Improving Livestock Indoor Localization in a Smart Agriculture Environment

Bachelor Thesis

Software System Science

Faculty of Information Systems and Applied Computer Sciences

Chair of Mobile Systems, University of Bamberg

Author: Jarod John Stock (2006947)
Supervisor: Simon Steuer
Examiner: Prof. Dr. Daniela Nicklas
Submission: May 2023

Contents

Abstract

Precision livestock farming can be of great value to animal welfare and the profitability of the livestock industry itself. The sensor-based livestock monitoring can specialize in the location and movement patterns of the animals. The techniques and technologies used for position determination, especially for indoor environments, are discussed in detail in this thesis. It is a usual occurrence that generated data sets show errors and noise. Both limit the information gain and thus the added value that Precision Livestock Farming can offer. In this context, established filters for indoor localization systems are presented, namely Kalman- and Particle Filtering. The work then covers the implementation of a customized filter specifically designed for the data of the project WeideInsight. This project focuses on providing a cost-effective, radio-based localization approach for livestock monitoring on pasture and in the barn. The developed filter focuses on outliers that are a frequent occurrence in the data sets. The subsequent evaluation confirms that the filter implementation can significantly improve the data quality.

Chapter 1

Introduction

1.1 Motivation

With the expansion of the Internet of Things, many sectors, including agriculture, are experiencing a shift towards digitalization. Sensor-based monitoring of livestock can be referred to as Precision Livestock Farming and is primarily concerned with optimization processes for animal welfare, production quantity and quality, and profitability. The fundamental change towards the usage of modern technology within livestock farming was made possible by the continuous development of hardware, which became available in smaller versions and required less acquisition and operation expenditures. [1]

Precision Livestock Farming monitors the status of feed and water supplies, temperature values, and audio and video recordings, but less often the location information of individual animals. However, position data does offer promising information gain, influencing the decisions of farmers. [2]

1.2 Problem Statement

Despite this trend, the explicit extraction of location information continues to be associated with expensive technology unsuitable for permanent application. Therefore, a cost-effective and efficient solution to the problem is needed, which the project WeideInsight tries to offer.[3]

If the installed infrastructure takes over the tasks that otherwise had to be performed by farmers on site, it has to extract reliable data to draw valid conclusions.[4] However, the systems in place do not always produce correct data. Data cleaning is a discipline determined to detect and repair errors and inconsistencies in data to maximize its data quality.[5] High data quality comes with a high density of information gain. So to take appropriate actions, farmers need to introduce some form of data cleaning to their observations.

In the case of indoor localization, usually, a filter solution is implemented. Therefore this paper analyzes what technologies are used in modern systems to track the indoor

[1] Zhao et al. [2010]
[2] Berckmans [2014]
[3] Berckmans [2014]
[4] Berckmans [2014]
[5] Rahm et al. [2000]

positions of livestock and approaches to improve their data quality.

1.3 Project WeideInsight

Although IoT-assisted livestock monitoring exists in many farms, attention rarely lies on location information. The reason for this is the high cost of existing localization technologies and their lack of ability to function under continuous operation with high data rates. The advantages of a sensor infrastructure are the time-independent readout of the location of an individual animal and obtaining movement information over time to draw appropriate conclusions. The project WeideInsight of the Chair of Mobile Systems at the University of Bamberg and various partners develop a cost-effective, radio-based localization solution for pasture and barn. Data sets that are analyzed and filtered in the process of the thesis originated within the project.

1.4 Thesis Structure

The thesis divides into a theoretical and a practical half. The theoretical part introduces the reader to Precision Livestock Farming and explains the benefits of monitoring livestock welfare. It then proceeds to cover the livestock localization from a technical point of view, going over important terms, techniques and technologies. The chapter finishes by gathering information about data filtering to lay the fundamentals of a successful implementation in the following chapter. The practical part is concerned with the implementation of a custom filter solution, discussing fundamental structure and assumptions that are used to make the filter perform in a suitable way. The thesis is wrapped up by an evaluation of the filtered data and a final conclusion.

Chapter 2

Background Knowledge

This chapter is divided into four sections, covering the review of literature, conceptual foundations, and related work. The first section defines the term Precision Livestock Farming and portrays two practical approaches. The second section then deals with behavioral aspects of livestock and the necessity of knowledge about their welfare. The third section is used to define multiple terms that are critical to comprehend the fundamentals of this work, as well as covering modern localization techniques, technologies, and how to measure their success. The final section presents details about the WeideInsight project and its data, discussing approaches and their effectiveness in achieving high data quality.

2.1 Precision Livestock Farming

Precision Livestock Farming PLF enables farmers to control production, health, reproduction, and environmental impact through customized management systems and real-time monitoring.

To view data on the status of livestock at any point in time opens up the possibility of taking appropriate action whenever needed. This is of great economic relevance to farmers because if the well-being of the animals is taken care of, the correlated production and quality will also be at an optimum. The technologies used to date mainly track external factors that affect animals, such as feed supply or temperature regulations - less often animals themselves. However, animal husbandry in the agricultural sector is in change due to the evolution of the technologies used and the invention of new technologies. This evolution is characterized in particular by two decreasing factors: the cost and the size of the equipment used, which ultimately means that more farmers are able to purchase and operate them, and time can be devoted to activities other than personal control on site. The devices used here consist mainly of cameras, microphones, various sensors, and other infrastructure links required to enable wireless communication, as well as the now invisible use of the Internet and data storage in the cloud.[1]

Therefore, the main objective of practicing PLF is the application of suitable hardware to collect data, followed by the use of intelligent software to obtain relevant and correct information that helps the farmers needs.[2]

[1]Berckmans [2014]
[2]Berckmans [2014]

Obviously, different data types provide other benefits and must be evaluated differently depending on the technology used. To gain a better understanding of the PLF concept, two application areas will be discussed. Popular examples include real-time sound analysis for monitoring health status and real-time image analysis for monitoring animal welfare.

Real-time sound analysis is widely used on pig farms, where an animal very likely suffers from a respiratory disease from rearing to slaughter weight. These diseases manifest as coughing and can be detected by hardware installed in the barn. The early detection of this type of disease is of great economic importance to the farmer, as the necessary investment in veterinarians in the event of a widespread disease outbreak makes livestock farming unprofitable. Problematically, however, it has been shown that even early detection of coughing sounds is still costly to treat due to the lack of categorization capabilities identify different diseases. In veterinary medicine, it is common to evaluate coughing sounds to better classify and ultimately diagnose the disease and, if necessary, isolate individual animals from the group. In order to automate this process, algorithms are now used in practice that have been trained to recognize the characteristics and differences of the various sounds produced by livestock and to classify them accordingly. For example, algorithms already exist that can correctly classify coughing sounds from pigs with a success rate of nearly 90 percent.[3]

Real-time image analysis uses cameras and appropriate monitoring algorithms focussed on inside barn usage. The behavior of the animals is recorded periodically by taking pictures or as permanent approach, via video recordings. This is important, for instance, because of legally required inspections within the EU, which are less time-consuming thanks to the installed technology. Other conclusions that can be drawn from this type of PLF mainly relate to the spatial distribution of the animals. For example, several cameras can monitor different locations in the barn and provide feedback on the number of individual animals. Again, alarms are set in case of conspicuous distribution patterns so the farmer can take appropriate action. Real-time image analysis models train on different scenarios and learn the distribution of animals in the top view. Based on this distribution, they compare already known patterns and identify problems such as high or low temperatures in certain zones, power outages, missing light sources, or even the status of water and food supplies.[4]

A third category of PLF, called real-time position analysis, is the main focus of this work since it is the fundament for indoor livestock localization, which is discussed in detail in the later sections.

2.2 Livestock Behaviour and Welfare

The main interest in PLF is thus livestock welfare, which raises the question of how one can measure it. Since the project WeideInsight deals with the location information of cattle, these animals are the focus of this section. Making assumptions from collected data sets require a basic understanding of behavioral livestock traits. Conclusions about the welfare of an animal can be drawn when the animal behaviour is studied. Figure 1 shows a classic cycle of how the various factors of health, physical condition, behavior,

[3] Berckmans [2014]
[4] Berckmans [2014]

and environmental factors are related.[5]

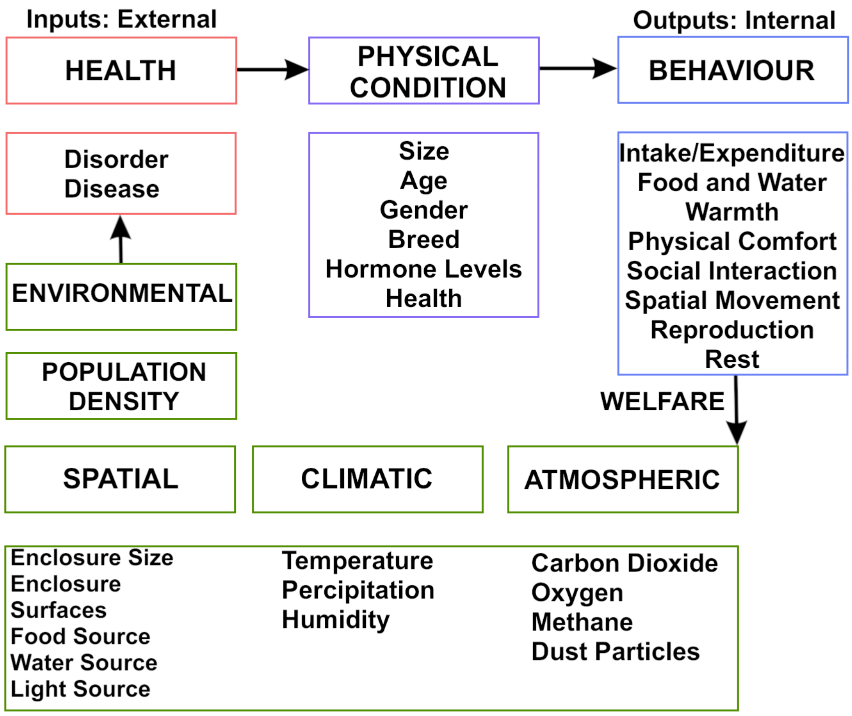

Figure 1: Livestock Behaviour Cycle, adapted from Tscharke & Banhazi [2016]

As with all living things, ones individual environment and interactions with it shape behaviour. An animal's behavior is always a direct response to physical stimuli, such as the need for food and water, or to external influences on the animal, such as responses to climate. The health status of an animal depends on the presence of disease, which strongly influences the physical condition. Multiple factors, as age, sex, and size, impact the physical condition. Depending on the attributes of health and physical condition, the actual behavior of an animal forms. For example, if an animal is in good health and physical condition, it will consume sufficient water and food, interact socially with other animals, and reproduce. Evaluating these behavioral attributes allows us to determine how good or bad an animal's well-being is. The last point on the diagram displays the environmental impact, which affects the animal health, starting the cycle once again. Therefore ethically and economically, it should be the goal of the farmer to intervene appropriately in the case of problems with the health or well-being of the animals.[6] Even if most countries have laws and legal restrictions that ensure a certain level of animal health and welfare, no universal threshold exists that needs to be achieved, making it vary depending on human moral.[7]

While most of the attributes of the behavioral category are self-explanatory, it is particularly interesting to take a closer look at the social interaction and spatial movement of cattle in the sense of localization. Even today far from all behavioral attributes of

[5] Tscharke & Banhazi [2016]
[6] Tscharke & Banhazi [2016]
[7] Phillips [2008]

livestock can be explained, although much can be inferred from recent research in this area. In their paper, Marino and Allen conducted a literature review on cattle psychology and came to the following conclusions: Cattle can discriminate between objects, humans, and conspecifics. In doing so, they can engage in a variety of complex social interactions and actions due to their equally present emotional capacities and individual personalities.[8] In herds, there are often animals that deliberately avoid each other because of personal dislike, or the exact opposite, where groups of animals are often found together because of affiliate relationships.[9]

To make further progress in the technology-focused field of PLF, it is also necessary that research in the context of animals and their behavior progresses, to achieve the optimal benefit for humans and animals. In summary, the cognitive abilities of cattle are more complex than most people believe, yet one can successfully monitor livestock using existing IoT because the data is sufficiently well interpreted.

2.3 Livestock Localization

The real-time positioning analysis has a positive impact on animal welfare, detecting diseases and sending alarms whenever human intervention is required. In addition, the time-consuming locating of animals is greatly facilitated, which has a high economic value. This section details how livestock is located and discusses indoor techniques and technologies and their related terminology.

2.3.1 Definition of Terms

This subsection defines basic terms used in the literature and practice related to livestock localization, with a special focus on indoor environments.

Positioning is concerned with determining the position of an object. In the context of livestock localization, this object is some type of sensor that is attached to the animals. [10]

Localization is the process of determining the position of an object based on wireless sensor networks and communicating nodes. In many cases the term emphasizes that an application necessitates topological correctness of the sensor locations, while the absolute coordinate position is not as important. Hence, this term is an approximate location estimate, especially for low-accuracy systems. In literature localization and positioning are often used synonymously.[11]

Tracking is concerned with determining the position of a moving object over a time interval. This mobile object is usually part of some positioning system.[12]

[8] Marino & Allen [2017]
[9] Gygax et al. [2010]
[10] Mautz [2012]
[11] Mautz [2012]
[12] Mautz, 2012

Positioning Systems are an infrastructure deployed locally to track mobile object positions.[13]

Beacons are the part of a Positioning System that track, e.g. receive signals from the tags put onto livestock.

Location Based Service LBS depends on the data gathered by localization of a device to offer a variety of of services, such as tracking or navigation. The service usually consists out of the software application, a network, a device that is concerned with positioning and the users own mobile device. There are a variety of ways to determine the location of a device and they differ depending on the environment.[14]

Environments for Livestock Localization can be subdivided into indoor environments and outdoor environments. Depending on the environment the used positioning system can be further characterized as either indoor/outdoor or a mixed type of positioning system.[15]

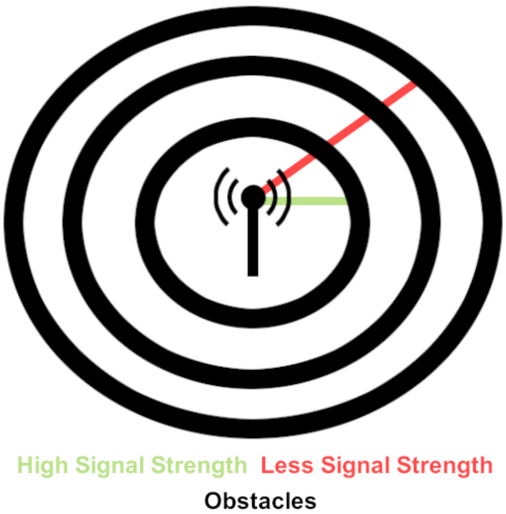

Figure 2: Decreasing Signal Strength, adapted from Farid et al. [2013]

Indoor Localization is therefore localization that takes place in closed surroundings with the goal of providing a precise value of the position of an object. This can be challenging especially in comparison to outdoor environments since indoor environments like buildings can have a complex interior, consisting of missing line of sights, obstacles like walls, doors, even animals or humans and more. Furthermore the fact that indoor environments are usually subdivided into multiple rooms allows the assumption of objects in the wrong room when the localization data is unprecise. A good positioning system must

[13]Mautz [2012]
[14]Farid et al. [2013]
[15]Farid et al. [2013]

therefore provide an infrastructure that can overcome these challenges. When considering indoor localization a term that is often to used to refer to the accuracy of the location data is signal strength. A simple Visualization of signal strength can be seen in Figure 2, where whenever multiple obstacles are passed and due to the higher range the signal strength decreases.[16]

Multipath effect happens whenever the propagation of a signal reaches the receiving beacon via multiple paths, causing interferences.[17] Most of the downsides related to indoor localization are based on this effect, which is caused by non existing line of sight conditions. [18]

2.3.2 Performance Metrics

When talking about localization, there are serveral performance metrics that can be measured to evaluate the success of a localization process.

Accuracy must be high in almost all positioning systems and is an important user requirement. Accuracy, sometimes referred to as position error, is the value that can be calculated when comparing position data with the actual position of the monitored object. In the development of positioning systems, there is often some form of ground truth data that is guaranteed to contain the correct positions so that the accuracy can be tracked.[19]

Coverage is strongly related to the accuracy metric. It is important in determining the effectiveness of the existing infrastructure, where the higher the coverage of an area, the better.[20]

Responsiveness is how often location data is collected when looking at a moving object.[21]

Adaptiveness is the ability of a localization system to perform under changing environmental conditions. An adaptive system will provide more accurate data. From an economic point of view, it may be an initial investment to build adaptive systems, but it comes with the avoidance of repeated recalibrations.[22]

Scalability is a highly anticipated feature, especially in livestock localization, where a large number of individual animals often need to be tracked. High scalability means that the existing system can work without restrictions even with a large number of tracked objects and a large area coverage.[23]

[16] Farid et al. [2013]
[17] National Institute of Standards and Technology [2023]
[18] De-La-Llana-Calvo et al. [2019]
[19] Farid et al. [2013]
[20] Farid et al. [2013]
[21] Farid et al. [2013]
[22] Farid et al. [2013]
[23] Farid et al. [2013]

Cost and Complexity are important economic considerations. Costs typically arise from building an infrastructure, adding additional bandwidth, replacing obsolete technology, and maintaining the system. The complexity of a system often changes with the application and can cause additional costs. The level of complexity is rooted in localization techniques and technologies, which are discussed next. [24]

2.3.3 Localization Techniques

As shown in Figure 3, many modern literature sources divide indoor localization techniques into three different categories: proximity detection, triangulation (of which most approaches use direction- or distance-based information) and scene analysis. This subsection explains the basics of each of these categories and then introduces some of their variants, that are commonly used in real life scenarios.

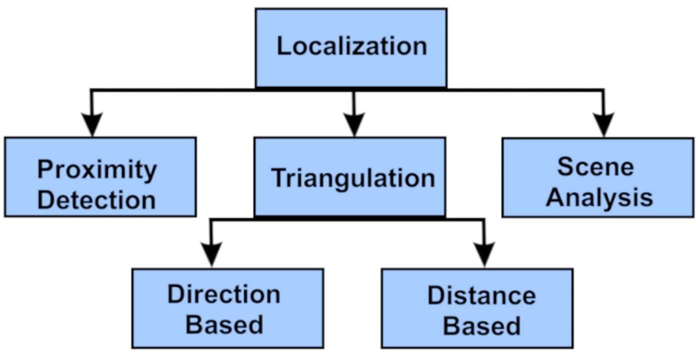

Figure 3: Location detection classification, adapted from Farid et al. [2013]

Proximity Detection is a simple way to implement localization and equals to connectivity-based positioning. The location of a mobile object is determined using the cell of origin method and beacons with limited range. If the mobile object is within range of more than one beacon, the beacon with the strongest signal is selected. The accuracy in this category is highly dependent on the number of beacons and their range capabilities. Proximity detection can be implemented using a variety of wireless localization technologies, such as Radio Frequency Identification (RFID) and Bluetooth, which reappear later.[25] Figure 4 shows that the object would be identified to be at the position of beacon 2, based on the proximity comparison between the two beacons in range.

Triangulation uses triangles and their geometric properties to determine the position of an object. Implementing triangulation, one can choose between a lateration or an angulation approach. Lateration approaches include time based and received signal strength based techniques, while angulation approaches utilize the angle of incoming communications. Formally, triangulation splits into the two categories direction-based and distance-based. Lateration is a time-based method that uses different distance measurements, such

[24] Farid et al. [2013]
[25] Farid et al., 2013

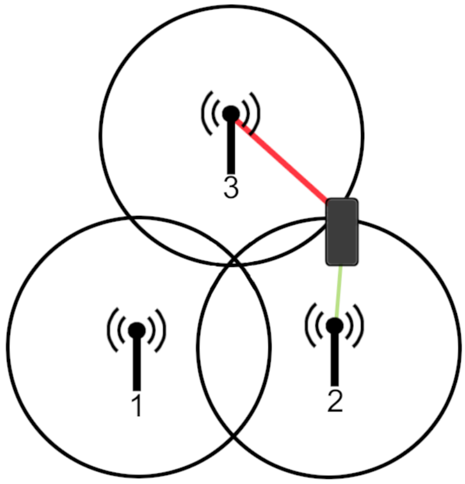

Figure 4: Localization via Proximity

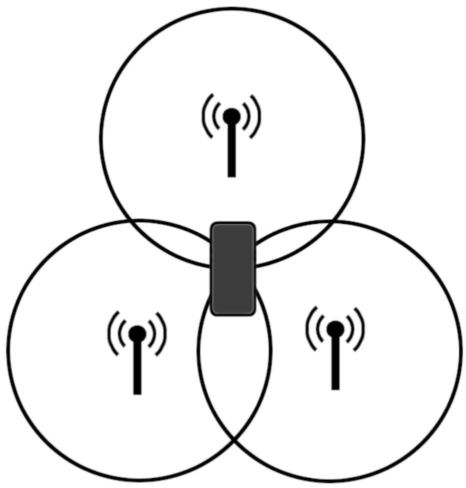

Figure 5: Localization via ToA, adapted from Farid et al. [2013]

as measuring the distance of a given object to multiple fixed reference points. In the case of trilateration, there are at least three different points used to determine the location; if there are more, it is called multilateration. Time of Arrival (ToA) systems focus on a synchronized arrival time at beacons of a signal sent by a mobile device. The received signal includes a time stamp so that the beacons receiving it can measure the distance to the mobile object via the time delay. Figure 5 shows how the location of the object is then determined by a intersection once the signal has been received and the distance has been measured. ToA is one of the best solutions for indoor environments, because it can filter out multipath effects, resulting in high accuracy. However, to perform at its best, ToA and the deployed positioning system must be implemented in such a way that they have precise knowledge of the start times of the transmitted signals. In addition, all beacons must be synchronized in time and distance to calculate values correctly. Setting up a system that has synchronized devices and measures time delays right will be very costly.

For instance, additional servers that calculate the time delay. Time delay measurement can also be a difficult task as it can be affected by dense environments caused by multiple livestock slowing down the signal between the beacons and the tracked device.[26]

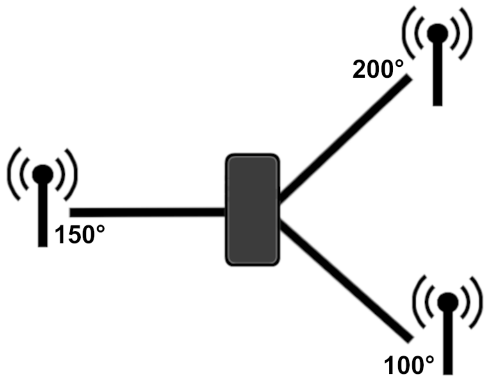

Figure 6: Localization via AoA, adapted from Farid et al. [2013]

Angulation often uses the Angle of Arrival (AoA) technique, which calculates the angle of an incoming signal from a known location. Using this method, AoA requires as few as two beacons to determine the location of an object in two-dimensional space. Adding more beacons increases the accuracy of the location. When tracking an object with three beacons, the object sends a signal to each of the beacons, the beacons can estimate the incoming angle, and find the object's location by finding the intersection of all lines that follow the measured angle. There are several drawbacks to this approach, especially when trying to track livestock indoors. Antennas must be added to the localization system to compute the incoming signal angles, and again to find the direction of a moving object. Also, all AoA methods require a line of sight between them to maintain reasonable accuracy, since all sorts of obstacles will reflect or alter the signals, making the accuracy of objects using AoA indoors unacceptable.[27]

Most modern wireless positioning systems use localization techniques that are either time or angle based, although both of these approaches suffer from the multipath effect, which leads to suboptimal accuracy when determining indoor locations. A more recent approach to locating an object is based on signal characteristics. More specifically, the strength of the incoming signal that can be measured when the signal consists of radio signals.[28] A multitude of positioning systems use some form of the so-called Received Signal Strength RSS and measure its value as Received Signal Strength Indicator RSSI, which is the actual signal power strength. Figure 7 again shows a simple representation of a system with three receiving beacons. The incoming signals are now referred to as RSSI 1, 2 and 3. The system is usually capable of performing complex signal propagation calculations based either on the different received strengths of signals 1, 2 and 3 alone, or using the strength of a signal from a known reference point to estimate the location of the object. The popularity of using RSSI values to determine a desired position can

[26] Farid et al. [2013]
[27] Farid et al. [2013]
[28] Farid et al., 2013, p. 3

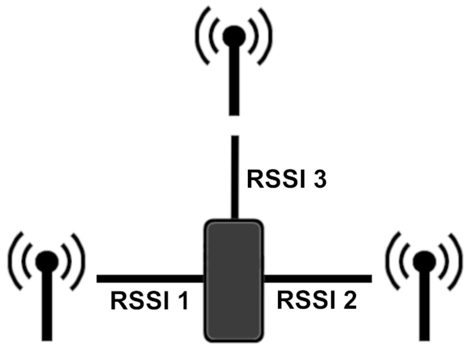

Figure 7: Localization via RSSI, adapted from Zafari et al. [2019]

be explained by the simplicity of the approach and the low cost intensity. However, even RSSI approaches can suffer from insufficient accuracy due to indoor noise and additional attenuation of the signal when encountering obstacles such as walls or livestock. To overcome the low accuracy, various filters and algorithms can be used to reduce and update or delete erroneously collected values.[29]

The final technique we will look at is scene analysis, also known as fingerprinting, which uses additional information gathered by environmental surveys to obtain features, called fingerprints, of the area where the positioning system is deployed. This is usually done using RSSI values collected before the positioning system is in active operation, referred to as offline data. When the system is in active use, incoming RSSI signals can be evaluated in real-time using the previously collected data to determine the position of the tracked object.[30] The mapping from the original stats to those measured by the localization in real-time can be done by a variety of algorithms, two of which are discussed below:

K-Nearest-Neighbour KNN is an algorithm that uses the real-time RSSI data and tries to find the K nearest neighbors in the collected offline RSSI data using the root mean square error. To obtain the location of the object, the nearest identified neighbors are averaged and then a suitable location is selected.[31]

Probabilistic methods work with the probability that the object is at a certain position. This likelihood is usually calculated using histograms and pattern matching algorithms, so that reference points within the system have certain probabilities, knowing the real-time RSSI values and evaluating offline data.[32]

Thus, no matter what methods are used for a scene analysis approach, the object position is always identified by mapping it to a specific part of a pre-selected grid. When analyzing the accuracy of this approach, one will come to the conclusion that it can lead to low accuracy depending on how detailed this grid is applied. Since a single part of the

[29] Zafari et al. [2019]
[30] Zafari et al. [2019]
[31] Zafari et al. [2019]
[32] Zafari et al. [2019]

grid is chosen as the location value, it can be seen as discrete instead of continuous. When trying to improve the coarse accuracy, a simple solution would be to use more beacons as reference points. In addition to the additional cost of providing more infrastructure, increasing the density of reference points also significantly reduces the difference in signal strength between two adjacent beacons compared to the natural indoor measurement noise, leading to an almost undecidable decision when determining a location. While fingerprinting approaches can still locate positions with good accuracy, they are also very sensitive to changes in the environment, making it important to keep offline data up to date. [33]

2.3.4 Localization Technologies

While the different techniques have their impact on building a positioning system and the data it produces, the technology used has a more significant role. This section covers a selection of existing technologies behind modern indoor positioning systems, as well as their advantages and disadvantages. A rough overview of the different commonly used localization technologies is given in Figure 8.

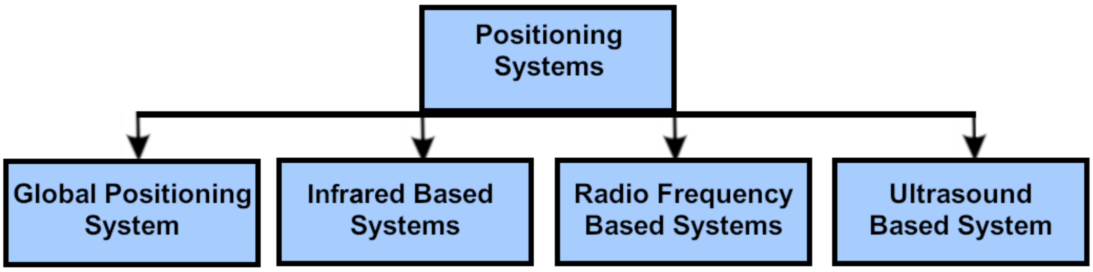

Figure 8: Localization Technologies, adapted from Farid et al. [2013]

Global Positioning System GPS is a popular approach for outdoor localization but struggles if used in indoor environments, because the satellite based technology is heavily influenced by obstacles that lie in the line of sight of the tracked object. Therefore, it is not further discussed in this thesis.[34]

Infrared Radiation IR is a wireless technology used in a variety of devices such as cell phones and televisions, and typically uses line-of-sight communication. It can also be used to detect and track objects, making it theoretically applicable to indoor localization.

[33] Zafari et al. [2019]
[34] Farid et al. [2013]

IR-based transmitters and receivers are typically small and lightweight, making them easy to deploy while maintaining high accuracy. However, this technology is costly to purchase and maintain, and is subject to interference from various light sources, such as sunlight or possibly even electric light sources.[35]

Radio Frequency Based Systems are often used for indoor positioning systems, primarily because one of the characteristics of the radio signal is that it can more easily penetrate obstacles such as walls, people, and livestock. This property achieves a high coverage metric with less installed infrastructure compared to other technologies. Radio frequency-based technologies can be further divided into narrow-band and wide-band. Since these signals are particularly suitable for indoor localization, they are given special attention in this thesis. This includes RFID, Bluetooth, WLAN and FM as representatives of the former category and UWB for the latter.[36]

Radio Frequency Identification RFID is a widely used wireless technology for locating objects via one-way radio signal communication between objects with different RFID tags and a reader. This technology is already being used in a variety of areas such as the automotive industry, supply chain networks or asset tracking, which makes it suitable for livestock tracking.[37] There are two different approaches how RFID systems can work:

- Active RFID: RFID tags periodically transmit a signal containing their ID over long distances, up to hundreds of meters from their reader. Although the advantage of working at these long ranges comes at the cost of loss of accuracy, making it nearly impossible to locate units below the meter mark, active approaches are low cost and can be easily added to objects.[38]

- Passive RFID: This approach has a much more limited range of just a few meters, but can operate without batteries. Compared to the active version, these are smaller and even lighter. Although they can be used for localization in some proximity-based scenarios, their limited range makes them unattractive for indoor localization.[39]

Ultra Wideband UWB is another radio signal technology with a short range but, as the name suggests, a wide bandwidth that is highly resistant to multipath, meaning it can penetrate a variety of obstacles. UWB achieves higher accuracy than the more established RFID with up to 10 cm localization accuracy. However, there are many drawbacks to UWB, the most important of which is the slow advancement of the technology due to the fact that the industry has yet to make much use of UWB in consumer products and mobile devices.[40] This also leads to expensive prices for UWB infrastructure when built into localization systems.[41]

WLAN or more specific WiFi has the main goal to connect devices to the Internet. The latest version of WiFi has a range up to 1 kilometer and is optimized for the IoT usuage. Most modern devices like smartphones or other mobile devices are able to use

[35] Farid et al. [2013]
[36] Farid et al. [2013]
[37] Farid et al. [2013]
[38] Zafari et al. [2019]
[39] Zafari et al. [2019]
[40] Zafari et al. [2019]
[41] Farid et al. [2013]

WiFi, making the technology suitable for indoor localization.[42] The localization process can take place without the need for additional software or changes of existing hardware, working without line of sight. As wlan is already strongly integrated in society, this is a cost effective approach for the general field indoor localization. Since the normal use of Wifi is designed for communication and not for indoor localisation, there is still a need for research into which algorithms and techniques in combination with Wifi will lead to optimal accury in indoor environments. Current research approaches such as RSS ToA, and AoA, as well as related techniques, have already led to very accurate values in practice, up to 23cm accuracy. [43]

Bluetooth, more specifically its latest version Bluetooth Low Energy BLE, also known as Bluetooth Smart, has an improved data rate and covers ranges up to 100 meters. The BLE technology can be paired with a variety of discussed techniques such as RSSI, AoA and ToA, while the most common BLE systems use some form of RSS. As with all systems based on RSS values, the location is therefore only estimated and of low accuracy. Overall, BLE is already a suitable approach for indoor localization, mainly because of its low power consumption and good range. Nevertheless, there are adaptations of it, such as Apple's iBeacons, a protocol specifically designed for proximity localization. The protocol allows beacons in the infrastructure of the positioning system to transmit signals at a periodic time interval. The messages sent include a 16-byte universally unique identifier (UUID) and optional major and minor values. For indoor livestock location, different barn buildings would be different major values and then different sections or the beacons themselves within a barn would be minor values. An application is deployed that receives the beacon messages and then uses the RSSI value to estimate the location of the object based on proximity detection.[44]

Hybrid positioning systems are the final selected radio frequency-based technology discussed in this section. As the name implies, hybrid localization is a combination of different positioning technologies. Most of the technologies that are good for indoor positioning do not work well or at all in outdoor environments and vice versa, e.g. GPS. In some areas, it is therefore advantageous to combine different technologies into a hybrid solution that achieves good performance metrics for both indoor and outdoor environments. This is especially useful for livestock, such as cattle, that are both indoors and outdoors.[45]

Ultrasound signals can also be used for localization and are the last category presented in this section. The position of objects is estimated by sending the signal from a tag placed on the object and then receiving it from beacons placed on the object. Unlike most of the introduced radio frequency based solutions, ultrasound is unable to penetrate walls and is reflected by most objects in indoor environments, making it a less suitable option for indoor livestock localization.[46]

Other indoor localization techniques that are less common include the usage of light and acoustic signals but will not be further discussed.

[42] Farid et al. [2013]
[43] Zafari et al. [2019]
[44] Zafari et al. [2019]
[45] Farid et al. [2013]
[46] Farid et al. [2013]

2.4 Data Filtering

In this subsection, the problem is demonstrated in a practical way by means of an introductory preparation of an example data set generated within the framework of the WeideInsight project. The reader is then introduced to the basics of data cleaning and a variation of common filters for indoor localisation is discussed. Before a filter solution for the project data is implemented in the next chapter, an evaluation of the work of Pastell et al. takes place, who applied several filter solutions for their livestock indoor localisation data.

2.4.1 Project Details and Data Exploration

As mentioned in the introduction, the WeideInsight project has taken an innovative approach to indoor localization for livestock. The solution involves the provision of a low-cost hybrid localisation system for the pasture and the barn. The chosen approach for monitoring the animals inside the barn, which is relevant for this work, is realised by placing Bluetooth beacons inside the barn using the iBeacon protocol. The cattle were technologically pre-positioned in order to get information about which beacon they are closest to, so that we can speak of a proximity-based approach. Figure 9 shows the monitored barn section with visualized, numbered beacons.

The data generated by this infrastructure provides information about the whereabouts of an animal by storing the closest beacon as a minor value with a timestamp containing the date and time of the measurement in a CSV file. This file contains a number of other attributes which are not considered for the purpose of this work.

The position is always measured at variable time intervals, where in the example dataset the first measurement took place on 2022-04-26 at 00:00:00 and the last measurement was recorded on 2022-04-27 at 23:59:59. So if you take the time difference between the first and last measurement of two days and multiply it by 86400 seconds on one day, you can divide that by the total number of measurements of 8594, so the average measurement is about every 20 seconds. The smallest intervals between measurements are in the single-digit seconds range, while the longest intervals between two measurements never exceed the one-minute mark.

This is important information when trying to comprehend the cattle movement to consider whether if the collected values are realistic behaviour or measurement errors. The following table displays the movement of one cattle over approximately a single minute by displaying the time at which the system located it and which beacon was the closest.

Time	Beacon
00:00:04	14
00:00:07	3
00:00:44	19
00:00:54	8
00:00:59	10
00:01:18	1
00:01:25	19

In this specific time interval the start beacon is number 14 and the end beacon is 19. The trajectory is represented as a red line in Figure 10. Although, the movement between

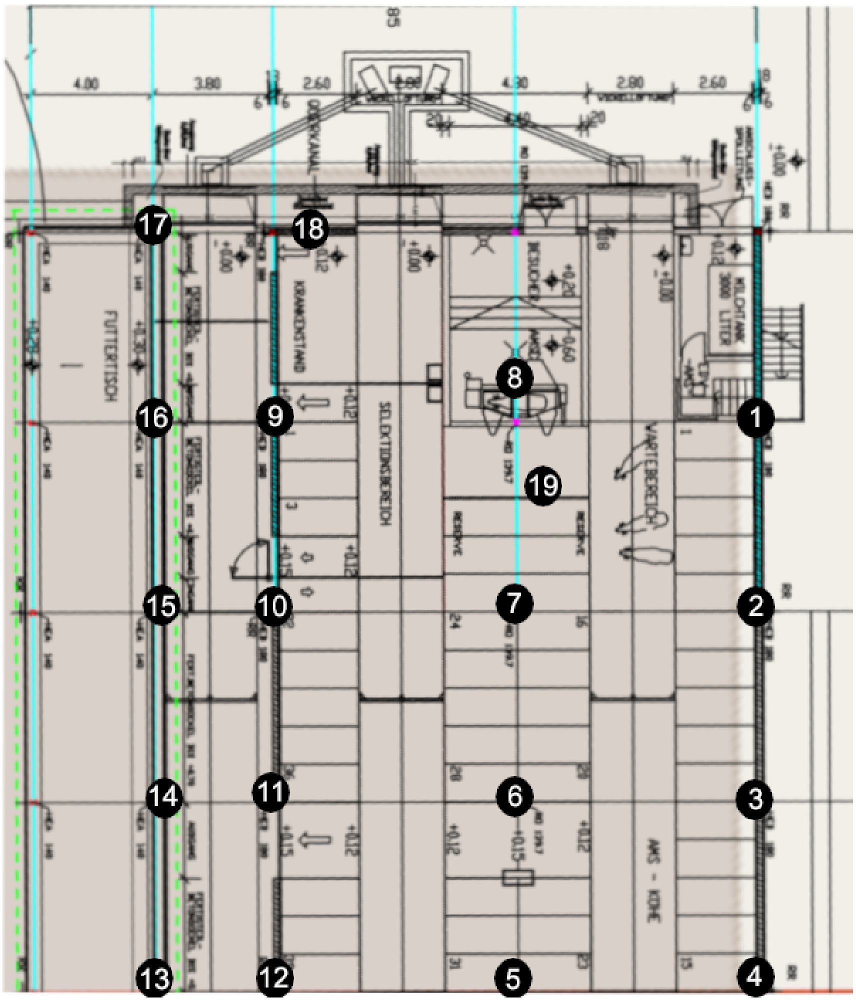

Figure 9: Beacon Distribution inside the Barn

beacons is portrayed as a straight line, the real route is unknown. This seems to be a lot of movement for the short time frame and leads to the assumption that the data set contains errors. The first location update takes place in a 3 second time interval (00:00:07-00:00:04) and travels to beacon 3 which is on the other sides of the barn. Judging by the barn plan this equals approximately a range of over 16 meters. The following measurements are all at beacon 19 and its neighbouring beacons with only smaller ranged updates, which suggests that the actual location might be somewhere around beacon 19 and is influenced of noise. This pattern proves to be the general occurrence for all of the observed cattle, leading to a starting point for a filter implementation.

When talking about improving the accuracy of localization data there needs to be some form of evaluation metric. In the case of the WeideInsight project camera footage from of three days got manually evaluated so that several cows have ground-truth available, narrowed down to specific locations, as can be seen in Figure 11, where AMS is the milking robot, T are drinking stations, F are feeding areas and A-E are lying areas.

The generated data set by this process stores information about the position with a time interval rather than each singular measurement. The following table shows an

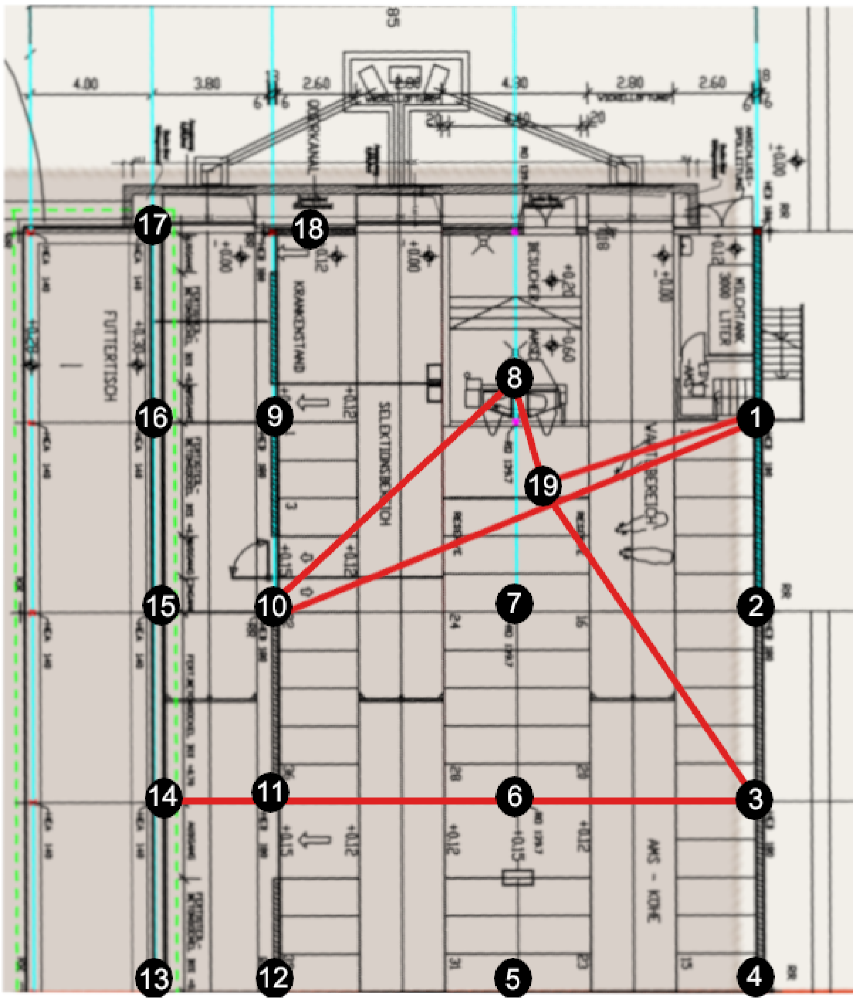

Figure 10: Cattle Movement based on the Table Data

example of the ground-truth, reduced to relevant information for this thesis:

Start	End	Observed Location
12:23:21	13:22:38	B9
13:05:07	13:09:39	F3-5
13:09:59	13:11:59	F3-5
13:12:16	13:14:38	F4-2
13:14:50	13:16:54	F4-1

There is multiple challenges attached to evaluate based on this ground-truth: First, time intervals have to be matched to the time points of the measurements and the specific locations have to be matched with minor values, determined by their proximity to the installed beacons. Additionally there are intersections of time intervals with differing observed locations as can be seen in row 1 and row 2 and time periods between recorded intervals, where no information is available, as can be seen in the remaining rows.

The realization of analyzing the ground-truth is that there is significantly less movement, even over longer periods of time than suggested by the raw data. From the second

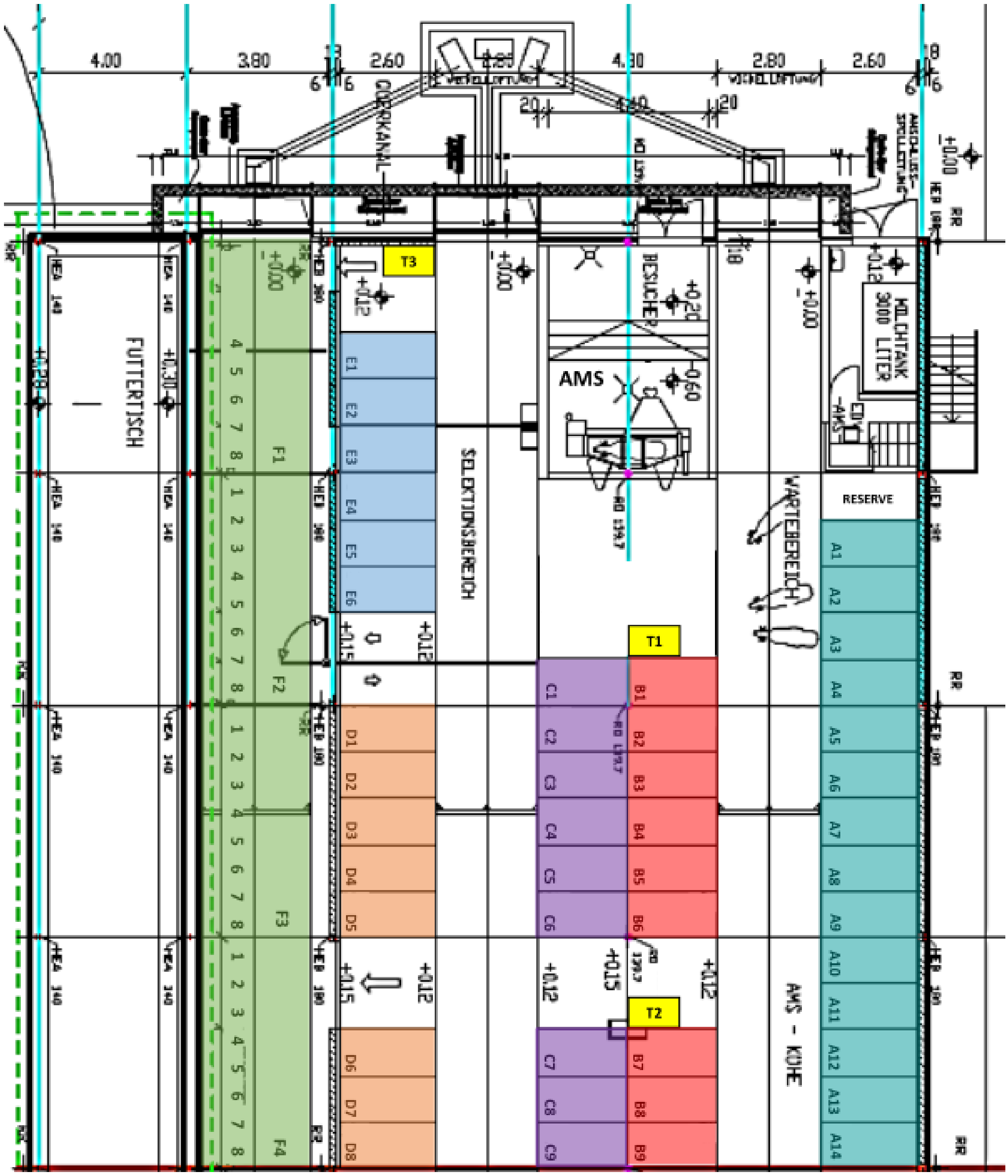

Figure 11: Barn Plan used for Groundtruth Data Set

row of the ground-truth table we can derive that the observed animal did not leave the food station 3 for four consecutive minutes, which could be translated to beacon 11,14,15 or 10. However, the unfiltered data, measured by the deployed infrastructure, for the same time period does not include one of those beacons, but seemingly their neighbouring beacons.

In summary, it can be assumed that the large distances that can be detected indicate large measurement errors, while strong noise is also present. The following subsections deals with the question of how to approach such scenarios.

2.4.2 Data Cleaning

Data cleaning is a discipline that specializes on detecting and removing errors or noise from data sets to improve the overall data quality. Data quality can be impaired by a variety of factors, in the case of indoor localization these mostly can be led back to used techniques and technologies miscalculations and the existence of the multipath effect.[47] Uncleaned data will lead to a wrong basis for decision-making and has the potential to damage livestock welfare and the farmer's profitability. [48]

A general approach to operating data cleaning consists of the two phases of error detection and error repairing. The error detection phase tries to answer the following three questions:

- 1. What to detect?
- 2. How to detect?
- 3. Where to detect?

The first question is concerned with the type of error, that can be detected by specifying some form of integrity constraints. These decide which values are considered errors. The second question deals with the type of the approach of detecting errors, examples being human involvement or a automated implementations. In the case of indoor localization data, where positions get estimated every couple of seconds over hours, days or even weeks a automated solution in form of a filter implementation is definitely required. The second question deals with the type of approach to detecting errors, examples being human involvement or automated implementations. In the case of indoor localization data, where positions get estimated every couple of seconds over hours, days, or even weeks an automated solution in the form of a filter implementation is required.

The error repairing phase tries to answer the following three questions:

- 1. What to repair?
- 2. How to repair?
- 3. Where to repair?

The first question is concerned with the target of the repair process. Repairing algorithms based on specified integrity constraints increase the data quality by removing identified values and updating them by reasonable values. This process usually takes only one type of error into account and replaces all of those - afterwards in a second process constraints can be changed to remove a different type of error. The second question deals with the automation of the repairing approach, which can be of a custom nature to fit a certain problem scenario or the usage of established solutions that are commonly used for similar types of errors, as can be seen by the filters that are introduced in the next section. The details of the final question are once again not relevant to the goal of this work and will be

[47] Rahm et al. [2000]
[48] Chu et al. [2016]

therefore not further discussed. The errors that will be repaired are the inside the minor column of the data sets. [49]

The introduced pattern of detecting errors and then repairing them will be used in the next chapter, when a filter solution is built for the WeideInsight project.

2.4.3 Filter Solutions

Before the implementation of an own filter solution takes place, this subsection is used to discuss frequently used filtering methods. When researching related literature on the topic of filtering data produced by indoor positioning systems, there are two filters specifically that are mentioned and implemented in most cases. These are the Kalman Filter and the Particle Filter, both being explained on a fundamental level in the following. Kalman and Particle filters are complex, however, there is also a variety of easier approachable filtering solutions discussed first.

RSSI-based localization via Bluetooth connections is prone to errors, to such an extent that the RSSI value changes even if the identical position is measured with the same infrastructure. There is a variety of possible types of RSSI filters that can be applied, one of the most common being the Moving Average Filter. This filter is easy to implement and its goal is to smooth out the differing signal strengths by getting rid of values that do not meet previously specified criteria. Values can be deleted by defining a dynamically moving bandwidth, where only signals inside the frequency are considered valid.[50] This smoothing process can also be achieved by the Kalman Filter.

Another simple filter approach is the one of a median filter, replaces values in a data set by calculating the median under a specified amount of neighboring measurements. This approach is of success when the observed animal is not moving since it removes all outliers. The higher the number of measurements taken into account for the calculation the more noise will be removed, ultimately leading to the possibility of cutting out the actual movement of the livestock.[51]

To some extent Kalman and Particle filters need to be customized to a specific scenario, however, more general custom approaches built to fix specific problems, e.g. the question of what to repair, is another type of a simple filter solution. The WeideInsight data shows noise as well as large errors, so custom approaches could focus on either removing one or the other. The next section shows such a filter, built to remove outliers, in detail.

The Kalman and the Particle Filter solutions can be categorized as Bayesian filters, even if their implementations and assumptions differ. Bayesian filtering is an approach based on probability theory to estimate the states of a system that is subject to noise.[52]

Kalman Filtering is treated as the optimal answer to improving the quality of data produced by most localization technology. Kalman's approaches use a technique that can be best described as a prediction and update model.[53] The filter recursively calculates positions based on a previous prediction and the actual measured data, which is beneficial

[49] Chu et al. [2016]
[50] Ordóñez-Camacho & Cabrera-Goyes [2018]
[51] Pastell et al. [2018]
[52] Chen et al. [2003]
[53] Shen et al. [2020]

for the necessary storage.[54]

The earliest version of the filter is also called the discrete Kalman Filter because it estimates discrete points in time in a linear model. The equations in detail are beyond the purpose of this work, which is why the following explanation is based on the Kalman Algorithm on a superficial level. [55] There are two equations, which lay the fundamental basis for the Kalman Filtering.

- Process equation:
$$x_{\mathrm{k+1}} = F_{\mathrm{k+1,k}} x_{\mathrm{k}} + w_{\mathrm{k}}$$

- Measurement equation:
$$y_{\mathrm{k}} = H_{\mathrm{k}} x_{\mathrm{k}} + v_{\mathrm{k}}$$

The underlying concept is the right definition of a state ($\mathrm{x_k}$), which requires information about the dynamic behavior of the system the filter is applied to. The discrete-time is referred to by the subscript k. Therefore the unknown state $\mathrm{x_k}$ can be described as the minimum required data to perform predictions on future behavior. This prediction is based on the measured data and denoted by vector $\mathrm{y_k}$. To calculate the state $\mathrm{x_{k+1}}$ the state ($\mathrm{x_k}$) is multiplied by the transition matrix $\mathrm{F_{k+1,k}}$, performing the change from time k to k+1. The noise is taken into account via $\mathrm{w_k}$. The Measurement equation is the actual measured value $\mathrm{y_k}$ at time k, being calculated by the measurement matrix $\mathrm{H_k}$ multiplied by the state $\mathrm{x_k}$ and the measurement of noise $\mathrm{v_k}$. Mutually solving these two equations is considered the Kalman filtering problem, since process and measurement calculations need to be optimized for an unknown state. Depending on the circumstances the problem can be categorized and be solved differently. Since the observed data values can be described as $\mathrm{y_1,y_2,...,y_k}$ the problem is either a real filtering with i=k, a prediction if i>k or a smoothing if $1 \leq \mathrm{i} < \mathrm{k}$ when dealing with state $\mathrm{x_i}$.[56] Those equations have the goal of ultimately leading to a minimal error covariance, a statistical measurement describing the correlation of values, eventually resulting in higher accuracy.[57]

In other words, the algorithm can be seen as a constant cycle of prediction and correction, where the prediction estimates the next measured value ahead of time, followed by a measurement update that adjusts the projected value by the actual measured value. This recursive approach takes past measurements into account so that an actual cycle put into words would include the following steps:

- 1. Predict the future Position
- 2. Predict the future error covariance
- 3. Calculate, for instance, the Kalman Gain
- 4. Update predicted Position with real Measurement
- 5. Update the error covariance

[54] Haykin [2004]
[55] Bishop et al. [2001]
[56] Haykin [2004]
[57] Bishop et al. [2001]

Where the Kalman Gain is one popular form of determining, how much the actual measurements will influence the predictions.[58]

The discrete Kalman Filter is a state space model that can be only applied to linear processes and relationships - however, the filter can be extended to fit nonlinear model by linearization to obtain the Extended Kalman Filter EKF.[59]

Particle Filtering in literature often also named sequential Monte Carlo SMC, consists of two steps of sequential importance sampling and resampling. A generic Particle Filter works with a set of randomly sampled values, called particles, to represent the initial state of a system. The behavior of these particles is predicted with a fitting propagation dynamic to come up with the general distribution for each point in time. Each particle has a weight based on its posterior probability. These decide which particles are useful to resample, with high values having a high resample rate and vice versa. Based on this method and its cyclic repetition the state of a system can be measured, using the estimated state as the start sample values to predict the next state. The exact measurements include once again a multitude of complex equations, which will not be further discussed in this work.[60]

2.4.4 Review of applied Filters on Livestock Data

After the presentation of modern localization systems and outlining their filters this section covers the application of filters to indoor livestock localization in real a life scenarios. As a reference, this section presents the achievements of Pastell et al. in Filtering Methods to Improve the Accuracy of Indoor Positioning Data for Dairy Cows, where suitable filter solutions were applied to the data of a UWB and AoA-based indoor positioning system monitoring cattle inside a barn.[61]

As a first step, data analysis for determining the lack of data quality took place, and three different types of errors got identified, of which each got a different repair approach assigned:

- Large Errors: A type of Error that is defined as a large jump that is unexplainable by realistic cow movement followed by the observed cattle location going back or close to the initial state in the next measurement. In their work, a custom approach called jump filter corrected this type of error.
- Normal Errors: A type of Error that is defined as a measurement inaccuarcy and cannot be categorized as large distance error. These were filtered by a Median filter and a Extended Kalman Filter - on the data, cleaned by the jump filter.
- Missing Data: A type of Error where the positioning system could not measure the location and therefore the data cell is empty.[62] These were filled in by linear interpolation because the signal usually returned near the last measurement.[63]

[58] Bishop et al. [2001]
[59] Haykin [2004]
[60] Speekenbrink [2016]
[61] Pastell et al. [2018]
[62] Pastell et al. [2018]
[63] Pastell et al. [2018]

The Jump Filter is a custom approach of Pastell et al. that updates localization values that appear to be long-distance errors. They observed three consecutive measurements and compared their distances. Fundamentally a value will be identified as a large error if the data point that is currently observed has traveled long distances from the data points that got measured right before and after, while those two points have a shorter distance in comparison. This approach needs constraints in place to specify which distance is categorized as too long, for instance, by setting a threshold for what distances can be achievable for cattle in the time window between measurements. This filter was always used before applying further filters.[64]

Extended Kalman Filter EKF is their approach to reducing existing noise. The Extended Kalman Filter in this scenario needs a cow movement model. However, this model is too complex to build precisely so an approximation took place. Their movement model of individual animals is a vector based on a three-dimensional position, facing direction, and speed. Further model details and assumptions are located in the original paper and will not be discussed further as their complexity is far beyond the scope of this work. [65]

Results of their work can be measured by looking at the effect the applied filter solutions had on the accuracy of the data. Figure 12 shows how many of the locations got calculated

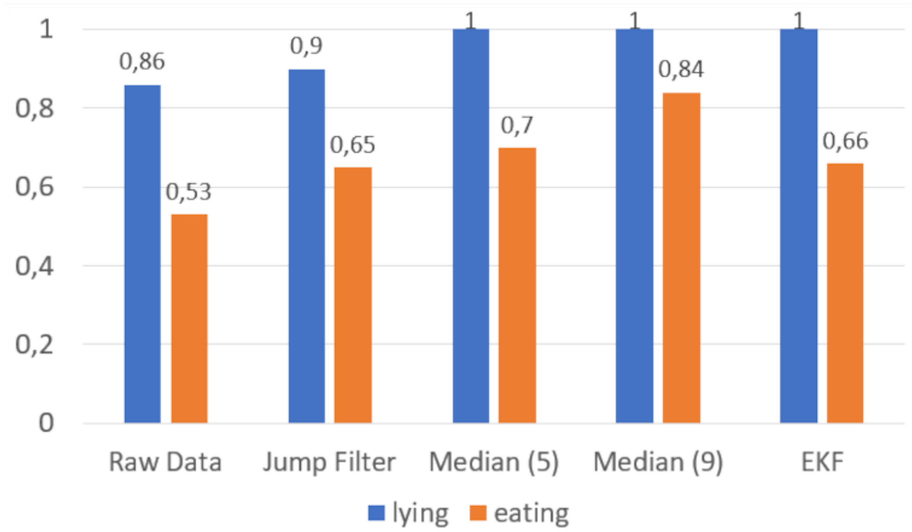

Figure 12: Percentages of correct Area Values by Categories lying and eating, adapted from Pastell et al. [2018]

to be in the right area of the barn. There is a distinction in their approach between a lying cow and an eating one at the feeding station, marked with different colors. Their original raw data has already a good hit rate of 86 Percent for the lying cows. In both cases as aforementioned, the Jump Filter was applied as a first step and then either a median filter

[64] Pastell et al. [2018]
[65] Pastell et al. [2018]

or an EKF was applied afterward. This first step led to a small increase of position values hitting the right area up to approximately 90 Percent. Step two included applying a five and a nine-point median filter, where in both scenarios all of the position data hit the real area of the monitored cows. An identical success was measured for the application of the EKF.

In the case of position validation of eating cows, the raw data hit rate was lower at 53 Percent. Again the application of the jump filter led to small increase of up to 65 Percent. The most successful approach to apply afterward was the nine-point median filter, achieving a hit rate of 84 Percent.

Pastell et al. also compared the trajectories of individual cow movement based on raw data and filtered data. The conclusion here was as expected that the median filter, especially the nine-point version undercuts the movement, whereas the EKF had a more accurate approximation of the real movement.

They concluded that they improved the data quality significantly using the combination of either Jump and Median Filter or Jump and Extended Kalman Filter. The median filter was of great success when the cows were not moving, while the EKF was the better option when trying to model the trajectory of moving cows. So choosing an optimal filter solution depends heavily on the application area. Especially the EKF variant's success depends heavily on a good implementation of their custom approach of the Jump Filter because the removal of outliers leads ultimately to the possibility of better location estimation.[66]

[66] Pastell et al. [2018]

Chapter 3

Implementation and Application

This chapter covers the approach to improve the gathered data of the WeideInsight project. The implementation is custom and based on the Jump Filter approach of Pastell et al. since it was their fundamental step that led to great success. Since the goal is to gain a higher data quality this chapter starts by answering the relevant questions of the data cleaning chapter. Afterwards, fundamental assumptions to calculate distances inside the barn are presented, followed by the discussion of necessary constraints.

3.1 Error Detection and Error Repairing

Detection Estimating positions based on radio frequency is prone to errors. From the data exploration in this work and the related literature, we know that larger miscalculations are a usual occurrence. Pastell et al. chose to eliminate the large outliers first before applying further filtering actions. So the goal of the filter implementation in this work is also to detect and remove outliers. Deciding what is considered an outlier can be formulated by a set of certain rules, defining realistic cattle behavior. The implementation of the detection and reparation takes place in Java code.

Reparation After the successful detection of outliers, they need to be updated with a more realistic value. Since the WeideInsight project works with proximity-based values of what beacon the cattle is closest to rather with exact positions, the error value will simply be replaced by setting it to the value of the last measurement, which should be accurate enough because of the frequency they take place.

3.2 Measuring Distances

To implement the Filter there has to be some fundamental form of measuring the distance between the beacons inside the barn. The chosen implementation will feature the beacons inside the barn as a complete, undirected, weighted graph structure.

- A complete graph is a graph where each node has edges to all other nodes.
- An undirected graph is a graph where an edge between two nodes is not directed, so the edge can be used in both directions.

- A weighted graph is a graph where each edge has a weight or some form of cost attached to it.

Using this type of graph the beacons are the nodes and every node is connected with each other, because of the complete property, making it possible to add weight values to edges between beacons.

To simulate distances a grid was applied to the barn-plan, making it possible to estimate coordinates for each beacon as a point inside the grid, as can be seen in Figure 13.

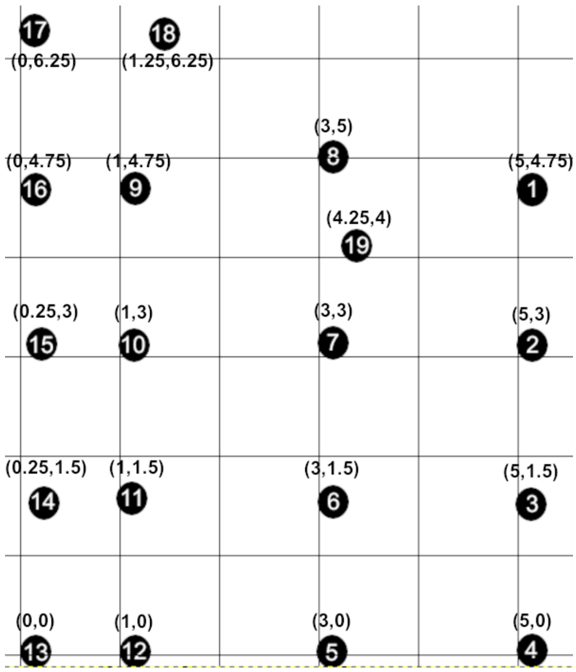

Figure 13: Beacon Coordinates

If a beacon touches the line of the grid it is considered to be in the row or column - otherwise it is determined proximity based on 0.25 steps. According to the barn-plan Beacon 13 and Beacon 12 have a distance of 3.8 Meters, choosing the grid to have a length of this distance open up the possibility to simulate real meter values. To find the distance between two points in meters the following equation was used:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} * 3.8m$$

3.3 Implementation of Constraints

Now that we can measure the distance between arbitrary beacons we have to decide what is considered realistic behavior and which position updates are considered to be errors because they surpass a certain distance threshold t. The constraint is based on the work of Pastell et al. and compares distances d of three points:

$$d(p_t, p_{t-1}) > t \wedge d(p_t, p_{t+1}) > t \wedge d(p_{t-1}, p_{t+1}) < ((d(p_t, p_{t-1}) + d(p_t, p_{t+1}))/2)$$

To be considered as an outlier, three conditions need to be fulfilled. First, the distance between the current point and the previous need to exceed a certain distance threshold. Second, this same threshold also needs to be cut by the distance between the current point and the next point in time. In the final step, the distance of the previous and the next point is measured, to be compared to the average distance of the previously calculated distances. A visualization of this approach can be seen in figure 14, where the red color signals a detected outlier and greed a valid measurement. In the example, the minor that contains the value of beacon 18 would be replaced by 15.

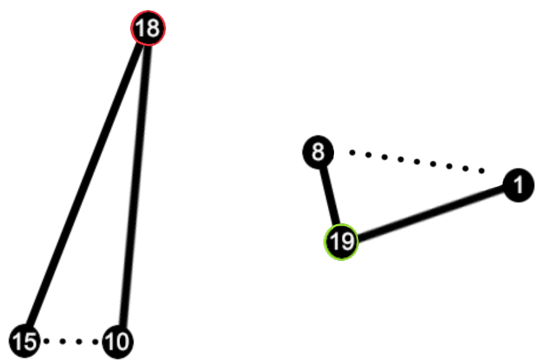

Figure 14: Outlier detection, adapted from Pastell et al. [2018]

The threshold used in the constraint is implemented as a variable but initially set to 3.8 meters. This can be justified by the placements of the closest beacons. For instance, it seems to be a realistic behavior that the cattle changes its closest proximity from Beacon 13 to Beacon 12, in the average 20-second time interval between measurements.

3.4 Summary

Described from a practical point of view, the final filter implementation can now read CSV data and match minor values with the discussed constraints. Any outliers found are replaced by the Beacon of the previous measurement, while all other values remain untouched. The next and final chapter will now measure what impact the filter solution has on the collected data.

Chapter 4

Evaluation and Conclusion

4.1 Difficulties

As mentioned in the data exploration, it seems to be very challenging to evaluate the implementation. An intuitive approach could be to match the ground-truth locations' proximity based on beacon names, then break the time intervals to time points to finally arrive at a format that can be matched with the raw data and the filtered data. This approach encounters several problems: The filter solution is specialized in filtering outliers, where the replacement will be the previous measurement. If these measurements are too noisy a comparison to the ground truth will still result in the same percentage for accurate positions. Replacing the observed location by the ground truth with beacons leads to undecidable situations where several beacons could be replacing the original measurement. Furthermore, some measurements cannot be validated since they took place between existing time intervals of the ground-truth. So it seems rather appropriate to check if the actual goal of the filter, to remove outliers, has worked successfully. The following section presents the achievements of the applied custom filter solution.

4.2 Evaluation

In the first step, we have a look at the comparison of raw data and filtered data, on the example of three different data sets. The following table includes the number of consecutive measurements and the count of how many outliers were changed. The mentioned anomaly was manually removed in the case of its appearance to ensure the filter is working properly. All data sets were filtered with a threshold of 3.8 meters.

Data sets:	1	2	3
Measurements:	8188	8552	7844
Original/Changed:	4656/3532	5032/3520	4213/3631
Difference:	43,13%	41,16%	46,29%

In all of the above cases, more than 40 percent of data was changed as it got identified as an outlier. Setting the threshold any lower would increase the changed values even further. A higher threshold would lead to less changed data, but at this point we need to evaluate further to see if changing the threshold is even necessary. One limitation of the filter that was noticed in connection with this comparison is that the first value in both data sets is always the same. This can be explained by the fact that the first measurement in a data set has no predecessor, so the first value that can be filtered is the second measurement.

Visualizing actual cattle movement would along with the introductory evaluation problem also include significant manual work, by mapping raw, filtered, and unfiltered data onto a barn plan. Figure 15 shows a one-hour comparison between the raw minor values MinorR and the filtered data MinorF. It is important to state that the lines between the measured time points are not interpretable as a real distance value and rather as the cattle leaving the proximity of beacon x and now it is closer to beacon y.

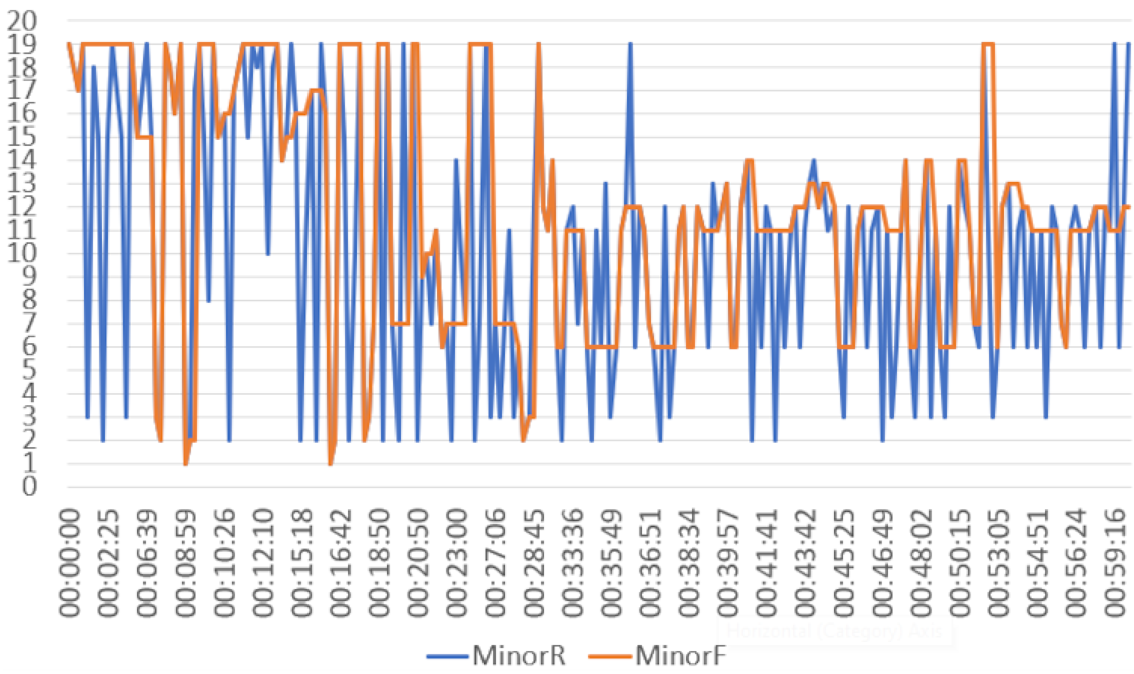

Figure 15: 60 Minute Comparison between Raw and Filtered Minor Values

This figure shows that there are significantly fewer changes in the proximity to beacons, as can be assumed to be more correct based on the ground truth. No change of the beacon value over multiple points in time can be identified on a straight horizontal line, which is displayed multiple times by the filtered values, however hardly for the raw data. This can be seen in further detail in Figure 16, which shows a more detailed view of the first 10 Minutes of the same data set.

The first half of the diagram seems to remove all the points that are considered to be unrealistic jumps. In the second part the orange line and the blue line overlap, which leads to the conclusion that no changes were made. The Verification that this is intentional and the correct action can be done by looking once again at the barn plan and its beacons, as in Figure 17. The change from Beacon 15 to Beacon 3 at approximately the 7 Minute mark, is followed by a measurement at Beacon 19 (blue line), which then is not considered to be an outlier by the stated constraint but still should be further investigated as a potential error source. The same scenario applies for the following overlapping measurements, while

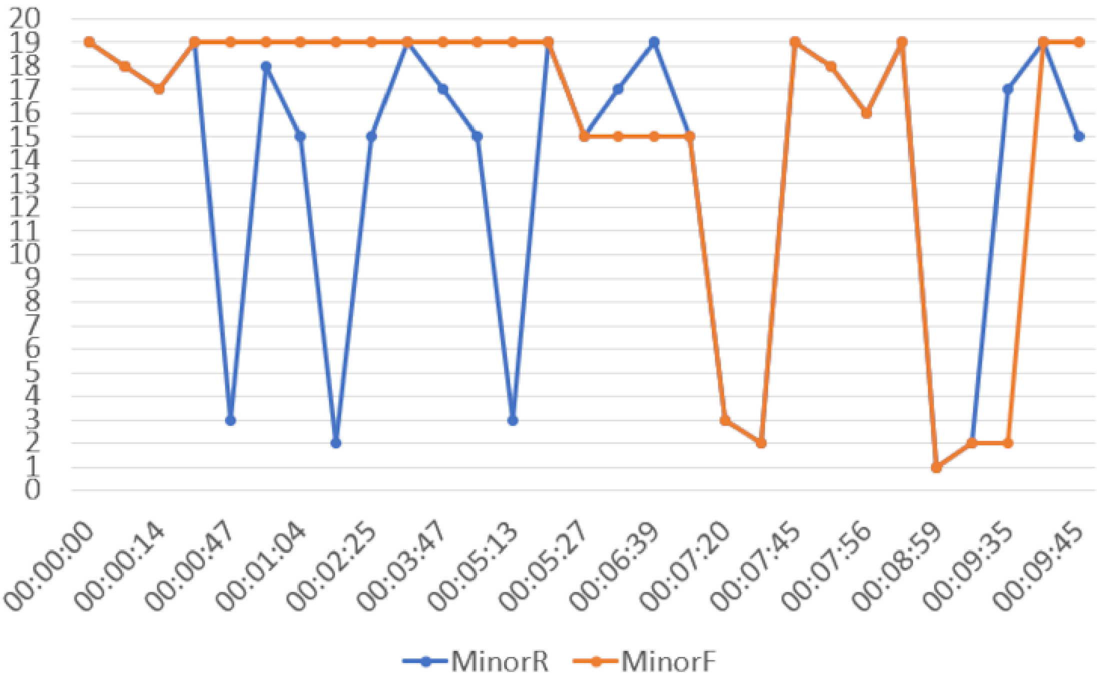

Figure 16: 10 Minute Comparison between Raw and Filtered Minor Values

the next position change demonstrates why the length of the line does not cohere with an actual real distance, as Beacon 2 and 19 (yellow line) are nearby.

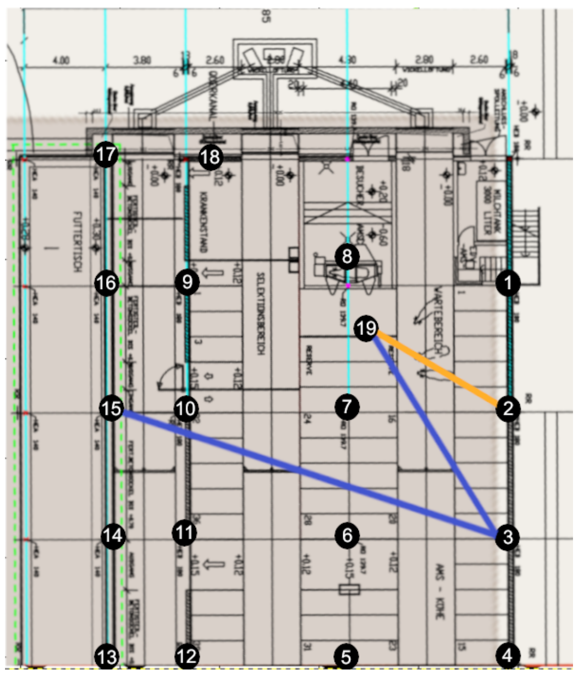

Figure 17: 10 Minute Comparison between Raw and Filtered Minor Values

Overall the filter seems to achieve its goal, filtering all the unrealistic jumps defined by the stated constraints, but as to be expected this does not remove all potential errors. However, the goal of this thesis is to improve the accuracy of the data sets has not been proven yet. This has to be done by involving the ground-truth in some way.

To combine already gained knowledge with a check of the actual accuracy of raw data and filtered data without significant mapping efforts, the following time intervals of the ground-truth are selected which show little or no movement for longer time intervals. In this case, the cow did not leave its lying area A5, which counts into the proximity of beacon 2.

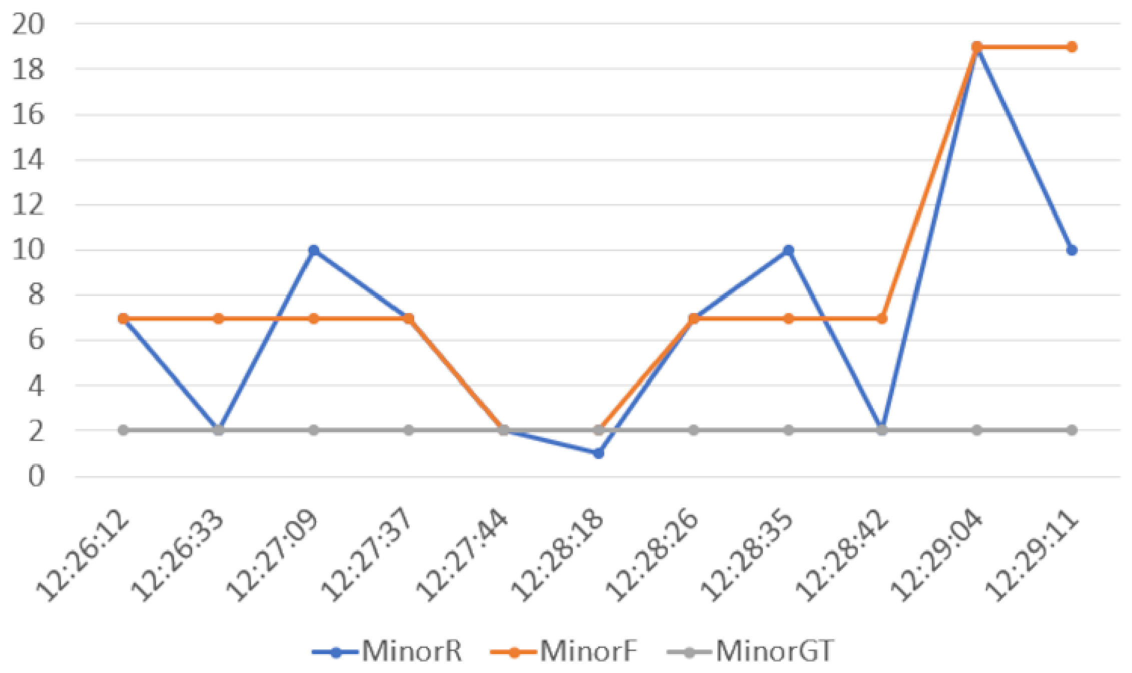

Figure 18: 3 Minute Comparison between Raw, Filtered and Ground-truth Minor Values

Figure 18 portrays the comparison of the three different available data types over a 3 Minute window. Although the cow never left the proximity of beacon 2, the raw data changes its minor value with each measurement. Of the eleven presented measurements the raw data tracks the position three times right. After the application of the filter, only two measurements can be validated to be true. However, the filter gets rid of the jumping measurements and tracks the object now only at beacons 7,2, and 19. Both, 7 and 19, are neighboring beacons to the real position. To further investigate this behavior another similar scenario with a significantly longer time interval will be analyzed.

This example covers approximately a 1-hour window, covering over 150 measurements without the cow leaving lying box B3, which closest beacon is number 7. Figure 19 shows the first 10 Minutes as a comparison of raw, filtered, and ground-truth data.

The figure shows similar behavior to what was already stated. Raw data jumps a lot, while filtered data jumps significantly less and either hit the right beacon or a neighboring one. In this special scenario, one could argue that beacons 19 and 7 are in such proximity that also beacons 19 values could be validated as true. Also, the filtered data and ground-truth data overlap significantly more than initially expected, promising higher accuracy.

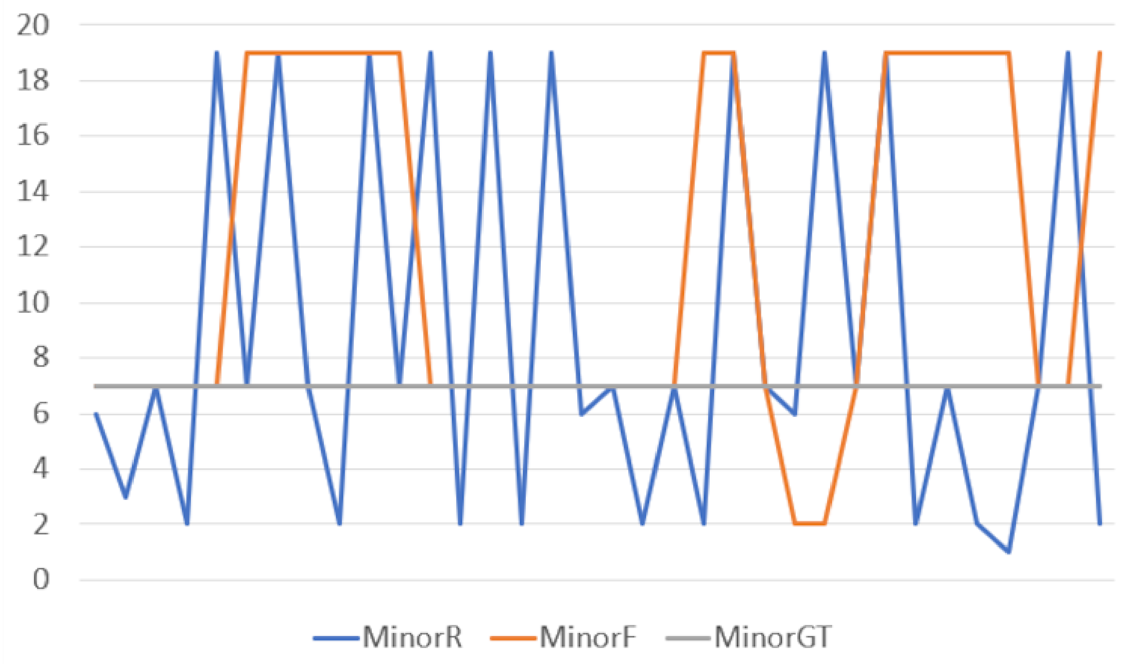

Figure 19: 10 Minute Comparison between Raw, Filtered and Ground-truth Minor Values

To prove this assumption, the remaining measurements of the time interval are presented in the table below.

Data sets:	Raw	Filtered	Filtered2
Measurements:	166	166	166
True/False:	50/116	89/77	128/38
Accuracy:	30,12%	53,61%	77,10%

In the one-hour time interval, the infrastructure estimated the positions of the observed animal 166 times. When comparing all the minor values of the different data sets, where Filtered2 is the filtered data set and Beacon 19 and 7 are both treated as accurate, a significant improvement was achieved. The raw data had only an approximate accuracy of 30 Percent, while the filtered versions show significant improvements of over 23 and 47 Percent. The remaining values that are still identified as false, once again show the property of being a neighboring beacon to the right one.

4.3 Conclusion

Precision Livestock Farming makes it possible to make all aspects of animal husbandry, such as animal welfare and profitability, significantly more efficient. Livestock Indoor Localization as a sub-discipline of PLF is already at a stage where many different technologies are available for successful livestock farming. Not all of these technologies are accessible to all farmers, mostly due to the economic aspects of acquisition or operation. Continued research in Smart Agriculture and related projects such as the WeideInsight project will in the long term ensure that these costs sink and a wider audience can successfully operate PLF with even better results.

However, the current mainly radio frequency-based infrastructures will remain dependent on suitable filtering solutions for the foreseeable future, as generated data sets are prone to errors. The resulting errors can usually be divided into large measurement errors and noise. A variety of established filters can help to reduce these errors and to increase the data quality, making it more informative, as this is the only way to use the full potential of PLF. In the past, solutions based on Particle and Kalman filters prevailed, but also custom filters show great potential. Inspired by the results from related literature, this thesis has developed a customized solution that aims to identify outliers and replaces them with more appropriate values.

The evaluation turned out to be challenging and pointed out several possibilities for future work. On the one hand, work can be done to automate the comparison of existing ground truth data and the data sets of the WeideInsight project to facilitate the quality control of applied filter solutions. These filter solutions should then focus on removing the noise since the implementation of this work successfully eliminated the irregular nature of the original raw data. Furthermore, assumptions of the evaluation are based ground truth of animals that do not move for long periods, which makes further investigation of filtering solutions on animals that are actually under movement interesting.

In summary, the implemented filtering solution achieves its goal by successfully filtering outliers, replacing them appropriately, and thus leading to a higher data quality. In a future step removing noise will lead to optimal accuracy.

Bibliography

Berckmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. *Rev. Sci. Tech*, 33(1), 189–196.

Bishop, G., Welch, G., et al. (2001). An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8(27599-23175), 41.

Chen, Z. et al. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics*, 182(1), 1–69.

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 international conference on management of data* (pp. 2201–2206).

De-La-Llana-Calvo, Á., Lázaro-Galilea, J.-L., Gardel-Vicente, A., Rodríguez-Navarro, D., Bravo-Muñoz, I., & Espinosa-Zapata, F. (2019). Characterization of multipath effects in indoor positioning systems by aoa and poa based on optical signals. *Sensors*, 19(4), 917.

Farid, Z., Nordin, R., Ismail, M., et al. (2013). Recent advances in wireless indoor localization techniques and system. *Journal of Computer Networks and Communications*, 2013.

Gygax, L., Neisen, G., & Wechsler, B. (2010). Socio-spatial relationships in dairy cows. *Ethology*, 116(1), 10–23.

Haykin, S. (2004). *Kalman filtering and neural networks*. John Wiley & Sons.

Marino, L. & Allen, K. (2017). The psychology of cows. *Animal Behavior and Cognition*, 4(4), 474–498.

Mautz, R. (2012). Indoor positioning technologies.

National Institute of Standards and Technology (2023). multipath. [Online; accessed 10.05.2023].

Ordóñez-Camacho, D. & Cabrera-Goyes, E. (2018). An adaptive-bounds band-pass moving-average filter to increase precision on distance estimation from bluetooth rssi. In *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)* (pp. 823–832).: Springer.

Pastell, M., Frondelius, L., Järvinen, M., & Backman, J. (2018). Filtering methods to improve the accuracy of indoor positioning data for dairy cows. *Biosystems Engineering*, 169, 22–31.

Phillips, C. (2008). *Cattle behaviour and welfare*. John Wiley & Sons.

Rahm, E., Do, H. H., et al. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23(4), 3–13.

Shen, Y., Hwang, B., & Jeong, J. P. (2020). Particle filtering-based indoor positioning system for beacon tag tracking. *Ieee Access*, 8, 226445–226460.

Speekenbrink, M. (2016). A tutorial on particle filters. *Journal of Mathematical Psychology*, 73, 140–152.

Tscharke, M. & Banhazi, T. M. (2016). A brief review of the application of machine vision in livestock behaviour analysis. *Agrárinformatika/Journal of Agricultural Informatics*, 7(1), 23–42.

Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3), 2568–2599.

Zhao, J.-c., Zhang, J.-f., Feng, Y., & Guo, J.-x. (2010). The study and application of the iot technology in agriculture. In *2010 3rd international conference on computer science and information technology*, volume 2 (pp. 462–465).: IEEE.

Erklärung

Ich erkläre hiermit gemäß § 9 Abs. 12 APO, dass ich die vorstehende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Des Weiteren erkläre ich, dass Inhalt und Wortlaut der beiden Fassungen (digital/in Papierform) identisch sind und zur Kenntnis genommen wurde, dass die digitale Fassung einer durch Software unterstützten, anonymisierten Prüfung auf Plagiate unterzogen werden kann.

Datum Unterschrift