 38 percent accuracy, whereas ResNet 50 performed with 40 percent accuracy. But the network size of ResNet 50 (Layer 50) is more abruptly huge than ResNet 18 (Layer 18), and computation time is highly excessive (app. four times) in these experiments. Thus, this study decided to consider ResNet 18 rather than ResNet 50 even though its performance is little improved.

To test the system performance differently, we combined different datasets and trained and validated our system. Our achieved accuracies were impressive in all cases. In figure 6.8, we placed some experiments based on combined datasets. Here we put system accuracy and loss together. Six datasets related to their aligned ones. Suppose LiDAR 24 data was combined with LiDAR 2 and 4 data for training and validation of the system. Further, we tested the system individually with LiDAR 24, 2, and 4 data. The same scenarios were performed with LiDAR 1234, 13, and 24. Though the system was trained with multiple groups of data and tested with individual ones, it was previously performed as a regular system. Figure 8 shows that test accuracy lies around 99 percent, whereas its loss remains minor as below 20 percent. Here, accuracy and loss curves also follow asymmetry, which indicates network performance.

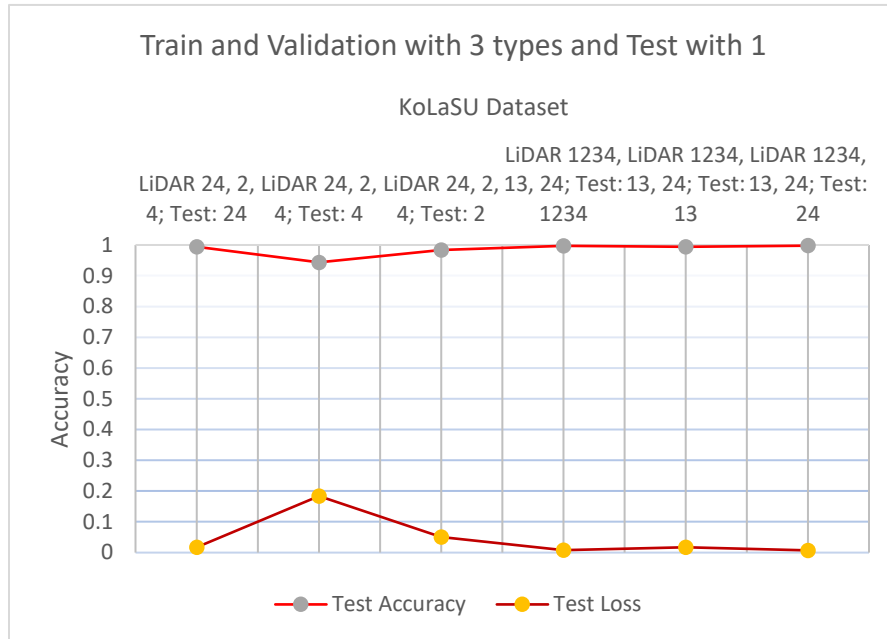


Fig. 6. 8 Combined dataset performance analysis

### 6.3.3 Comparison with contemporary studies

To the best of our knowledge, no such research was performed with 2D LiDAR sensors to identify a person based on gait analysis. There were some great initiatives conducted with a 3D LiDAR sensor. All the sensor setups, experimental complexities, methods, and even datasets were different, so an accurate comparison cannot be made with actual precisions. A conceptual description is given in figure 6.9. Benedek et al. [104] initiated the research for lidar-based gait analysis. They prepared their dataset, SZTAK-LGA, with 28 participants. They used CNN (convolutional neural network) and MLP (multi-layer perceptron) for training and testing the system. They used different people in their experiments; an increased number of people degraded the system performance from 92 percent (for five people) to 75 percent (for 28 people). Yamada et al. [103] performed a thorough experiment on lidar-based gait analysis. This research was also conducted with a 3D LiDAR sensor. They also prepared their dataset, PCG (point cloud gait), with 30 participants. A CNN and LSTM (long short-term memory) neural network model was applied for training and testing the system. Though they get different accuracies in different input patterns, here, l=1:8 gave a maximum of 72 percent in general. As datasets and scenarios are entirely different, even sensors and methods are separate, so the accuracies

cannot be compared eventually. We randomly used utterly unbiased datasets categorized into three classes (train, test, and validation). The average system performance is greater than 98 percent of expected training data, emphasizing a wide use of 2D sensors in different applications.

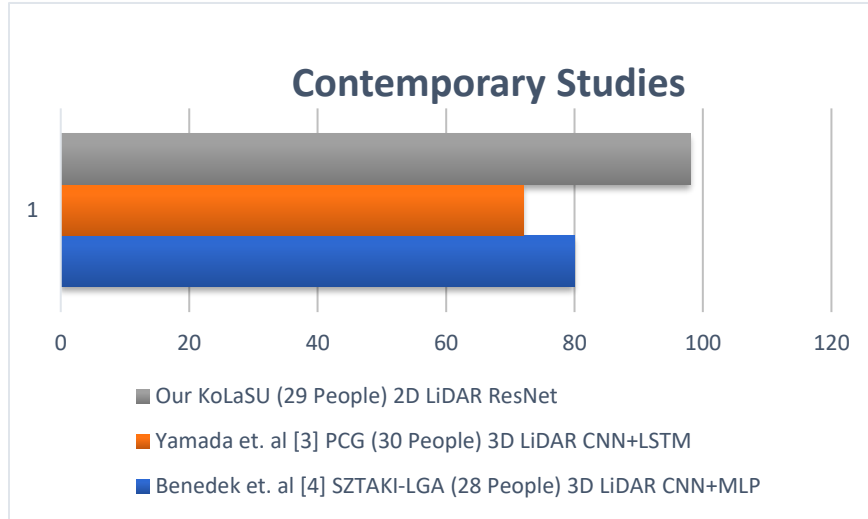


Fig. 6. 9 Contemporary studies with state-of-the-art technologies

## 6.4: Conclusion

This paper presents a way of person identification with a LiDAR sensor. Very little research has been done in this arena those were used only in this sensor setup. We tried to elaborate the use of 2D LiDAR sensors with real-time calculations and enhanced accuracies. A pre-trained deep neural network, ResNet effectively fit the data and found its losses for gait recognition. The overall precision of the system is awe-inspiring. A combined application of person tracking and identification with this sensor could be applied in autonomous robot movements.

# Chapter 7

## Conclusion and Future Work

### 7.1 Conclusion

This thesis demonstrates a comprehensive understanding of people's behavior apart from traditional sensor-based applications. When recent studies have bottleneck understanding of privacy and security, our LiDAR-based solution carefully addresses the shortcomings and the possibilities. A generous surveillance system could handle all its necessities effectively, but privacy remains unhampered. This dissertation showed the multifaceted functions of a 2D LiDAR-based monitoring system.

We rigorously exhibited the tracking system that could be a showcase for privacy-measured ideal people tracking system. Ankle scan with 2D LiDAR effectively understands people's behavior with reasonable accuracy. Besides conventional hierarchical clustering algorithms, density-based algorithms ensure the best classifications, especially in LiDAR-dense data. Real-time processing, low installation cost, and impartial sensing make the model authentic and applicable. Though its precisions were impressive, some constraints downgrade the tracking performances in occlusion, over-dense, or under-dense areas. We proposed two density-based clustering algorithms (EDBSCNA and EOPTICS) that could be applied independently to classify the LiDAR data. These two algorithms nicely deal with data overfitting and underfitting.

To enhance the usability of the LiDAR-based applications, we improved our study to estimate the properties of pedestrians. This study could distinguish people based on height (tall and short) and age (elder and young). Dataset of 2D LiDAR with newly developed

scanning system shows feasibility for Property estimation. We additionally required a next-stage solution for subcategory classification with new data collection. These property estimations are also essential in defense recruitment, clothing, and age-based sports. People identification is another integral research domain in recent person Re-ID. We extended our study for individual identification through Gait estimation with our new scan data. We considered people identification based on 2D LiDAR data, which must not confuse with specific identity disclosure. We emphasized identifying a person moving different positions on the premises. A multi-LiDAR sensor setup was considered to get different angular data that enhanced the reachability of the dataset. We are trying to summarize the study with group pedestrian recognition, which efficiently solves most camera issues and can be used in support robot applications.

Finally, we can conclude that with the emergence of video camera-based applications in surveillance systems, its drawbacks also should be optimized. Besides privacy, disasters and even lighting restrains could be solved by LiDAR-based applications. Our proposed 2D Lidar-based monitoring system is a promising novel innovation in people recognition, a parallel or standalone application.

## **7.2 Future Work**

This research covered four main properties of a person's behavior understanding through LiDAR. Video or 3D LiDAR-based applications deal with more detailed data, quickly determining people's properties. We faced some limitations in continuing our study. In the future, we will try to overcome the following issues.

- In the future, we will try to enhance our study to find more properties of a person so that he/she can be more accurately traced.
- Our impending focus is on group recognition and predicting their behavior without compromising individual identity.

Finally, we want to enhance our dataset and make those accessible so that this 2D LiDAR-based behavior understanding research will proceed next.

# Publication List

## Published Journal:

- (1) **Hasan, M.**, Hanawa, J., Goto, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: Tracking People Using Ankle-Level 2D LiDAR for Gait Analysis. In: Ahram T. (eds) Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2020. Advances in Intelligent Systems and Computing, vol 1213. Springer, Cham. (2020).
- (2) **Hasan, M.**, Hanawa, J., Goto, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: Person Tracking Using Ankle-Level LiDAR Based on Enhanced DBSCAN and OPTICS. In: IEEJ Trans Elec Electron Eng. (2021).
- (3) **Hasan, M.**, Goto R., Hanawa J., Fukuda H., Kuno Y., Kobayashi Y.: Person Property Estimation Based on 2D LiDAR Data Using Deep Neural Network. In: Huang DS., Jo KH., Li J., Gribova V., Bevilacqua V. (eds) Intelligent Computing Theories and Application. ICIC 2021. Lecture Notes in Computer Science, vol 12836. Springer, Cham. (2021)
- (4) **Hasan, M.**, Hanawa, J., Goto, R., Suzuki, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: LiDAR-based Detection, Tracking, and Property Estimation: A Contemporary Review. In: Neurocomputing. 506. 393-405. (2022).

## Under Review:

- (1) **Hasan, M.**, Hanawa, J., Goto, R., Suzuki, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: Person Identification by Evaluating Gait using 2D LiDAR and Deep Neural Network. (Under Review)

# References

1. Iguernaissi, R., Merad, D., Aziz, K. et al. People tracking in multi-camera systems: a review. *Multimed Tools Appl* 78, 10773-10793 <https://doi.org/10.1007/s11042-018-6638-5> (2019)
2. S. Kumar, T. K. Marks and M. Jones, Improving Person Tracking Using an Inexpensive Thermal Infrared Sensor, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 217-224, doi: 10.1109/CVPRW.2014.41. (2014)
3. C. Conaire, E. Cooke, N. O'Connor, N. Murphy, and A. Smeardon. Background modelling in infrared and visible spectrum video for people tracking. In *CVPR Workshops*, pages 20-20, (2005)
4. C. Dai, Y. Zheng, and X. Li. Pedestrian detection and tracking in infrared imagery using shape and appearance. *Computer Vision and Image Understanding*, 106(2):288-299, (2007)
5. E. Goubet, J. Katz, and F. Porikli. Pedestrian tracking using thermal infrared imaging. *Infrared Technology and Applications XXXII*, pages 62062C-1, (2006)
6. A. Leykin and R. Hammoud. Pedestrian tracking by fusion of thermal-visible surveillance videos. *Machine Vision and Applications*, 21(4):587-595, (2010)
7. D. Olmeda, A. de la Escalera, and J. M. Armingol. Contrast invariant features for human detection in far infrared images. In *Intelligent Vehicles Symposium (IV)*, IEEE, pages 117-122. IEEE, (2012)
8. P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *CVPR*, pages 304-311, (2009)
9. Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In *CVPR*, pages 2411-2418, (2013)
10. B. Babenko, M. Yang and S. Belongie, Robust Object Tracking with Online Multiple Instance Learning, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619-1632, doi: 10.1109/TPAMI.2010.226. Aug. (2011)
11. M. J. Black. EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation. *IJCV*, 26(1):63-84, (1998)
12. R. T. Collins, Y. Liu, and M. Leordeanu. Online Selection of Discriminative Tracking Features. *PAMI*, 27(10):1631-1643, (2005)
13. S. Hare, A. Sári, and P. H. S. Torr. Struck: Structured Output Tracking with Kernels. In *ICCV*, (2011)
14. J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. In *ECCV*, (2012)
15. X. Jia, H. Lu, and M.-H. Yang. Visual Tracking via Adaptive Structural Local Sparse Appearance Model. In *CVPR*, (2012)
16. Z. Kalal, J. Matas, and K. Mikolajczyk. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In *CVPR*, (2010)
17. B. Liu, J. Huang, L. Yang, and C. Kulikowski. Robust tracking using Local Sparse Appearance Model and K-Selection. In *CVPR*, (2011)
18. A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4), (2006)
19. R. Gade, A. Jorgensen, and T. Moeslund. Long-term occupancy analysis using graph-based optimisation in thermal imagery. In *CVPR*, pages 3698-3705, (2013)
20. A. Berg, J. Ahlberg, and M. Felsberg. A thermal object tracking benchmark. In *Advanced Video and Signal Based Surveillance (AVSS)*, 2015a, (2015)
21. R. Ippalally, S. H. Mudumba, M. Adkay and N. V. H. R., Object Detection Using Thermal Imaging, *IEEE 17th India Council International Conference (INDICON)*, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342179. (2020)
22. M. San-Biagio, M. Crocco and M. Cristani, Recursive segmentation based on higher order statistics in thermal imaging pedestrian detection, *2012 5th International Symposium on Communications, Control and Signal Processing*, pp. 1-4, doi: 10.1109/ISCCSP.2012.6217877. (2012)
23. A. Bañuls, A. Mandow, R. Vázquez-Martín, J. Morales, and A. García-Cerezo, Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes, *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 380-386, doi: 10.1109/SSRR50563.2020.9292593 (2020)
24. Hou et al. Human tracking over camera networks: a review. *EURASIP Journal on Advances in Signal Processing*, 43 DOI 10.1186/s13634-017-0482-z (2017)
25. M Paul, SM Haque, S Chakraborty, Human detection in surveillance videos and its applications-a review, *EURASIP Journal on Advances in Signal Processing*, no. 1, 176 (2013)
26. C Rasmussen, G Hager, Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(6), 560-576 (2001)



27. SM Naqvi, L Mihaylovay, JA Chambers, Clustering, and a joint probabilistic data association filter for dealing with occlusions in multi-target tracking, *Int. Conf. Information Fusion*, Istanbul, Turkey, Jul. (2013)
28. S Hamid Rezato\_ghi, et al., Joint probabilistic data association revisited, *IEEE International Conf. Computer Vision*, Santiago, Chile, Dec. (2015)
29. SW Joo, R Chellappa, A multiple-hypothesis approach for multiobject visual tracking. *IEEE Trans. Image Process.* 16(11), 2849-2854 (2007)
30. L Zhang, Y Li, and R Nevatia. Global data association for multi-object tracking using network flows, *IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, USA, Jun. (2008)
31. DS Jang, HI Choi, Active models for tracking moving objects. *Pattern Recogn.* 33(7), 1135-1146 (2000)
32. DS Jang, SW Jang, HI Choi, 2D human body tracking with structural Kalman Filter. *Pattern Recogn.* 35(10), 2041-2049 (2002)
33. C Liu, C Hu, JK Aggarwal, Eigenshape kernel based mean shift for human tracking, *IEEE Int. Conf. Computer Vision Workshops*, Barcelona, Spain, Nov. (2011)
34. J Fang, J Yang, H Liu, Efficient and robust fragments-based multiple kernels tracking. *Int. J. Electron. Commun.* 65(1), 915-923 (2011)
35. B Yang, R Yang, Interactive particle filter with occlusion handling for multitarget tracking, *IEEE Int. Conf. Fuzzy Systems and Knowledge Discovery*, Zhangjiajie, China, Aug. (2015)
36. X Zhang, W Hu, S Maybank, A smarter particle Filter, *Asian Conf. Computer Vision Computer*, Xi'an, China, Sep. (2009)
37. AW Smeulders, DM Chu, R Cucchiara, S Calderara, A Dehghan, M Shah, Visual tracking: an experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 36(7), 1442-1468 (2014)
38. A Yilmaz, O Javed, M Shah, Object tracking: a survey. *ACM Comput. Surv. (CSUR)* 38(4), 1-13, (2006)
39. B Y Lee et al. Occlusion handling in videos object tracking: A survey. *IOP Conf. Ser.: Earth Environ. Sci.* 18 012020. doi:10.1088/1755-1315/18/1/012020 (2014)
40. C. Ertler, H. Posseger, M. Optiz, and H. Bischof, Pedestrian detection in rgb-d images from an elevated viewpoint, in *22nd Computer Vision Winter Workshop*, (2017)
41. Ahmed, A. Ahmad, F. Piccialli, A. K. Sangaiah, and G. Jeon, A robust features-based person tracker for overhead views in industrial environment, *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1598-1605, (2018)
42. Ahmed and J. N. Carter, A robust person detector for overhead views, in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, pp. 1483-1486. (2012)
43. M. Rauter, Reliable human detection and tracking in top-view depth images, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 529-534. (2013)
44. Ahmad M. et al. Person Detection from Overhead View: A Survey. *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 4, (2019)
45. Hosokawa T., Kudo M., Person Tracking with Infrared Sensors. In: Khosla R., Howlett R.J., Jain L.C. (eds) *Knowledge-Based Intelligent Information and Engineering Systems. KES 2005. Lecture Notes in Computer Science*, vol 3684. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/1155402895>:(2005)
46. S. Honda, K. Fukui, K. Moriyama, S. Kurihara and M. Numao, Extracting Human Behaviors with Infrared Sensor Network, 2007 Fourth International Conference on Networked Sensing Systems, pp. 122-125, doi: 10.1109/INSS.2007.4297404. (2007)
47. S. Tao, M. Kudo, B. Pei, H. Nonaka, and J. Toyama, Multiperson Locating and Their Soft Tracking in a Binary Infrared Sensor Network, in *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 550-561, Oct. doi: 10.1109/THMS.2014.2365466. (2015)
48. T. Hosokawa, M. Kudo, H. Nonaka, and J. Toyama, Soft authentication using an infrared ceiling sensor network, *Pattern Anal. Appl.*, vol. 12, no. 3, pp. 237-249, (2009)
49. A. Berg, J. Ahlberg, and M. Felsberg. A thermal infrared dataset for evaluation of short-term tracking methods. In *Swedish Symposium on Image Analysis (SSBA)*, (2015)
50. Berg, K. • Ofj• all, J. Ahlberg, and M. Felsberg. Detecting rails and obstacles using a train-mounted thermal camera. In *Image Analysis*, volume 9127 of *Lecture Notes in Computer Science*, pages 492-503. Springer International Publishing, (2015)
51. Berg, J. Ahlberg, and M. Felsberg. Channel coded distribution\_eld tracking for thermal infrared imagery. Submitted to *IEEE PETS Workshop*, (2016).
52. Berg, J. Ahlberg, and M. Felsberg. Enhanced analysis of thermographic images for monitoring of district heat pipe networks. Submitted to *Pattern Recognition Letters (PRL)*, (2016)
53. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallower: principles and practice of background maintenance. In, *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 255-261 vol.1, doi: 10.1109/ICCV.1999.791228. (1999)

54. A. Berg and J. Ahlberg. Classification and temporal analysis of district heating leakages in thermal images. In *The 14th International Symposium on District Heating and Cooling (DHC)*, (2014)
55. Tro\_mova, A., Masciadri, A., Veronese, F. and Salice, F. Indoor Human Detection Based on Thermal Array Sensor Data and Adaptive Background Estimation. *Journal of Computer and Communications*, 5, 16-28. doi: 10.4236/jcc.2017.54002. (2017)
56. O. B. Tariq, M. T. Lazarescu and L. Lavagno, Neural Networks for Indoor Person Tracking with Infrared Sensors, in *IEEE Sensors Letters*, vol. 5, no. 1, pp. 1-4, Jan., Art no. 6000204, doi: 10.1109/LSSENS.2021.3049706. (2021)
57. Sixsmith, A. and Johnson, N., A Smart Sensor to Detect the Falls of the Elderly. *IEEE Pervasive Computing*, 3, 42-47. <https://doi.org/10.1109/MPRV.2004.1316817> (2004)
58. Erickson, V.L., Beltran, A., Winkler, D.A., Esfahani, N.P., Lusby, J.R. and Cerpa, A.E. Thermosense: Thermal Array Sensor Networks in Building Management. *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, Roma, 11-15 November, 87. <https://doi.org/10.1145/2517351.2517437> (2013)
59. Basu, C. and Rowe, A. Tracking Motion and Proxemics Using Thermal-Sensor Array. *arXiv:1511.08166* (2015)
60. Qian, R., Garg, D., Wang, Y., You, Y., Belongie, S., Hariharan, B., Campbell, M., Weinberger, K., Chao, W.L. End-to-End Pseudo-LiDAR for Image-Based 3D Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2020)
61. Ye, M., Xu, S., Cao, T. HVNet: Hybrid Voxel Network for LiDAR Based 3D Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2020)
62. Tu, J., Ren, M., Manivasagam, S., Liang, M., Yang, B., Du, R., Cheng, F., Urtasun, R., Physically Realizable Adversarial Examples for LiDAR Object Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2020)
63. Tian, H., Chen, Y., Dai, J., Zhang, Z., Zhu, X., Unsupervised Object Detection with LIDAR Clues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5962-5972). (2021)
64. Fang, J., Zuo, X., Zhou, D., Jin, S., Wang, S., Zhang, L., LiDAR-Aug: A General Rendering-Based Augmentation Framework for 3D Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4710-4720). (2021)
65. Li, Z., Wang, F., Wang, N., LiDAR R-CNN: An Efficient and Universal 3D Object Detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7546-7555). (2021)
66. Un, P., Wang, W., Chai, Y., Elsayed, G., Bewley, A., Zhang, X., Sminchisescu, C., Anguelov, D. RSN: Range Sparse Net for Efficient, Accurate LiDAR 3D Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5725-5734). (2021)
67. Wang, Y., Chao, W.L., Garg, D., Hariharan, B., Campbell, M., Weinberger, K. Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2019)
68. Zhu, M., Ma, C., Ji, P., Yang, X. Cross-Modality 3D Object Detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 3772-3781). (2021)
69. Andreas Geiger, Philip Lenz, Raquel Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. (2012)
70. Alnagar, Y., A\_\_, M., Amer, K., and ElHelw, M. Multi Projection Fusion for Real-Time Semantic Segmentation of 3D LiDAR Point Clouds. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* pp. 1800-1809. (2021)
71. Koide K, Miura J, Menegatti E. A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems*. March doi:10.1177/1729881419841532 (2019)
72. Oishi, S., Kohari, Y., Miura, J. Toward a robotic attendant adaptively behaving according to human state. 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). doi:10.1109/roman.2016.7745236 (2016)
73. Bellotto, N., Huosheng Hu., Multisensor-Based Human Detection and Tracking for Mobile Service Robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1), 167-181. doi:10.1109/tsmcb.2008.2004050 (2009)
74. Lindstrom, M., Eklundh, J.-O. (n.d.). Detecting and tracking moving objects from a mobile platform using a laser range scanner. *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the Next Millennium (Cat. No.01CH37180)*. doi:10.1109/iros.2001.977171 (2001)
75. K. Misu, J. Miura, Specific person tracking using 3D LIDAR and ESPAR antenna for mobile service robots, *Advanced Robotics*, 29:22, 1483-1495, DOI: 10.1080/01691864.2015.1093429. (2015)
76. D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita, Person Tracking in Large Public Spaces Using 3-D Range Sensors, in *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522-534, Nov. doi: 10.1109/THMS.2013.2283945. (2013)

77. Sahba, R., Sahba, A., Jamshidi, M., Rad, P. (2019). 3D Object Detection Based on LiDAR Data. IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON). doi:10.1109/uemcon47517.2019.8993 (2019)
78. A. Lang, et al., PointPillars: Fast Encoders for Object Detection from Point Clouds, in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 12689-12697. doi: 10.1109/CVPR.2019.01298. (2019)
79. J. Lambert et al., Performance Analysis of 10 Models of 3D LiDARs for Automated Driving, in IEEE Access, vol. 8, pp. 131699-131722, doi: 10.1109/ACCESS.2020.3009680. (2020)
80. Y. Wu, Y. Wang, S. Zhang and H. Ogai, Deep 3D Object Detection Networks Using LiDAR Data: A Review, in IEEE Sensors Journal, vol. 21, no. 2, pp. 1152-1171, 15 Jan.15, doi: 10.1109/JSEN.2020.3020626. (2021)
81. Y. Fan, B. Wu, C. Huang and Y. Bai, Environment Detection of 3D LiDAR by Using Neural Networks, 2019 IEEE International Conference on Consumer Electronics (ICCE), pp. 1-2, doi: 10.1109/ICCE.2019.8662037. (2019)
82. Masahiro Shiomi Norihiro Hagita Finding a person with a wearable acceleration sensor using a 3D position tracking system in daily environments, *Advanced Robotics*, 29:23, 1563-1574, DOI: 10.1080/01691864.2015.1095651. (2015)
83. M. Sualeh and G. -W. Kim, Visual-LiDAR Based 3D Object Detection and Tracking for Embedded Systems, in IEEE Access, vol. 8, pp. 156285-156298, doi: 10.1109/ACCESS.2020.3019187. (2020)
84. Z. Yan, T. Duckett and N. Bellotto, Online learning for human classification in 3D LiDAR-based tracking, 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 864-871, doi: 10.1109/IROS.2017.8202247. (2017)
85. Guerrero-Higueras \_AM, \_Alvarez-Aparicio C, Calvo Olivera MC, Rodríguez-Lera FJ, Fernandez-Llamas C, Rico FM and Matellan V Tracking People in a Mobile Robot From 2D LIDAR Scans Using Full Convolutional Neural Networks for Security in Cluttered Environments. *Front. Neurorobot.* 12:85. doi: 10.3389/fnbot.2018.00085. (2019)
86. Aguirre, E., Garcia-Silvente, M., and Plata, J. Leg detection and tracking for a mobile robot and based on a laser device, supervised learning, and particle Filtering, in *ROBOT2013: First Iberian Robotics Conference, Vol 252*, eds M. Armada, A. Sanfeliu and M. Ferre (Cham: Springer), 433-440. doi: 10.1007/978-3-319-03413-331: (2014)
87. Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. ROS: an open-source robot operating system," in *ICRA Workshop on Open-Source Software, Vol. 3 (Kobe)*, 5-10. (2009)
88. Premebida, C., Ludwig, O., and Nunes, U. Lidar and vision-based pedestrian detection system. *J. Field Rob.* 26, 696-711. (2009)
89. Wang DZ, Posner I, Newman P. Model-free detection and tracking of dynamic objects with 2D lidar. *The International Journal of Robotics Research.*34(7):1039-1063. doi:10.1177/0278364914562237. (2015)
90. Christoph Mertz, Luis E. Navarro-Serment, Robert MacLachlan, Paul Rybski, Aaron Steinfeld, Arne Suppe, Christopher Urmson, Nicolas Vandapel, Martial Hebert, Chuck Thorpe, David Duggins, and Jay Gowdy. Moving object detection with laser scanners. *J. Field Robot.* 30, 1, 17-43. DOI: <https://doi.org/10.1002/rob.21430>. (2013)
91. Becker, Marcelo et al. 2D laser-based probabilistic motion tracking in urban-like environments. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, v. 31, n. 2, pp. 83-96. 25. <https://doi.org/10.1590/S1678-58782009000200001>. (2009)
92. A. Leigh, J. Pineau, N. Olmedo and H. Zhang, Person tracking and following with 2D laser scanners, 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 726-733, doi: 10.1109/ICRA.2015.7139259. (2015)
93. Hasan, M., Hanawa, J., Goto, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: Tracking People Using Ankle-Level 2D LiDAR for Gait Analysis. In: Ahram T. (eds) *Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2020. Advances in Intelligent Systems and Computing*, vol 1213. Springer, Cham. (2021)
94. Hasan, M., Hanawa, J., Goto, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: Person Tracking Using Ankle-Level LiDAR Based on Enhanced DBSCAN and OPTICS. In: *IEEJ Trans Elec Electron Eng.* (2021)
95. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, pp. 34-42 (2015)
96. Li, S., Nguyen, V.H., Ma, M. et al.: A simplified nonlinear regression method for human height estimation in video surveillance. In: *J Image Video Proc.*, 32 (2015)
97. Gunel, S., Rhodin, H., Fua, P.: What face and body shapes can tell us about height. In: *Proceedings of IEEE International Conference on Computer Vision (ICCV) Workshops* (2019)
98. Bieler, D., Gunel, S., Fua, P., Rhodin, H.: Gravity as a Reference for Estimating a Person's Height from Video. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8569-8577 (2019)
99. Lee, D. et al.: Human Height Estimation by Color Deep Learning and Depth 3D Conversion. *Applied Sciences*. 10, 16, 5531 (2020).

100. Gunathilake KMTB, Manjika MS, Vidanapathirana M, Siddhisena KAP. Estimation of the height by using hand-span; a clinical forensic study. *Medico-Legal Journal of Sri Lanka*, 6(2):64-69. DOI: <http://dx.doi.org/10.4038/mlj.v6i2.7376>. (2018)
101. Merve Gullu, Eyup Burak Ceyhan, and Ceren Ulucan. A New Approach: Predicting Height of a Person from Joint Ratio of Fingers. In *Proceedings of the Fifth International Conference on Network, Communication and Computing (ICNCC '16)*. Association for Computing Machinery, New York, NY, USA, 182-187. doi: <https://doi.org/10.1145/3033288.3033326>. (2016)
102. Yan, Z., Duckett, T. Bellotto, N. Online learning for 3D LiDAR-based human detection: experimental analysis of point cloud clustering and classification methods. *Auton Robot* 44, 147-164 (2020)
103. Yamada, H., Ahn, J., Mozos, O. M., Iwashita, Y., Kurazume, R.: Gait-based person identification using 3D LiDAR and long short-term memory deep networks. *Advanced Robotics*, 1-11 (2020)
104. Benedek, C., Galai, B., Nagy, B., Jankó, Z.: Lidar-Based Gait Analysis and Activity Recognition in a 4D Surveillance System. In: *Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 101-113 (2018)
105. Tu, J., Ren, M., Manivasagam, S., Liang, M., Yang, B., Du, R., Cheng, F., Urtasun, R.: Physically Realizable Adversarial Examples for LiDAR Object Detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13716-13725 (2020)
106. Hasan, M., Goto, R., Hanawa, J., Fukuda, H., Kuno, Y., Kobayashi, Y.: Person Property Estimation Based on 2D LiDAR Data Using Deep Neural Network. In: Huang DS., Jo KH., Li J., Gribova V., Bevilacqua V. (eds) *Intelligent Computing Theories and Application. ICIC 2021. Lecture Notes in Computer Science*, vol 12836. Springer, Cham. (2021)
107. Rusu R.B. and Cousins S: 3D is here: Point Cloud Library (PCL), *Proc. of IEEE International Conference on Robotics and Automation*, Shanghai, pp. 1-4 (2011)
108. Bogoslavsky I and Stachniss C: Fast range image-based segmentation of sparse 3d laser scans for online operation, *Proc. of IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp.163-169 (2016)
109. Zermas D, Izzat I and Papanikolopoulos N: Fast segmentation of 3D point clouds: A paradigm on LiDAR data for autonomous vehicle applications, *Proc. of IEEE international conference on robotics and automation (ICRA) Singapore*, pp. 5067-5073 (2017)
110. Wen W, Hsu L-T, Zhang G. Performance Analysis of NDT-based Graph SLAM for Autonomous Vehicle in Diverse Typical Driving Scenarios of Hong Kong. *Sensors*. 18(11):3928. <https://doi.org/10.3390/s18113928>. (2018)
111. Geyer J. et. al. A2D2: Audi Autonomous Driving Dataset. *CoRR*, abs/2004.06320. (2004)
112. Huang, X., Wang, P., Cheng, X., Zhou, D., Geng, Q., Yang, R. The ApolloScape Open Dataset for Autonomous Driving and Its Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10), 2702-2719. (2020)
113. Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, James Hays. *Argoverse: 3D Tracking and Forecasting with Rich Maps*. in *CVPR 2019*, (2019)
114. Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. *arXiv:1805.04687*, (2020)
115. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, Bernt Schiele. *The Cityscapes Dataset for Semantic Urban Scene Understanding*. *arXiv:1604.01685*, (2016)
116. Harald Schafer, Eder Santana, Andrew Haden, Riccardo Biasini. *A Commute in Data: The comma2k19 Dataset*. *arXiv:1812.05752* (2018)
117. Tobias Weyand, Andre Araujo, Bingyi Cao, Jack Sim. *Google Landmarks Dataset v2: A Large-Scale Benchmark for Instance-Level Recognition and Retrieval*. In *CVPR 2020*, (2020).
118. Jean-Luc Deziel, Pierre Merriaux, Francis Tremblay, Dave Lessard, Dominique Plourde, Julien Stanguennec, Pierre Goulet, Pierre Olivier. *PixSet: An Opportunity for 3D Computer Vision to Go Beyond Point Clouds With a Full-Waveform LiDAR Dataset*. *arXiv:2102.12010*, (2021)
119. John Houston, Guido Zuidhof, Luca Bergamini, Yawei Ye, Long Chen, Ashesh Jain, Sammy Omari, Vladimir Igloukov, Peter Ondruska. *One Thousand and One Hours: Self-driving Motion Prediction Dataset*. *arXiv:2006.14480*, (2020)
120. Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, Oscar Beijbom. *nuScenes: A multimodal dataset for autonomous driving*. *arXiv:1903.11027*, (2020)
121. Dan Barnes, Matthew Gadd, Paul Murcutt, Paul Newman, Ingmar Posner. *The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset*. *arXiv:1909.01300*, (2020)

122. P. Xiao et al., PandaSet: Advanced Sensor Suite Dataset for Autonomous Driving, 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 3095-3101, doi: 10.1109/ITSC48978.2021.9565009. (2021)
123. Yueming Zhang, Xiaolin Song, Bing Bai, Tengfei Xing, Chao Liu, Xin Gao, Zhihui Wang, Yawei Wen, Haojin Liao, Guoshan Zhang, Pengfei Xu. 2nd Place Solution for Waymo Open Dataset Challenge - Real-time 2D Object Detection.arXiv:2106.08713, (2021)
124. Abhishek Patil, Srikanth Malla, Haiming Gang, Yi-Ting Chen. The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes.arXiv:1903.01568 (2019)
125. Dan Jia, Alexander Hermans, Bastian Leibe. Domain and Modality Gaps for LiDAR-based Person Detection on Mobile Robots.arXiv:2106.11239. (2021)
126. Yin Zhou and Oncel Tuzel. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In CVPR, (2017)
127. Bo Li Yan Yan, Yuxing Mao. SECOND: Sparsely Embedded Convolutional Detection. Sensors, (2018)
128. Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. STD: Sparse-to-Dense 3D Object Detector for Point Cloud. In ICCV, (2019)
129. Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From Points to Parts: 3D Object Detection from Point Cloud with Part-aware and Part aggregation Network. PAMI, (2020)
130. Kai O Arras, Oscar Martinez Mozos, and Wolfram Burgard. Using Boosted Features for the detection of People in 2D Range Data. In ICRA, (2007)
131. Caroline Pantofaru. ROS leg detector package. <https://wiki.ros.org/legdetector>; Accessed (2018)
132. Angus Leigh, Joelle Pineau, Nicolas Olmedo, and Hong Zhang. Person tracking and Following with 2D Laser Scanners. In ICRA, (2015)
133. Lucas Beyer, Alexander Hermans, Timm Linder, Kai Oliver Arras, and Bastian Leibe. Deep Person Detection in 2D Range Data. RA-L, 3(3):2726-2733, (2018)
134. Dan Jia, Alexander Hermans, and Bastian Leibe. DRSPAAM: A Spatial-Attention and Auto-regressive Model for Person Detection in 2D Range Data. In IROS, (2020)
135. Dan Jia, Mats Steinweg, Alexander Hermans, and Bastian Leibe. Self-Supervised Person Detection in 2D Range Data using a Calibrated Camera. In ICRA, (2021)
136. Martin Engelcke, Dushyant Rao, D.Wang, C. Tong, and I. Posner. Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks. In ICRA, (2017)
137. Y. Zhou, P. Sun, Y. Zhang, Dragomir Anguelov, J. Gao, Tom Ouyang, J. Guo, J. Ngiam, and Vijay Vasudevan. End-to-End Multi-View Fusion for 3D Object Detection in LiDAR Point Clouds. In CoRL, (2019)
138. Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center based 3D Object Detection and Tracking. In CVPR, (2021)
139. Wu Zheng, Weiliang Tang, Sijin Chen, Li Jiang, and Chi-Wing Fu. CIA-SSD: Confident IoU-Aware Single-Stage Object Detector From Point Cloud. In AAAI, (2021)
140. Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud. In CVPR, (2019)
141. Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep Hough Voting for 3D Object Detection in Point Clouds. In ICCV, (2019)
142. Yilun Chen, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Fast Point R-CNN. In ICCV, (2019)
143. Lee Y, Park S. A Deep Learning-Based Perception Algorithm Using 3D LiDAR for Autonomous Driving: Simultaneous Segmentation and Detection Network (SSADNet). Applied Sciences. 10(13):4486. <https://doi.org/10.3390/app10134486>. (2020)
144. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. arXiv 2015, arXiv:1411.4038. (2015)
145. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Intelligent Tutoring Systems; Springer Science and Business Media LLC: Berlin, Germany, Volume 9351, pp. 234-241. (2015)
146. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 91-99, (2015)
147. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy; pp. 2999-3007. 22-29, (2017)
148. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. arXiv 2018, arXiv:1804.02767. (2018)
149. Yang, B.; Luo, W.; Urtasun, R. PIXOR: Real-time 3D object detection from point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 7652-7660. (2018)

150. Milioto, A.; Vizzo, I.; Behley, J.; Stachniss, C. RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Venetian, Macao, 3-8, pp. 4213-4220. (2019)
151. Wang, Y.; Shi, T.; Yun, P.; Tai, L.; Liu, M. PointSeg: Real-time semantic segmentation based on 3D LiDAR point cloud. arXiv arXiv:1807.06288. (2018)
152. Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers. People Tracking with Mobile Robots Using Sample-Based Joint Probabilistic Data Association Filters. *IJRR*, 22(2):99-116, (2003)
153. Ajo Fod, Andrew Howard, and Maja J. Mataric. A laserbased people tracker. In *ICRA*, (2002)
154. Matthias Scheutz, J. McRaven, and Gyorgy Cserey. Fast, reliable, adaptive, bimodal people tracking for indoor environments. In *IROS*, (2004)
155. Angus Leigh, Joelle Pineau, Nicolas Olmedo, and Hong Zhang. Person tracking and Following with 2D Laser Scanners. In *ICRA*, (2015)
156. Caroline Pantofaru. ROS leg detector package. <https://wiki.ros.org/legdetector>; 2010: Accessed(2018)
157. Lucas Beyer, Alexander Hermans, and Bastian Leibe. DROW: Real-Time Deep Learning based Wheelchair Detection in 2D Range Data. *RA-L*, 2(2):585-592, (2016)
158. Dan Jia, Mats Steinweg, Alexander Hermans, and Bastian Leibe. Self-Supervised Person Detection in 2D Range Data using a Calibrated Camera. In *ICRA*, (2021)
159. Yong Wu, Kun Zhang, Di Wu, Chao Wang, Chang-An Yuan, Xiao Qin, Tao Zhu, Yu-Chuan Du, Han-Li Wang, D.S.Huang, Person reidenti\_cation by multiscale feature representation learning with random batch feature mask, *IEEE Transactions on Cognitive and Developmental Systems*, 13(4): 865-874, (2021)
160. Di Wu, Chao Wang, Yong Wu, Qi-Cong Wang and D.S.Huang, Attention deep model with multiscale deep supervision for person re-identi\_cation, *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 1, pp. 70-78.(2021)
161. Van-Thanh Hoang, D.S.Huang and Kang-Hyun Jo, 3-D Facial Landmarks Detection for Intelligent Video Systems, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 578-586, (2021)
162. Xianpeng Liang, Di Wu, D.S.Huang, Image co-segmentation via locally biased discriminative clustering, *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 11: 2228-2233,(2019)
163. Di Wu, Si-Jia Zheng, Xiao-Ping Zhang, Chang-An Yuan, Fei Cheng, Yang Zhao, Yong-Jun Lin, Zhong-Qiu Zhao, Yong-Li Jiang and D.S.Huang, Deep learning based methods for person re-identification: A comprehensive review, *Neurocomputing*, vol.337: 354-371, (2019)
164. Di Wu, Kun Zhang, Si-Jia Zheng, Yong-Tao Hao, Fu-Qiang Liu, Xiao Qin, Fei Cheng, Yang Zhao, Qi Liu, Chang-An Yuan and D.S.Huang, Random occlusion recovery for person re-identification, *Journal of Imaging Science Technology*, 63(3), pp. 30405-1-30405-9(9) , (2019)
165. Di Wu, Hong-Wei Yang, D.S.Huang, Chang-An Yuan, Xiao Qin, Yang Zhao, Xin-Yong Zhao, Jian-Hong Sun. Omnidirectional feature learning for person reindentation, *IEEE Access*, vol. 7, pp. 28402-28411, (2019)
166. DiWu, Si-Jia Zheng, Chang-An Yuan and D.S.Huang, A deep model with combined losses for person re-identi\_cation. *Cognitive Systems Research*, 54: 74-82, (2019)
167. Di Wu, Si-Jia Zheng, Wen-Zheng Bao, Xiao-Ping Zhang, Chang-An Yuan and D.S.Huang, A novel deep model with multi-loss and efficient training for person reidentification, *Neurocomputing*, 324, pp. 69-76, (2019)
168. Xianpeng Liang, Lin Zhu, and D.S.Huang, Image segmentation fusion using weakly supervised trace-norm multi-task learning method, *IET Image Processing*, 12.7: 1079-1085. (2018)
169. Xianpeng Liang, Lin Zhu, and D.S.Huang, Multi-task ranking SVM for image segmentation, *Neurocomputing*, 247, pp. 126-136, (2017)
170. Wen Jiang, D.S.Huang, Shenghong Li, Random-walk based solution to triple level stochastic point location problem, *IEEE Trans. on Cybernetics*, vol.46, no.6, pp.1438- 1451, (2016)
171. Yang Zhao, and D.S.Huang, Completed local binary count for rotation invariant texture classification, *IEEE Trans. on Image Processing*, vol.21, no.10, pp. 4492 - 4497, (2012)
172. D.S.Huang, and Wen Jiang, A general CPL-AdS methodology for fixing dynamic parameters in dual environments, *IEEE Trans. on Systems, Man and Cybernetics -Part B*, vol.42, no.5, pp.1489-1500, (2012)
173. Xiao-Feng Wang, D.S.Huang and Huan Xu, An efficient local Chan-Vese model for image segmentation, *Pattern Recognition*, vol. 43, no.3, pp. 603-618, (2010)
174. Xiao-Feng Wang, D.S.Huang, A novel density-based clustering framework by using level set method, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no.11, pp 1515-1531, (2009)
175. D.S.Huang, Ji-Xiang Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, *IEEE Transactions on Neural Networks*, vol. 19, no.12, pp 2099-2115, (2008)
176. Linder T., Arras K.O. People Detection, Tracking and Visualization Using ROS on a Mobile Service Robot. In: Koubaa A. (eds) *Robot Operating System (ROS). Studies in Computational Intelligence*, vol 625. Springer, Cham. <https://doi.org/10.1007/978-3-319-26054-98>:(2016)

177. S. Hwang, N. Kim, Y. Choi, S. Lee, and I. S. Kweon, Fast multiple objects detection and tracking fusing color camera and 3D LIDAR for intelligent vehicles, 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 234-239, doi: 10.1109/URAI.2016.7625744. (2016)
178. Dendorfer, P., Osep, A., Milan, A. et al. MOT Challenge: A Benchmark for Single-Camera Multiple Target Tracking. *Int J Comput Vis* 129, 845-881. <https://doi.org/10.1007/s11263-020-01393-0>. (2021)
179. Babaei, M., Li, Z., Rigoll, G. A dual CNN-RNN for multiple people tracking. *Neurocomputing*, 368, 69-83. (2019)
180. Andriluka, M., Iqbal, U., Insafutdinov, E., Pishchulin, L., Milan, A., Gall, J., Schiele, B. PoseTrack: A benchmark for human pose estimation and tracking. In *Conference on computer vision and pattern recognition*. (2018)
181. Bae, S.-H., Yoon, K.-J. Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking. *Transactions on Pattern Analysis and Machine Intelligence*, 40(3), 595-610. (2018)
182. Baisa, N. L. Online multi-target visual tracking using a HISP filter. In *International joint conference on computer vision, imaging and computer graphics theory and applications*. (2018)
183. Baisa, N. L. Online multi-object visual tracking using a GM-PHD filter with deep appearance learning. In *International conference on information fusion*. (2019)
184. Bergmann, P., Meinhardt, T., Leal-Taix\_e, L. Tracking without bells and whistles. In *International conference on computer vision*. (2019)
185. Bochinski, E., Eiselein, V., Sikora, T. High-speed tracking-by-detection without using image information. In *International conference on advanced video and signal-based surveillance*. (2017)
186. Dicle, C., Camps, O., Sznai, M. The way they move: Tracking targets with similar appearance. In *International conference on computer vision*. (2013)
187. Henschel, R., Zou, Y., Rosenhahn, B., Multiple people tracking using body and joint detections. In *Conference on computer vision and pattern recognition workshops*. (2019)
188. Kim, C., Li, F., Rehg, J. M. Multi-object tracking with neural gating using bilinear LSTM. In *European conference on computer vision*. (2018)
189. Kristan, M., et al. The visual object tracking VOT2014 challenge results. In *European conference on computer vision workshops*. (2014)
190. Lan, L., Wang, X., Zhang, S., Tao, D., Gao, W., Huang, T. S. Interacting tracklets for multi-object tracking. *Transactions on Image Processing*, 27(9), 4585-4597. (2018)
191. Zhang Y, Sun P, Jiang Y, Yu D, Yuan Z, Luo P, Liu W, Wang X ByteTrack: Multi-Object Tracking by Associating Every Detection Box, arXiv:2110.06864 (2021)
192. Leal-Taixe, L., Canton-Ferrer, C., Schindler, K. Learning by tracking: Siamese CNN for robust target association. In *Conference on computer vision and pattern recognition workshops*. (2016)
193. Lee, S., Kim, E. Multiple object tracking via feature pyramid Siamese networks. *Access*, 7, 8181-8194. (2019)
194. Long, C., Haizhou, A., Chong, S., Zijie, Z., Bo, B. Online multi-object tracking with convolutional neural networks. In *International conference on image processing*. (2017)
195. Long, C., Haizhou, A., Zijie, Z., Chong, S. Real-time multiple people tracking with deeply learned candidate selection and person re-identification. In *International conference on multimedia and expo*. (2018)
196. Sanchez-Matilla, R., Cavallaro, A. A predictor of moving objects for first-person vision. In *International conference on image processing*. (2019)
197. Song, Y., Yoon, K., Yoon, Y., Yow, K., Jeon, M. Online multi-object tracking with GMPHD Filter and occlusion group management. *Access*, 7, 165103-165121. (2019)
198. Tang, S., Andres, B., Andriluka, M., Schiele, B. Multi-person tracking by multicast and deep matching. In *European conference on computer vision workshops*. (2016)
199. Tang, S., Andriluka, M., Andres, B., Schiele, B. Multiple people tracking with lifted multicut and person re-identification. In *Conference on computer vision and pattern recognition*. (2017)
200. Tao, Y., Chen, J., Fang, Y., Masaki, I., Horn, B. K. Adaptive spatio-temporal model based multiple object tracking in video sequences considering a moving camera. In *International conference on universal village*. (2018)
201. Tian, W., Lauer, M., Chen, L. Online multi-object tracking using joint domain information in tra\_c scenarios. *Transactions on Intelligent Transportation Systems*, 21(1), 374 - 384. (2019)
202. Wang, G., Wang, Y., Zhang, H., Gu, R., Hwang, J.-N. Exploit the connectivity: Multi-object tracking with trackletnet. In *International conference on multimedia*. (2019)
203. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, (2020)
204. Bolle RM, Connell J, Pankanti S, Ratha NK, Senior AW. *Guide to Biometrics*. Springer Science Business Media. (2013)
205. Wan C, Wang L, Phoha VV. A survey on gait recognition. *ACM Computing Surveys*. 51(5). 89. <https://doi.org/10.1145/3230633>. (2018)

206. Dargan S, and Kumar M. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*; Volume 143, doi: 10.1016/j.eswa.2019.113114. (2020)
207. Hasan, M., Hanawa, J., Goto, R., Suzuki, R., Fukuda, H., Kuno, Y., Kobayashi, Y.: LiDAR-based Detection, Tracking, and Property Estimation: A Contemporary Review. In: *Neurocomputing*. 506. 393-405. (2022).