PHD DISSERTATION

Pattern Recognition Methods for Reduction of Human-Wildlife Conflicts

by Kim Arild Steen

AARHUS UNIVERSITY

Pattern Recognition Methods for Reduction of Human-Wildlife Conflicts

Pattern Recognition Methods for Reduction of Human-Wildlife Conflicts

PhD Thesis by

Kim Arild Steen

Aarhus University Department of Engineering, Denmark

AARHUS UNIVERSITY
DEPARTMENT OF ENGINEERING

River Publishers
Aalborg

ISBN 978-87-93237-06-3 (e-book)

Published, sold and distributed by:
River Publishers
P.O. Box 1657
Algade 42
9000 Aalborg
Denmark

Tel.: +45369953197
www.riverpublishers.com

"A person who never made a mistake never tried anything new"

-Albert Einstein

Abstract

Conflicts between human activities and wildlife are an increasing problem, and in many parts of the world damage caused by wildlife creates significant economic challenges to human communities. Methods for reducing human-wildlife conflicts are either ineffective, time consuming or costly. Both lethal and non-lethal techniques are used for reduction of human-wildlife conflicts. However, the use of lethal methods is often controversial, as there is a public desire to co-exist with wildlife.

This thesis is focused on human-wildlife conflicts in agriculture. The research has been driven by two very different conflicts between wildlife and agricultural activities, namely the problems with large flocks of birds in agricultural fields and wildlife mortality during mowing operations. The scientific contributions of this Ph.D. thesis focus on how sensor technologies and pattern recognition methods can be applied in the design of solutions for the reduction of human-wildlife conflicts in agriculture and thereby contribute to more ethical and efficient wildlife damage management. The result is a collection of contributions to the design of pattern recognition and signal processing algorithms to enable the use of smart sensing in the solution for human-wildlife conflicts. The achieved results are a significant step towards a more efficient sensor based solutions for wildlife-friendly farming and reduction of human-wildlife conflicts within agriculture.

Resumé

Konflikter mellem menneskelige aktiviteter og dyreliv er et stigende problem, og i mange dele af verden skaber skader, forårsaget af dyreliv betydelige økonomiske udfordringer. Metoder til at reducere konflikter mellem menneskelige aktiviteter og dyr er enten ineffektive, tidskrævende eller dyre. Både dødbringende og udskadelige metoder anvendes til reduktion af menneskelige aktiviteter og dyreliv. Brugen af dødbringende metoder er dog ofte kontroversiel, da der er et folkeligt ønske om tolerance og beskyttelse af dyr.

Denne afhandling fokuserer på konflikter mellem dyr og menneskelige aktiviteter i forbindelse med landbrugsproduktion. Forskningen har været drevet af to meget forskellige konflikter mellem dyreliv og landbrugsaktiviteter, nemlig problemer med store fugle flokke i marker og påkørsler af dyr i forbindelse med høst. De videnskabelige bidrag i denne Ph.D. afhandling er fokuseret på, hvordan sensorteknologier og metoder fra mønstergenkendelse kan anvendes i design af løsninger til reduktion af konflikter mellem dyr og menneskelige aktiviteter i landbruget, og dermed bidrage til en mere etisk og effektiv løsning. Der præsenteres adskillige bidrag til design af signalbehandlings - og mønstergenkendelsesalgoritmer, der kan anvendes til intelligent brug af sensorer til løsning af konflikter mellem mennesker og dyr. De opnåede resultater er et væsentligt skridt i retning af en mere effektiv sensorbaseret løsning på vildtvenlig landbrug og reduktion af konflikter mellem dyr og menneskelige aktiviter i landbruget.

Acknowledgements

I would like to thank my supervisor Henrik Karstoft from the Department of Engineering, University of Aarhus, for providing valuable feedback and motivation during my Ph.D. project. It has been a great pleasure to work with Henrik, who has always left the door open for theoretical discussions and personal guidance. I would also like to thank my co-supervisor Ole Green from the Department of Engineering, University of Aarhus, for giving me the opportunity and challenge of pursuing a Ph.D. degree. Both Henrik and Ole have been an inspiration to constantly perform well during the Ph.D. project. For this, I thank you.

Ole Roland Therkildsen has also been an important part of this Ph.D. project. His dedication to research within human-wildlife conflicts and his skills within scientific writing have been very rewarding. I would also like to thank Rasmus Nyholm Jørgensen who has helped me during my work regarding wildlife-friendly farming. His constant flow of new ideas and critical questions have been an eye-opener for my approach to this research.

I would like to thank my colleagues Morten Stigaard Laursen, Thomas Jensen, Martin Christiansen and Peter Ahrendt for their contributions to technical discussion, feedback and motivational working environment. I have enjoyed discussing my work with you, and I am thankful for your patience during this.

I would like to thank James H. McClellan for giving me the opportunity to visit Georgia Institute of Technology during my Ph.D. project. This experience was rewarding both personally and professionally.

I would also like to thank my parents Jytte and Bjørn for their constant support. You have always let me make my own choices and never tried to convince me otherwise. This has been a gift, and for that I am very grateful.

Last, but not least, I would like to thank my family Mette, Johannes and Aksel for their patience and understanding during this Ph.D. project. You were willing to leave everything behind to join me when my research led me abroad. An experience I will always remember. For that, I thank you. You are all a very important part of my life.

Contents

Part I

Overview

1

Introduction

Human-wildlife conflicts are defined by the interaction between wild animals and people, which result in a negative impact on people or their resources, or wildlife or their habitat. These conflicts occur when growing human populations or activities overlap with established wildlife territory or migration. Conflicts between wildlife and humans are increasing [88], and in many parts of the world damage caused by wildlife creates significant economic challenges to human communities.

Wildlife is often referred to as a resource [35]. Humans may benefit positively from wildlife at various levels including ecologically, economically, scientifically and recreationally, like hunting and enjoying nature. Negative values include loss of agricultural production, destruction of property and wildlife-related human injuries (collisions etc.). Hence, the goal of wildlife damage management may be defined as the attempts to reduce the negative value, and thereby increase the net value of wildlife resources [35].

Human-wildlife conflicts are not an isolated phenomenon, and there are many scenarios where wildlife can cause serious problems for human activities. This includes air travel, industry, wind power, tourism, sports events, agriculture and many more. Both the nature of the conflict and the complexity of the wildlife damage management setup can be very different in these scenarios, e.g. in industry the problems can be isolated to a smaller area, whereas airports cover a much larger region.

Effective management of wildlife is vital to reduce the negative impact of human-wildlife conflicts. A wide range of devices and methods are used in wildlife damage management; however, their effectiveness is often highly variable due to habituation or limited impact [50, 133]. Habituation is the gradual adaptation to frightening stimuli, and it is a major limitation to current frightening devices. The use of frightening devices is a popular approach to wildlife damage management. However, both lethal and non-lethal techniques have been used in the struggle to reduce human-wildlife conflicts. The use of lethal control methods is often controversial to control wildlife damage [50]. The public accepts the use of lethal methods when there are no alternatives. However, they also believe that continued research towards non-lethal methods is needed [110]. This motivates research towards more efficient, ethical and wildlife-friendly methods for wildlife damage management.

To narrow the scope within human-wildlife conflicts, this thesis concerns human-wildlife conflicts within agricultural production. The scientific contributions of the Ph.D. project focus on how sensor technologies and pattern recognition methods can be applied in the design of solutions for the reduction of human-wildlife conflicts in agriculture and thereby contribute to more ethical and efficient wildlife damage management.

1.1 Human-Wildlife Conflicts in Agriculture

During the last decades, strong competition in the agricultural sector has resulted in the need for high efficiency agricultural production, and the development of high-efficiency farming equipment. These developments, along with a dramatic increase in a number of animal populations, have led to serious conflicts between wildlife and agricultural activities all over the world.

1.1.1 Conflicts with bird flocks

Large flocks of birds, such as geese, starlings, gulls and rooks, may damage fruit trees, feed on livestock food supplies, and cause severe damage to newly sown crops [50, 108]. This is an economic challenge to agricultural production, and it has increased over the past decades due to an exponential increase in animal populations[1]. An example of a large flock of birds is shown in figure 1.1, where a large flock of geese has landed on an agricultural field. These flocks may consist of thousands of birds, which are able to damage large areas within a short time.

Gas exploders or visual stimuli, such as pop-up scarecrows, are often used as frightening devices in this context. However, both methods are subject to habituation, which is often the major limitation associated with the use of frightening devices [103]. When habituation occurs, the birds perceive the disruptive stimuli as part of the acoustic or visual scene, which no longer pose as a threat. Gas exploders may also disturb nearby residents due to high noise levels [50]. Animal activated methods have also been used in agriculture. In most cases, these frightening devices are non-specific, so they can be activated by any animal, and not only by the target species. This increases

[1] This is especially the case for goose populations

the risk of habituation. Methods to delay habituation include changing the location of devices, altering the periodicity of stimuli [72] or the use of a combination of devices [50]. This can be very time consuming, which is undesirable in an efficient agricultural production. However, a type of stimuli that are promising for future frightening devices is bioacoustics [50].

Figure 1.1: *A large flock of geese are foraging in a field with crops. Photo by Domen Stanis*

1.1.1.1 Bioacoustics

Bioacoustics refers to animal communication signals, which includes calls like alarm and distress calls. Birds and other animals are less likely to habituate to their own alarm and distress calls [24, 50]. Furthermore, animals are more sensitive to alarm and distress calls, which affects them physiologically [126]. This means that alarm and distress calls are more meaningful to animals, even at lower intensities, than other sounds [50]. Frightening devices may therefore apply these stimuli at lower sound pressure compared to noisy gas exploders.

Frightening devices using bioacoustic-based stimuli have been used in various applications including cornfield protection [51], soybean field protection [21] and urban roosts [53]. A commercial system that utilize bioacoustics is *The GooseBuster*[2], which is specifically designed for Canada geese. The system is based on alarm, alert and distress calls which are played back from multiple speakers. The calls are altered in sequence of play, frequency, duration and interval, thus providing variability in the frightening stimuli. In

[2] http://www.bird-x.com/goosebuster-products-50.php?page_id=104

[136], the effect of the system is studied in three controlled experiments. The author concludes that the use of timed alarm and distress calls alone experience habituation, however, ”on-demand” playback and reinforcement (using screamers and bangers) proved to be efficient to avoid habituation.

Habituation to bioacoustics have been reported in [16, 50, 137]. In [137], the authors argue that this may be a result of the fact that the geese, used in the experiment, were not able to escape the enclosed study site. In [24], the authors conclude that alarm and distress calls are more resistant to habituation than other sounds, but a pest controller needs to be able to identify species, as most calls are species-specific.

The use of bioacoustics seems promising, however, more research is required within effective management of bird flocks. The reported strategies and limitations of existing methods have been part of the literature review in this Ph.D. project and have led to the concept of an adaptive frightening device.

1.1.1.2 An adaptive frightening device

This Ph.D. thesis includes work and scientific contributions regarding the development of an adaptive frightening device. Here, sensor technologies and pattern recognition methods have been applied to improve existent frightening strategies and thereby reduce habituation. These contributions are presented in Chapter 2 of this thesis.

1.1.2 Mowing operations and wildlife

The increased need for high-efficiency agricultural production has resulted in efficiency improvement of agricultural machinery, including machines for grass mowing. This involves working speeds beyond 15 km/h and working widths of more than 14 meters (see figure 1.2). Furthermore, many species nest or seek cover in cultivated forage fields, where mowing operations take place repeatedly and at a time when there is overlap with breeding season. This contributes to wildlife mortality, as it almost impossible for the farmers to see, and react to, the presence of the animals in the fields.

Several species are likely to be negatively affected by mowing operations. These include not only common farmland species, but also endangered species like the corncrake [56]. In particular, the nests of ground nesting bird species like grey partridge or pheasant are vulnerable to farming operations in their breeding habitat both as a result of the nests being destroyed [92] or the incubating female being killed or injured [71]. In mammals, the natural instinct of e.g., leverets of brown hare and fawns of roe deer, to lay low and

Figure 1.2: *An example of mowing operations. The machine in the picture has a typical working width of 14.6 meters and working speeds from 18 km/h to 22 km/h. Photo by Ole Green.*

still in the vegetation to avoid predators increase their risk of being killed or injured in farming operations [71] (see figure 1.3). As a result of the increase in both working speed and width, adults of otherwise mobile species, e.g., fox and roe deer, are now at risk of being killed or injured in farming operations as they may be unable to escape the approaching machinery.

Although the extent to which wildlife populations may be affected negatively by farming operations is difficult to assess, there is no doubt that the risk of wild animals being accidentally injured or killed during routine farming operations has increased dramatically over the years. Besides the potential effects on wildlife populations, fodder contaminated with carcasses of animals may impose a health hazard for livestock from infection by the bacteria *Clostridium botulinum* causing botulism [48]. This may lead to substantial commercial loss.

Moreover, an aspect that has only received little attention is the mental stress imposed on the farmers, who occasionally face an injured animal during farming operations. The health and safety issue associated with the farmer having to do a mercy killing without the professional expertise should not be ignored.

Various methods and approaches have been used to reduce wildlife mortality resulting from farming operations. Delayed mowing date, altered mowing patterns (e.g., mowing from the center outwards [55]) or strategy (e.g., leaving edge strips), longer mowing intervals, reduction of speed or higher cutting height [55] have been suggested to reduce wildlife mortality rates.

Figure 1.3: *Here is an example of five fawns being killed during mowing operations. Trained dogs where used prior to mowing, however due to human error, one field was not checked for animals. Photo by Chresten Bergh.*

Likewise, searches with trained dogs prior to mowing may enable the farmer to remove e.g., leverets and fawns to safety, whereas areas with bird nests can be marked and avoided. Alternatively, various scaring devices such as flushing bars [55] or plastic sacks set out on poles before mowing [67] have been reported to reduce wildlife mortality.

However, wildlife-friendly farming often results in lower efficiency. Therefore, attempts have been made to develop automatic systems capable of detecting wild animals in the crop without unnecessary cessation of the farming operation. For example, a detection system based on infrared sensors has been reported to reduce wildlife mortality in Germany [58]. In [64] a UAV-based[3] system for roe deer fawn detection is presented. The authors show that thermal imaging can be used to detect roe deer fawns based on aerial footage. However, the detection is performed manually, and should be automated to increase efficiency.

1.1.2.1 Wildlife-friendly farming

This Ph.D. thesis includes work and scientific contributions regarding the development of methods for wildlife-friendly farming. Here, sensor technolo-

[3] Unmanned Aerial Vehicle

gies and digital image processing methods have been applied to in the attempt to develop automatic detection of the animals hiding in the vegetation. This is presented in Chapter 3 of this thesis.

1.2 Smart Sensing in Agriculture

The use of sensor technology and pattern recognition is not a novel concept in agriculture. Both microphones, cameras (including multi-spectral and thermal cameras), accelerometers, lasers and other sensors have been used to increase efficiency or gain vital knowledge of agricultural production and operations.

Figure 1.4: *An example of using microphones to measure pig welfare - http:www.soundtalks.be*

1.2.1 Acoustics

States of mood or emotion are often accompanied by specific behaviors, with emission of sounds being one of them. This means that farm animal vocalizations may work as an indication of the well-being of livestock in agricultural production [81]. In agricultural production, knowledge of changes in the well-being of farm animals is very important. Abnormal changes in behavior may indicate emerging disease, which, if not detected, may spread and result in serious economic challenges. The advantage of using acoustic measurements to estimate the health and behavior of farm animals is that it is inexpensive, non-invasive and continuous. An example of a commercial application, which measures health status of pigs, is shown in figure 1.4.

In research regarding acoustic monitoring of livestock welfare [81, 82], three important tasks, which should be applied in a framework for audio-based recognition of animal welfare, is defined: 1) Expert knowledge[4] of the relationship between a specific vocalization and the emotional or health state of an animal 2) Descriptive features of the vocalizations, and 3) Statistical methods to compare these features. These tasks are similar to methods applied in the field of pattern recognition.

A common characteristic of using microphones as smart sensors in agriculture is that the recorded data needs to be processed in order to gain value from the efforts. In most cases, the processing step involves some degree of pattern recognition, and research within this field have been highly influenced by human speech recognition methods, and have incorporated both feature extraction and pattern recognition methods from this [25, 76].

The vocalizations of animals range from periodic vocal-fold vibration to completely atonal turbulent noise [81]. Therefore, a wide variety of acoustic features have been used to describe the vocalizations of animals. This include *time domain features*, such as energy and duration, *frequency domain features*, such as fundamental frequency, harmonics and bandwidth, *cepstral features*, known from human speech recognition [22, 42] and *coding models*, such as linear predictive coding [82]. For recognition based on these acoustic features, pattern recognition algorithms such as *Hidden Markov Models* [25, 73, 81], *Gaussian Mixture Models* [25, 130] and *Artificial Neural Networks* [81, 82] have been utilized.

1.2.2 Vision

Like acoustic measurements, visual measurements offer a non-invasive method to monitor livestock; or crops. Given the appropriate camera technology it is possible to record and recognize the behavior and health status of livestock [17, 114], or even distinguish between plant and weed during weed control [119]. Cameras have a limited range compared to microphones, however, the level of information and the number of potential applications is rich.

Camera technology has been used in a variety of different applications in agriculture. The development of image processing algorithms is often data driven, and there is no rigorous theoretical framework. However processing steps such as segmentation, thresholding, tracking and pattern recognition is often part of the solution [17, 93, 105]. Object recognition in images is not

[4] Can be derived from experiments

a trivial task, however, in agriculture, another challenge is the uncontrollable environment. An example of this is shown in figure 1.5, where a closed box is used to control lighting conditions in experiments regarding automated detection and recognition of crops and weed.

Figure 1.5: *Here is an example of a sophisticated computer vision setup. This field robot detects and recognizes plants from weeds to reduce the use of pesticides. Apart from multispectral camera technology, the robot is equipped with Real Time Kinematic-GPS for navigation. Photo by Morten Stigaard Laursen.*

1.2.3 Other Sensors

Other sensors used in agriculture include devices like Global Positioning Systems (GPS) [121] or other wireless transmitters in a wireless sensor network [99], or accelerometers, measuring the movement of specific parts of farm animals [98]. These sensors are efficient, however invasive, and, therefore, not suitable when dealing with wildlife.

1.2.4 Pattern recognition

The fundamental problem in pattern recognition is to provide a reasonable answer for a given input. The output is bound to be uncertain, due to statistical variations, and is often defined as the "most likely" answer. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of classes. However, pattern recognition is a more general problem that encompasses other types of output as well, such as regression, sequence labeling and parsing [23].

Pattern recognition can be roughly divided into two sub-categories: *supervised learning* and *unsupervised learning*. Supervised learning assumes that a set of training data is available. Training data consists of data that have been properly labeled with the correct output for a given input. The task in supervised learning is then to generate a model that performs as good as possible on the training data, and generalize as good as possible to new data. If the model perform very good on training but poor in new data, the model might have been over-fitted to the training data [23].

Unsupervised learning, on the other hand, use data that has not been labeled, and attempts to find patterns in the data that can then be used to determine the correct output value for new data. A third category in pattern recognition is *semi-supervised learning*, which is a combination of the two. Here, the learning is based on both labeled and unlabeled data. Typically, this method is used if it is very time consuming or expensive to get enough labeled training data.

There are examples of pattern recognition in a wide variety of research and commercial applications. Pattern recognition is used in speech recognition, optical character recognition[5], face recognition in commercial cameras and online image tagging software, spam filters and many more. The great interest in pattern recognition, in both research and commercial applications have resulted in a large toolbox of ready to use algorithms, which each have their advantages and disadvantages. One of the important tasks in applied pattern recognition, is to choose the right method, and choose the appropriate parameters of the model to solve the problem at hand.

1.2.5 Smart sensing and wildlife

Smart sensing in agriculture is mostly focused on agricultural production and its livestock. There has not been much attention to use this technology to reduce problems with wildlife damage in the agricultural production. It is clear, that there exist a technological gap in the development of effective and cost-efficient methods to reduce human-wildlife conflicts in agriculture, and other industries. The concepts of smart sensing could be utilized to close this gap. This Ph.D. thesis contributes with theoretical frameworks and methods for using smart sensing in wildlife damage management in agriculture.

[5] Used in postal services for automatic handwriting recognition

1.3 Research Methods

The approach that has been taken in this Ph.D. project focus on how to apply sensor technologies and pattern recognition methods in the context of wildlife damage management. The research has been focused on proof-of-concept algorithm development to design theoretical frameworks for reduction of human-wildlife conflicts in agriculture.

The work has been driven by two very different conflicts between wildlife and agricultural activities, namely the problems with large flocks of birds in agricultural fields and wildlife mortality during mowing operations. This work shows the potential of utilizing commercially available sensor technology and pattern recognition algorithms in wildlife damage management. In both case studies, field experiments have been conducted to record data, and evaluate perfomance[6], in the natural environment. This has included design and development of experimental systems.

The scientific contributions of the Ph.D. project are mostly focused on how pattern recognition methods can be applied in the design of solutions for reducing human-wildlife conflicts. Methods from human speech recognition, bioacoustics, visual- and acoustic surveillance have been adapted and extended to fit the needs within the topic of the project.

1.4 Research Hypothesis and Objectives

The overall hypothesis of the Ph.D. is formulated on the basis of the limitations of existing solutions for reduction of human-wildlife conflicts, and the opportunity to investigate the effect of using intelligent sensor strategies in the interaction with wildlife:

Wildlife damage management can be performed in a more ethical, efficient and wildlife-friendly manner, if based on new sensor technology, pattern recognition and automation within tools and methods for wildlife damage management

[6] Performance of reduction of human-wildlife conflict in mowing operations have not been evaluated as these conflicts are seasonal and; therefore, it was not possible to fit the schedule of the Ph.D. project

Based on the research hypothesis, the objective of the Ph.D. project is to define a theoretical framework, based on sensor technology and pattern recognition methods, and investigate the effect of using this framework in wildlife damage management. The three main objectives are identified as:

(1) Design a framework for utilizing sensor technologies and pattern recognition within wildlife damage management

(2) Develop algorithms for sensor based detection and recognition of wildlife

(3) Investigate the effect of the proposed theoretical framework in controlled field experiments

1.5 Published and Submitted Work

This section presents the work published and submitted during this Ph.D. project. To distinguish these publications from other references in the thesis, they are prefixed with the letter "P", e.g. [P2].

1.5.1 Publications

The publications listed here are all included in this thesis in Part II and III.

[P1] Kim Arild Steen, Henrik Karstoft and Ole Green (2011). *A Multimedia Capture System for Wildlife Studies*. Paper presented at The Third International Conference on Emerging Network Intelligence, Lissabon, Portugal.

[P2] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2012). *A Vocal-Based Analytical Method for Goose Behaviour Recognition*. Sensors 12(3), pp. 3773-3788

[P3] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2013). *An Audio Based Adaptive Goose Scaring Device*. Paper presented at CIOSTA XXXV Conference, Billund, Danmark.

[P4] Kim Arild Steen, Ole Roland Therkildsen, Ole Green and Henrik Karstoft (2013). *Audio-Visual Recognition of Goose Flocking Behavior*. International Journal of Pattern Recognition and Artificial Intelligence. 27(7), pp. 21

[P5] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2014). *Audio-Based Detection and Recognition of Conflict Species in Outdoor Environments Using Pattern Recognition Methods*. Applied Engineering in Agriculture vol. 30(1), pp. 89-96

[P6] Kim Arild Steen, Andrés Villa-Henriksen, Ole Roland Therkildsen and Ole Green (2012). *Automatic Detection of Animals in Mowing Operations Using Thermal Cameras*. Sensors 12(6), pp. 7587-7597

1.5.2 Submitted work

The submitted work here are all included in Part II and III.

[P7] Kim Arild Steen, James H. McClellan, Ole Green and Henrik Karstoft. *Acoustic Source Tracking in Long Baseline Microphone Arrays*. Submitted for publication in Applied Acoustics, March 2014.

[P8] Kim Arild Steen, Rasmus Nyholm Jørgensen, Ole Green and Henrik Karstoft. *Detection and Recognition of Wildlife in Thermal Images*, Submitted to IEEE International Conference on Image Processing, January 2014

1.5.3 Other publications

The publications listed here have not been selected for inclusion in this thesis but are all available from the respective publishers.

[P9] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2012) *Video-based detection of goose behaviours*. In International Conference on Agricultural Engineering, Valencia, Spain

[P10] Kim Arild Steen, Andrés Villa-Henriksen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2012). *Automatic detection of animals using thermal imaging*. In International Conference on Agricultural Engineering, Valencia, Spain

[P11] Maiken Bjerg Møller, Kim Arild Steen, Jesper Gaarsdal and Torben Gregersen (2013). *Acoustic pattern recognition on android devices*.In Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pp. 279-284. IEEE.

[P12] Kim Arild Steen (2012). *Brug af varmefølsomt kamera ved høst*. Abstract from Plantekongres 2012 : produktion, plan og miljø, Herning, Danmark

[P13] Kim Arild Steen and Morten Stigaard Laursen (2011). *Identification and tracking of sows*. Abstract from CIGR/NJF seminar 441 – Automation and system technology in plant production , Herning, Danmark.

[P14] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2012). *Wildlife Communication : Electrical and Computer Engineering*. Aarhus University, Department of Engineering. (Technical report; Nr. ECE-TR-7)

1.6 Outline and Reading Guide

This thesis is structured in three parts, and it is written as a *collection of papers*. Part I, gives an overview of the research topic and the contributions based on a selection of the publications carried out as part of this Ph.D. project. All contributions are numbered e.g. [C1], and framed. Part II and Part III, contains a selected subset of the actual publications that is the base of my contributions.

The publications introduced in Part I all fall within the topic of *"Pattern Recognition Methods for Reduction of Human-Wildlife Conflicts"*. The work carried out in the Ph.D. has been divided into two subcategories within this topic: *Adaptive Frightening Device* and *Wildlife-friendly Farming*. The purpose of Part I is to give an overview of the publications and how they contribute to the topic while introducing relevant background material and related work. Part I introduces a total of 8 publications, where 6 have been published, and 2 have been submitted.

Part I is structured as follows: Chapter 1 contains a short introduction of the Ph.D thesis. Chapter 2 presents the publications: [P1, P2, P3, P4, P5, P7] all concerning work within an adaptive frightening device. The chapter starts with an introduction to current wildlife damage management strategies, followed by the motivation for developing an adaptive frightening device. This is followed by a description of the proposed theoretical framework, and an overview of how the contributions relate to this. In the description of these contributions, related work regarding methodologies and topic is also presented. The contributions described in this chapter is focused on acoustic and visual signal processing and pattern recognition.

Chapter 3 introduces work regarding wildlife-friendly farming, and presents the publications: [P6, P8]. This chapter starts with an introduction to wildlife mortality in mowing operations, followed by the motivation for the research within this. This is followed by a description of the proposed theoretical framework for automatic detection and recognition of animals in vegetation, and an overview of how the contributions relate to this. The contributions and related work is focused on digital image processing of thermal images.

Chapter 4 concludes the work within the Ph.D. and discusses the contributions made. The contributions are compared to similar or related work, and a critical review of the methodologies is presented. The conclusion contains a discussion of how the contributions fulfill the research hypothesis and objectives, and presents possible future work.

Part II and III lists a selection of scientific papers written by the author of this Ph.D. thesis, in collaboration with others. Each chapter presents a publication and starts by listing the bibliography entry for the publication followed by the publication in its original form.

2

Adaptive Frightening Device

This chapter presents the work and contributions regarding the development of an adaptive frightening device. Current wildlife damage management strategies are presented, followed by the motivation for developing an adaptive frightening device. A theoretical framework for implementing this is presented, and the scientific contributions to this are presented according to the identified components of the framework.

The publications: [P1, P2, P3, P4, P5, P7] are presented in this chapter. These publications include a total of seven contributions, which are framed at the end of each section concerning a specific publication. The contributions are mostly focused on acoustic and visual signal processing and pattern recognition.

2.1 Wildlife Damage Management

Wildlife damage management involves the timely use of a variety of cost-efficient control methods to reduce wildlife damages to tolerable levels. Frightening devices are an important tool used in wildlife damage management to reduce the impacts of animals [50], and the goal of using frightening devices is to prevent or reduce the damage of animals by reducing their desire to enter or stay in an area [72, 103].

Visual and acoustic stimuli are among the frequently used methods in the effort to reduce wildlife damage caused by birds such as geese, rooks, gulls, blackbird and startlings [50]. Systems include gas exploders, mylar ribbon, moving and reflective objects, firecrackers, models of predators, ultrasonic devices and distress/alarm calls [72]. The effectiveness of these devices ranges from a few days to a few weeks, at best.

In [50], a combination of stimuli is recommended to increase the effectiveness. However, also the timing of activation of frightening devices is often a critical factor, and random or animal-activated devices may reduce habituation [72, 103]. Here radar, or motion sensors can be utilized [122], however, these methods are not very cost-efficient and non-specific.

2.1.1 Bioacoustics in wildlife damage management

A type of acoustic stimuli that are promising for future frightening devices is bioacoustics [50]. Bioacoustics are animal communication signals, and this communication includes alarm or distress calls. Alarm calls are vocalizations used to warn other animals of danger. An example is the loud calling of a disturbed Canada goose [16]. The communication signals are usually species-specific [24].

Frightening devices using bioacoustic-based stimuli have been used in various research applications. In [16], the authors used bioacoustics for management of Canada geese, and found that the geese moved up to a 100 meters away from the device but never left the area. In [95] they reported a reduction of 71% in goose numbers when using bioacoustics. In [100], the author compared the use of species-specific distress calls to using suspended crow carcasses for wildlife damage management. It is concluded that the use of distress calls proved to be very effective, whereas the carcasses had no effect. In [53], the authors also concluded that treatment with tape-recorded distress calls were able to scare crows away from their roosts.

There exist a few commercial systems, which utilize bioacoustics. The *GooseBuster*[1] is specifically designed for Canada geese. The system is based on alarm, alert and distress calls which are played back from multiple speakers. The calls are altered in sequence of play, frequency, duration and interval, thus providing variability in the frightening stimuli. In [136], the effect of the system is studied in three controlled experiments. The author concludes that the use of timed alarm and distress calls alone experience habituation, however, "on-demand" playback and reinforcement (using screamers and bangers) proved to be efficient to avoid habituation.

Another, more diverse system is the *Scarecrow Premier 1500*, together with the *Ultima*[2]. This system is based on manual operations and is specifically designed for airports. The system uses a roof mounted loud speaker system, together with an arsenal of alarm and distress calls, which can be played back if the operator sees the birds. The Ultima includes a visual description of the birds of interest, which makes it easy for an operator to recognize the birds. This system is not suitable for agricultural production,

[1] http://www.bird-x.com/goosebuster-products-50.php?page_id=104
[2] http://www.scarecrow.eu/

however, it has proven efficient in airports, where cost-efficiency is surpassed by flight security.

The LRAD system[3] also utilize bioacoustics to protect airports/runways, wind turbines and agricultural activities. The system i based on a directional system, which can playback predator sounds at great distance. The activation of the system is either based on manual operation or radar technology. This makes the system too expensive in most cases of agricultural production.

Habituation to bioacoustics have been reported in [16, 50, 137]. In [137], the authors argue that this may be a result of the fact that the geese, used in the experiment, were not able to escape the enclosed study site. In [24], the authors conclude that alarm and distress calls are more resistant to habituation than other sounds, but a pest controller needs to be able to identify species, as most calls are species-specific. These observations raises the questions:

Is it possible to design an automated system to avoid habituation?

and

How can sensor technology be used in such a system?

2.2 Wildlife Communication Framework

Based on reported results and a review of present frightening devices, a number of properties, which an effective frightening device should satisfy, has been identified. The device must be able to alter the periodicity of stimuli and make it possible to utilize a combination of stimuli. When frightening stimuli is based on bioacoustics, which seems most promising, the device should be able to detect and recognize specific species. Thereby, the stimuli can be targeted towards these species most effectively. Furthermore, the device should enable reinforcement, if needed.

These characteristics have led to the theoretical framework shown in figure 2.1. The framework is based on *perception* and *action*, which is the fundamental design of an intelligent agent [113]. In this design, a model is used to interpret incoming signals, and act accordingly. The model can be a simple *if-then* (e.g. the activation of an infrared sensor leads to a specific action), or a more sophisticated model, which perceives the world in a statistical manner, and base decisions on learning algorithms, such as pattern recognition algorithms.

[3] http://www.lradx.com/

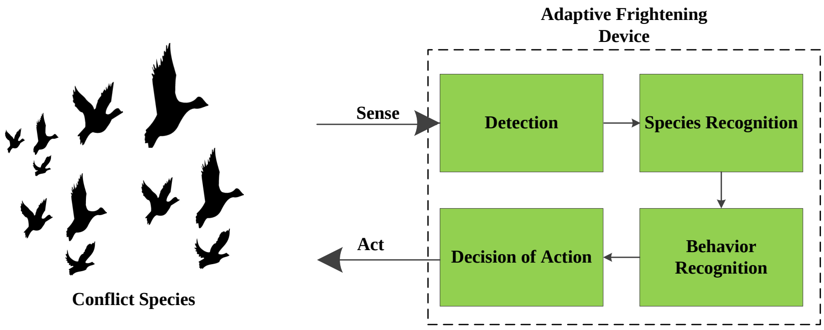


Figure 2.1: *The proposed theoretical framework of an adaptive frightening device*

The framework enables detection and recognition of species, which promotes timely use of bioacoustic, or other, stimuli. Futhermore, behavior recognition could monitor subsequent changes in behavior when frightening stimuli has been applied. This will make it possible to act accordingly, if reinforcement is needed. This framework, based on the design of an intelligent agent, features the components of an adaptive frightening device.

The identified components of the framework are: **Detection**, **Species Recognition**, **Behavior Recognition** and **Decision of Action**. The published work, presented in this chapter, contains contributions to one or more of these components. In table 2.1, an overview of the published work and their contributions to the framework, is shown. It is seen that some papers include more components and some papers contribute to the same component, but based on different methods.

A brief description of each component, together with the rationale for including the component in the framework, will be presented in the following sections. The proposed methods, and a short presentation of the results and contributions is also included.

Table 2.1: *Overview of publications and their contributions within the framework of an adaptive frightening device. The component* ***Data Collection*** *is not a part of the actual framework. However, it is critical in the design of the algoritms within the framework.*

		Component				
		Data Collection	Detection	Species Recognition	Behavior Recognition	Decision
Ref.	[P1]	x				
	[P5]		x	x		
	[P2]				x	
	[P7]				(x)[a]	
	[P4]		x		x	
	[P3]					x

[a] Behavior recognition was not implemented, however acoustic source tracking could be utilized for recognition of behavior

2.3 Reseach Contributions

This section presents contributions from the published or submitted work. The contributions are presented according to the identified components of the framework for an adaptive frightening device as presented in table 2.1. This section starts with a brief description of acoustic pattern recognition, as this is an important part of the Ph.D. thesis.

2.3.1 Acoustic pattern recognition

The main focus of the work regarding an adaptive frightening device is within acoustic pattern recognition. This section contains a brief description of the tasks within acoustic pattern recognition.

The main task in acoustic pattern recognition is to extract low-dimensional acoustic features from a high-dimensional acoustic waveform, and thereby utilize classifier algorithms to recognize a specific call-type, a species or an individual based on recorded acoustic data. These processing steps are referred to as *acoustic feature extraction* and *pattern recognition* in this thesis.

The first important step is to choose the acoustic data of interest, which is known as segmentation [45]. Both manual and automated segmentation has been utilized in research [34, 45, 129]. Manual segmentation is not suited for automatic recognition of vocalizations, however, the method can be useful when designing a classification scheme for later use. In this thesis, the task of segmentation is presented as the first component (detection) of the frightening device, which is presented in Section 2.3.3.

2.3.1.1 Acoustic feature extraction

The goal of feature extraction, with respect to classification, is to represent different vocalizations or bird calls in such a way that they are distinguishable in the given feature space.

Feature extraction for animal vocalization recognition is inspired by research within human speech and speaker recognition [25, 76]. Among the most frequently used features for speech processing are model based features such as *Linear Prediction Coding* (LPC) and *Mel-Frequency Cepstrum Coefficients* (MFCC). This has also been utilized in this thesis, where MFCC has been the method of choice. The process of MFCC feature extraction

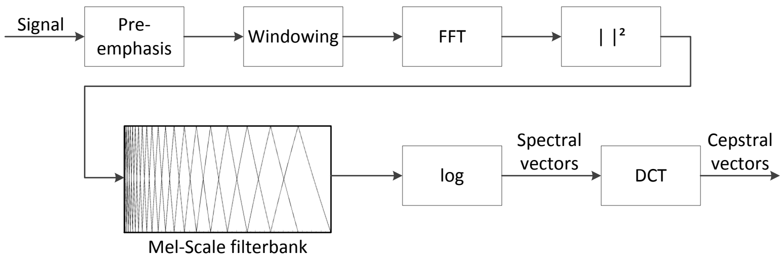


Figure 2.2: *Flow of MFCC feature extraction*

is shown in figure 2.2. In MFCC feature extraction, the frequency scale is warped according to the Mel-scale, which is a logarithmic scale designed to mimic the auditory perception of humans. The calculation of MFCC is often carried out using a Mel-scale filter bank, consisting of a number of critical band filters with center frequencies adjusted to the Mel-scale. The log-energy of each critical band is represented by spectral vectors, and a cosine transform converts the spectral vectors into cepstral vectors, according to:

$$c_n = \sum_{k=0}^{K\text{-}1} S_k \, cos\left(n\left(k\text{-}\frac{1}{2}\right)\frac{\pi}{K}\right) \quad n = 0, \ldots, K\text{-}1 \tag{2.1}$$

Here c_n is the *n*th cepstral coefficients and S_k is the spectral log-energy of the *k*th band.

Both LPC and MFCC are frequently used in research regarding animal vocalizations, however other, so called *descriptive* features have also been

Table 2.2: *Summary of related research papers within audio based recognition of animal species and behavior*

Ref.	*Animal*	*Segmentation*	*Features*	*Algorithm(s)*
[25]	Whales	-	Cepstral	HMM & GMM
[130]	Birds	STE/ZCR	MFCC & PLP[a]	DTW & GMM
[34]	Birds & Elephants	Manual	GFCC	HMM
[33]	Elephants	Manual	MFCC	HMM
[45]	Birds	Threshold in Energy domain	MFCC Descriptive	SVM
[66]	Cows	-	MFCC	HMM
[73]	Zebra Finch	Manual	MFCC	HMM & DTW
[76]	Birds	Threshold in frequency domain	aMFCC & aLPC	LDA[b]& HMM &
[81]	Farm animals	-	MFCC & LPC	ANN & HMM
[82]	Pigs	Manual	LPC	ANN
[115]	Pigs	-	LPC	SOM[c]
[129]	Antbirds	Manual	MFCC & LPC	HMM
[132]	Birds	-	Descriptive	SVM
[89]	Monkeys	Manual	MFCC	ANN

[a] Perceptual Linear Prediction
[b] Linear Discriminant Analysis
[c] Self-Organizing Map

used in recognition of animal vocalizations. Descriptive feature includes duration, signal bandwidth, short time energy etc.

These acoustic features are utilized in the subsequent pattern recognition algorithm. In this context, a number of acoustic features are often comprised into a *feature vector* which is used in the mathematical framework of a specific pattern recognition algorithm.

2.3.1.2 Pattern recognition

Recognition of animal vocalization, based on extracted features, is a pattern recognition problem. As with feature extraction, the pattern recognition algorithms used in animal vocalization research, are highly influenced by methods proven successful in human speech recognition, the most popular being the *Hidden Markov Model* (HMM) [25, 33, 73, 81, 107]. However other models such as *Support Vector Machines* (SVM), *Dynamic Time Warping* (DTW), *Artificial Neural Networks* (ANN) and *Gaussian Mixture Models* (GMM) have proven useful in classification of animal vocalizations [25, 45, 81, 130].

In table 2.2 there is a summary of research papers, which was part of the literature review of this thesis. The papers describe methods for recognition of animal vocalizations or behavior, based on auditory information. It is seen that LPC and MFCC are frequently used for feature extraction, and the HMM is by far the most popular pattern recognition algorithm.

This section includes a brief overview of three different approaches to pattern recognition of acoustic features. The algorithms HMM, SVM and GMM have all been used in animal vocalization research, and both GMM and SVM have been utilized during this Ph.D. project. I have decided to include a short description of HMM as well, as it is by far be most popular algorithm in vocalization recognition, due to its ability to capture both the stochastic and temporal variability in similar vocalizations.

Gaussian Mixture Models

The GMM is a statistical model capable of representing the probability distribution of the acoustic features for a given class. The probability distribution is modeled as a mixture of N Gaussian distributions (2.2)

$$p(\mathbf{x}) = \sum_{i=1}^{N} \pi_i \mathcal{N}(\mathbf{x} \mid \mu_i, \Sigma_i) \tag{2.2}$$

Here, the parameters μ and Σ are found from training data, which is manually labeled. The parameters π_i are the *mixing coefficients*, which must satisfy $\sum_{i=1}^{N} \pi_i = 1$. The parameters describe the statistical properties of the GMM, and much training data is needed to capture the true statistics of the acoustic features. The GMM is trained using known vocalizations as training data, and the parameters of the GMM are estimated using the *expectation maximization* (EM) algorithm. The EM-algorithm is an iterative algorithm for deriving the maximum likelihood solutions for models with latent variables [23].

In figure 2.3 an illustration of a two class problem is shown. The two classes are modeled with two different GMMs in $\mathcal{R}^2$; however, the GMM can be defined in higher dimension, based on the dimensionality of the acoustic features. The *red* class requires two Gaussians to model the statistics of the data, whereas the *blue* class only needs one. The contours show the resulting models.

Given a new observation, which may consist a number of extracted feature vectors: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$, the probability of each class can be calcu-

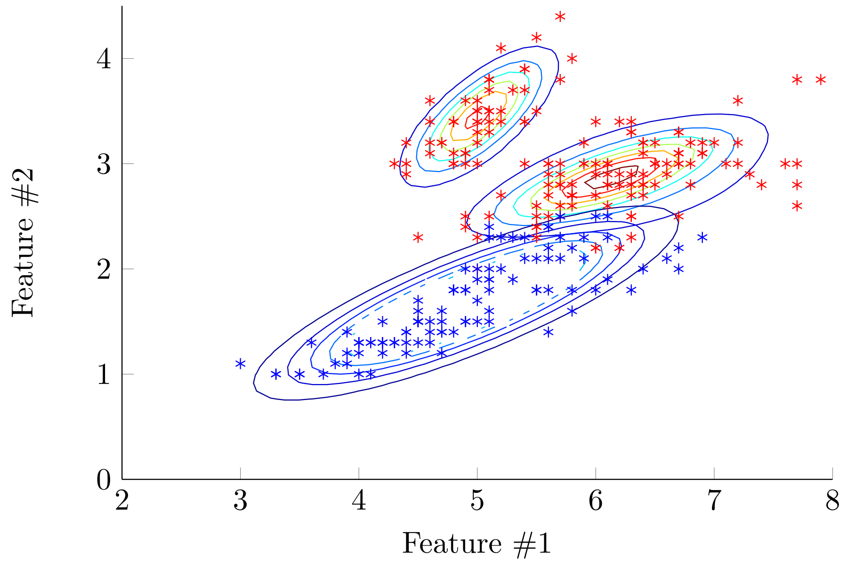


Figure 2.3: *Illustration of GMM based pattern recognition*

lated. This is usually calculated as a log-likelihood (2.3)[4] for each class. The subsequent classification is based on maximum likelihood.

$$log\ p(\mathbf{X}) = \frac{1}{M}\sum_{i=1}^{M} log\ p(\mathbf{x}_i) \tag{2.3}$$

Hidden Markov Models
The HMM is a state based model, where hidden states follow the properties of a Markov chain. The Markov model is often used to model sequential data, where the probability of an observation is influenced by prior observations. Expressed here as the first-order markov chain:

$$p(X_{n+1} = x \mid X_1 = x, X_2 = x_2, \ldots, X_n = x_n) = p(X_{n+1} \mid X_n = x_n) \tag{2.4}$$

[4] The log-transform allows for the use of the sum operator which is not as sensitive to rounding and close to zero estimates as the product operator

This makes the model excellent for modeling speech and specific animal vocalizations, where the sequential information of spectral data is important in the recognition algorithm.

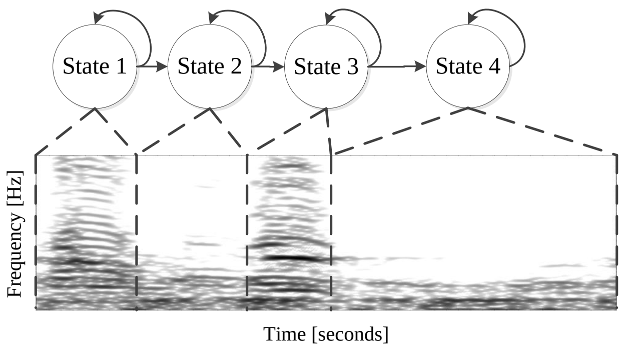


Figure 2.4: *Illustration of a HMM for a single vocalization*

In figure 2.4 an example of a HMM used for animal vocalization recognition is shown. The hidden states (1-4) follow the left to right model, often used in speech recognition [107]. They are called hidden states, as only the state outputs are observable. In vocalization recognition, these observations could be the extracted, often cepstral, features. To incorporate temporal information, the probability of observing a given feature vector is dependent on the current state, and thereby a sequence of observations is bound to a sequence of hidden states. This probability of a given observation for a given state is often modeled by a GMM [23].

The parameters of the HMM include state transition probabilities and observation probability distributions. These parameters can be estimated via the *Baum-Welch* method [107], and the training of the HMM is within the category of supervised learning, as samples of known vocalizations needs to be available. The Baum-Welch method is an iterative procedure which maximizes the probability of the model parameters given an observation of known label.

Given a new observation, which may consist of a sequence of extracted feature vectors (as with GMM), it is possible to compute the probability of the observation sequence given a specific model[5], and thereby assign a class to a new observation. An excellent tutorial in HMM for speech recognition can be found in [107].

Support Vector Machines

The SVM is a supervised learning algorithm which can be used in both linear and non-linear pattern recognition problems [26]. The models are based on structural risk minimization principle, which improves the generalization ability of the classifier [124]. Since the introduction of the model in the 1990s [134], the SVM has become a popular method of choice for many applications, including behavior recognition and speaker identification [27, 83].

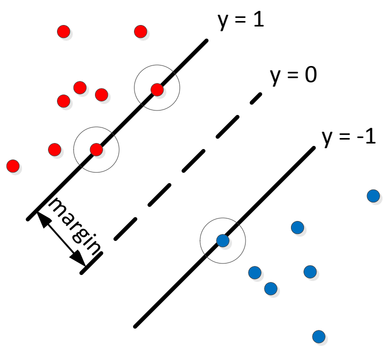


Figure 2.5: *Illustration of SVM, where the model is based on maximizing the margin. The class labels are $y = 1$ or $y =$ -1 and the decision boundary is given by $y(\mathbf{x}) = 0$*

In training the SVM, an optimization algorithm tries to find the line, plane or hyperplane (2.5) which best separates data from two classes. This is accomplished by finding the largest possible margin (see figure 2.5). Independent on the dimension of the data, a hyperplane can be expressed as

$$y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{2.5}$$

and the SVM training includes estimation of $\mathbf{w}$ and b given labeled data.

[5] An example could be a HMM for each call-type

As only the data points near the line, plane or hyperplane are used in the final model; the SVM is useful when limited training data is available as it not necessary to estimate distributions. The data points in the final model are called support vectors. In figure 2.5 this principle is shown. Here, the data points marked with circles are the support vectors, which are used in the model for classification between classes. In the figure these classes are labeled as $y = 1$ or $y =$ -1 and the decision boundary is given by $y(\mathbf{x}) = 0$.

SVMs can handle data that are not linearly separable. This is accomplished by the use of kernels (2.6), where data is transformed by $\phi(\mathbf{x})$ to a higher dimension, where non-linear data becomes linearly separable. This is not a computationally costly expansion of the SVM, as a kernel function (2.6) provides that only the dot product needs to be calculated. The SVM is one of the many pattern recognition algorithms which utilize this kernel trick, and kernel parameters are an important part of SVM model training.

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}^T)\phi(\mathbf{x}') \tag{2.6}$$

Here, $\mathbf{x}'$ is the data to be classified by the SVM.

The structure of the SVM leads to a binary classification, where the output is either one class, or another. However, it can also be used for multiclass purposes [104, 124]. A more detailed description of SVM can be found in [26].

2.3.2 Data collection

Collection of training and test data is important when developing supervised learning algorithms. There exist on-line sound libraries of thousands of species, including potential conflict species in agriculture[6,7]. However, as presented later in this chapter, both the link between vocalizations and behavior and the opportunity to use multiple sensor inputs, is of interest in this Ph.D. thesis. Therefore, a recording system was designed to record synchronized audio and video in a wildlife setting. The system is presented in [P1], where requirements such as remote access, uncompressed data and standalone power source are taken into account in the design.

The system recorded for one month, and 4-5 hours of useful data was captured during that time. This data have been used for algorithm development in [P2, P4, P5]. Furthermore, the experience gained during recording, have been utilized to design the prototype of an adaptive frightening device used in [P3].

[6] http://macaulaylibrary.org/
[7] http://www.animalsoundarchive.org/

2.3.3 Detection

As described in Section 2.3.1 the process of acoustic pattern recognition involves manual or automatic segmentation of signals of interest. In speech or vocalization recognition the segmentation task is to find voice segments in a continuous audio recording [109]. In outdoor environments, both high degrees of background noise and other sudden changes in the acoustic scene are present. This increases the complexity of automatic detection. Therefore, the question to be answered is: *Is it possible to achieve robust detection of conflict species in an outdoor environment, based on audio recordings?*

There exist limited work regarding robust detection of bird species, as most research is based on manually labeled data. In [45], the author uses a threshold defined by an estimate of the background noise energy level, in his work regarding automatic bird species recognition. However, in the context of outdoor devices, this may be difficult to estimate accurately, due to the non-stationary noisy environments with sudden noise and varying levels of this, such as gusts of wind. Acoustic arrays are utilized in [132], and this allows for higher threshold values of the energy measurements in the detection stage. The detection scheme used in the array is presented in [18]. Here, they model the background noise as a Gaussian distribution and set the threshold above the mean of the estimated distribution. In [130] a mixture of energy and zero-crossing-rate is used to separate bird calls from silent periods in the recorded data. Despite the similarity of analyzing bird vocalizations, none of the research mentioned has been working with bird flocks or the acoustic scene of agricultural fields.

An algorithm for detection of conflict species based on acoustic measurements is presented in [P5]. The algorithm is based on a probabilistic framework, where the probability of conflict species versus background is evaluated by GMMs. The GMM is used to model the density of acoustic features and has proven to be a good model in human speaker recognition, where only the spectral information is used for recognition [111], compared to speech recognition where temporal information is important. The purpose of both detection and the subsequent species recognition is not to recognize specific calls, but rather to detect the presence and recognize the species based on the measured soundscape. The GMMs are capable of modeling the multimodality of animal vocalizations. Hence, GMMs were chosen over HMM to model both conflict species and background.

The flow of the proposed algorithm is shown in figure 2.6. The detection algorithm is given by a *Multiple Hypothesis Model*, which evaluates the prob-

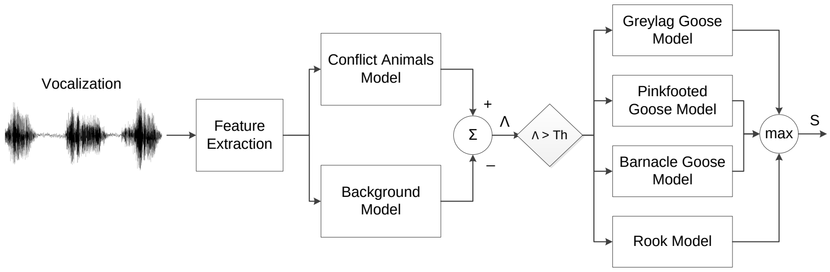


Figure 2.6: *Flow of the proposed algorithm in [P5]*

ability of conflict species versus the probability of background (no conflict species), and make a decision based on pre-defined threshold ($\Lambda > Th$). The algorithm includes both detection and species recognition, based on acoustic feature extraction and GMM evaluation. In figure 2.7, the concept of GMM based detection is visualized. In the figure, a background model and a conflict species model are shown. These are trained on the acoustic features of labeled data. Given a new observation, which consist of feature #1 and feature #2, the probability of this being a conflict species or part of the background, is calculated based on the trained models. The models shown in the figure are manually generated for visualization purposes, as the a 2-dimensional visualization of the models in [P5] is not possible[8].

As presented in Section 2.3.1.1 MFCC features are widely used in animal vocalization research, and these are also utilized in [P5]. Acoustic pattern recognition algorithms which utilize MFCC feature extraction has yielded good results across a variety of taxa including frogs, crickets, birds, cows and fish [30]. This generic feature of MFCC is attractive with respect to the proposed framework, which should be capable of managing various conflict species.

The paper includes four conflict species: Rooks, Barnacle-, Pinkfooted- and Greylag geese, which are all concatenated into one conflict species model (with multiple mixtures, as seen in figure 2.7). The performance of the detection algorithm is found via a five-fold cross-validation [23]. In a five-fold cross-validation, the extracted feature vectors are randomized, and $4/5$ are chosen for training data, and the remaining $1/5$ is test data. The division of

[8] The models are of a higher dimension, and a simplified 2-dimensional visualization would not illustrate the concept in a good manner

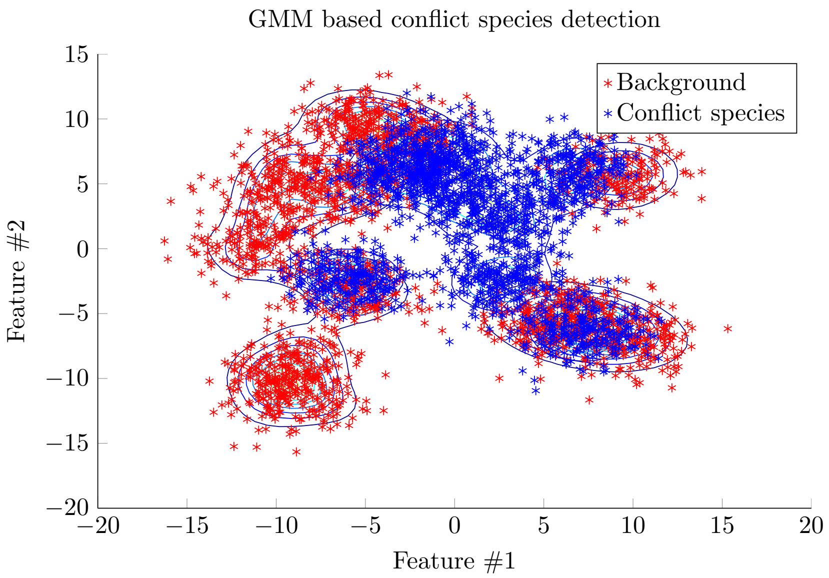


Figure 2.7: *Illustration of the concept in conflict species detection based on GMM models. The figure shows an example of a background model and a conflict species model in a 2-dimensional feature space*

data is continued five times until all data has been both training and test data. The proposed algorithm achieves a detection rate of 0.98, which is decreased as signal-to-noise ratio decreases, as seen in table 2.3. The SNR given in the table is simulated by adding additive white gaussian noise to the original data, which consist of various wildlife recordings of the conflict species.

The algorithm is based on a background model, which should model all other sounds than the conflict species. This requires more data, than used in the paper, and an evaluation of the true performance of the algorithm is therefore difficult to present. It is not possible to acquire a complete background model of all possible sounds, however a local background model[9] could be used. Here, a model update scheme, based on semi-supervised learning, could be utilized to train new models based on new measurements [41, 94]. This will be further discussed in Chapter 4.

[9] A model representing the common sounds in a specific geographic location

Table 2.3: *Performance of conflict species detection*

	Predicted	
Observed	Conflict Species	Background
Conflict species (SNR = 5 dB)	0.74	0.26
Conflict species (SNR = 10 dB)	0.81	0.19
Conflict species (SNR = 12 dB)	0.86	0.14
Conflict species (SNR = 15 dB)	0.88	0.12
Conflict species (SNR = 18 dB)	0.93	0.07
Conflict species (original data)	0.98	0.02

Contribution 1. A framework for detection of bird flocks based on acoustic measurements and statistical modeling

Video recordings have been used in surveillance applications where detection of specific events is an important part of this. Here, background subtraction or motion estimation are among the frequently used methods. In [P4], an algorithm based on audio-visual fusion is presented. The main contribution of the paper is within automatic behavior recognition, and the contribution will be presented in more detail in Section 2.3.5.2. However, the class of no activity is included in the algorithm. This could be utilized to detect activity, and thereby aid in the detection of the presence of specific species. The audio and video recording supplements each other and the algorithm achieve good performance. However, due to the limited field of view of the camera, the proposed method is, at this point, not suitable in an agricultural setting, where large areas should be monitored.

2.3.4 Species recognition

The next component in the proposed framework is species recognition. The framework is focused on the use of alarm calls as disruptive stimuli, and also base the decision of when to frighten the birds, on observed behavior (is presented in Section 2.3.5). The communication signals within a bird flock, or between individual birds, are usually species-specific [24]. Hence, the proposed framework should include the feature of species recognition if bioacoustics are used as frightening stimuli, and the question that drives the contribution included in this thesis is: *Is it possible to utilize acoustic measurement to perform robust recognition of conflict species?*

Acoustic based species recognition is comparable to the task of speaker recognition. The result is a hypothesis of who, or what, is talking or vocalizing, based on spectral or other acoustic information in the recorded audio. Different species have different acoustic characteristics, and acoustic pattern recognition can be utilized to recognize between different animals, species and even individuals within the same species.

As presented in table 2.2, different feature extraction methods and pattern recognition algorithms have been used for animal vocalization recognition, including work on species recognition.

In [129], authors use HMM to distinguish songs from five species of antbirds. Here, the various songs have very different spectral and temporal characteristics, and the HMM is able to capture both the spectral and temporal information in the songs, resulting in classification performance above 90%. In [115] LPC feature extraction is utilized based on a brief analysis of pig sound production, which is similar to human sound production. Even though songbirds vocalization include harmonics, which is comparable to the formants found in LPC, birds like geese and rooks produce a more non-tonal sound which is not captured by the LPC model.

The study performed in [45] include bird species comparable to the scope of this thesis; namely the hooded crows and greylag goose. However, only individuals, and not flocks, were investigated. The author achieves the best results with a mixture of MFCC and descriptive features. He argues that the resulting feature vector is very high-dimensional; however, classification is based on SVM, which are less sensitive to the curse of dimensionality [23, 124].

Based on the similarities between human speaker recognition and animal species recognition, and the literature review presented in table 2.2, the algorithm presented in [P5] is based on MFCC feature extraction and GMM for classification of specific conflict species. The proposed species recognition algorithm is included in the paper regarding detection of conflict species. In figure 2.6 the individual species recognition is the last step of the block diagram. When a conflict species is detected, the probability of each species is evaluated. Each species model is a GMM, trained on labeled data (presented in the paper). The species with the highest probability is the output of the algorithm.

In figure 2.8, the conflict species data from figure 2.7 are labeled according to species. The species recognition algorithm models each species as a GMM, as seen in the figure. Like GMM based detection of conflict species,

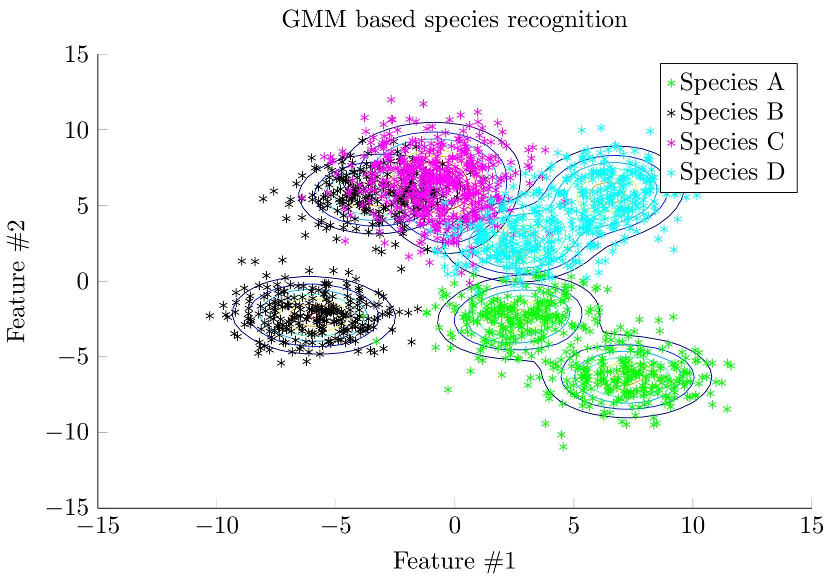


Figure 2.8: *Concept GMM based species recognition. Here the conflict species data from figure 2.6 is divided into specific species. The evaluation of probability follows the same concept as with conflict species detection*

the acoustic features represent the soundscape of the individual species, and the purpose is to classify species and not individual call types[10].

The proposed algorithm is evaluated via five-fold cross-validation and achieves classification precision between 87% and 95%. The proposed method has not been tested in real life scenarios, and it is, therefore, difficult to conclude if these results are good enough for the proposed application of an adaptive frightening device. However, it can be concluded that it is possible to utilize vocalizations to automatically recognize specific conflict species in a flock. The algorithm for species recognition was not tested in different SNR as the detection part of the algorithm. In [129] the authors conclude that classification performance is decreases as SNR decreases, which is also expected from the proposed algorithm. Simulations with various levels of SNR should have been conducted to strengthen the contribution and conclusion of [P5] with respect to species recognition.

[10] As with human speaker recognition, which should be invariant to the choice of words

More species could be included in the framework, however, with a potential cost in performance. An increase of models would increase the possibility of overlap within the feature space, which could result in more erroneous classifications.

GMMs are used to model the distribution of features within the feature space. In the paper, recognition of barnacle geese has the strongest performance, and this species is also the species of which most training data was available, which is important in modeling probability distributions. Even though, feature selection was performed to reduce the curse of dimensionality [23], more data could be used for GMM training to increase performance.

Contribution 2. GMM based framework for recognition of specific conflict species based on their vocalizations

2.3.5 Behavior recognition

The initial hypothesis of the proposed framework (see figure 2.1) is that an action from the system is capable of altering animal behavior. In a frightening scenario, the intended result is flight, based on fear. Therefore, an adaptive system need to be able to monitor change in behavior, based on the ability to recognize behavior, and react accordingly.

Methods used within animal behavior research include attached tracking devices like GPS [121] or other wireless transmitters in a wireless sensor network [99], or accelerometers, measuring the movement of specific parts of the animal body [98]. Acoustic information has also been used in chewing behavior recognition of cows [131], however, these methods also rely on attaching a device on the animals. These methods are not suitable when the purpose of the animal behavior recognition, is to utilize the results in a wildlife damage management system, as it is not possible to attach these devices on the animals. Therefore, non-invasive sensors, like microphones and cameras, are a necessity in this context.

In [132] acoustic measurements, within an array, are utilized to recognize vocalizations for source identification and localization, and thereby recognize bird behavior. However, their study is focused on individuals. Other studies utilizing acoustic measurements include recognition of dolphin behavior [125], measuring pig and chicken welfare [36, 81], and real-time stress monitoring of piglets [97]. This research show that it is possible to recognize behavior based on acoustic measurements, whether it being recognition of specific calls or the soundscape of multiple animals, which is being presented in this thesis.

A more frequently used non-invasive technique for behavior recognition is video recordings. In video recordings, digital image processing techniques and tracking algorithms can be utilized to detect and recognize specific movements, which are linked to certain behaviors. Compared to acoustic measurements, the range of visual information may be lower. However, the link between visual information, like movement or posture, and behavior is more straightforward. A popular application in automated video based behavior recognition is laboratory experiments, where changes in mouse or fish behavior is important for medical research or behavioral research [84, 138]. More domain related applications include monitoring of livestock behavior, including pigs [17, 127], chickens [36] and cows [80]. These applications are either focused on controlled experiments or indoor applications, which is not the case with wildlife in an agricultural setting.

During the Ph.D. project various methods for automatic behavior recognition, have been proposed based on both audio and video recordings. The most recent approach, based on using an array of microphones, has not been used for behavior recognition within the Ph.D. project. However, results from this work may be utilized in future work.

In the following three subsection, three different strategies for automated behavior recognition are presented.

2.3.5.1 Single microphone setup

A certain behavior is a mixture of responses to internal and external stimuli, and a full description of behavior would include internal as well as external responses [127]. Therefore, the task of measuring behavior can be a difficult one. However, acoustic information may provide useful information about animal behavior, as animals may use different vocalizations based on their state of mind. This is especially true for birds within a flock, as they utilize acoustic communication within the flock to express this. Thereby recognition of interspecies communication could be used for automatic behavior recognition. This was the hypothesis in [P2], where audio recordings of wild geese during different behaviors are utilized to train a supervised learning algorithm, and thereby develop an algorithm for audio based behavior recognition.

The data, utilized in the paper, is acquired by the system described in [P1]. Here two occurrences of three behaviors: landing, foraging and flushing, were recorded. The amount of data is limited; hence the SVM has been chosen for classification of behavior.

In the proposed algorithm, the SVM is used in a multiclass setup using a directed acyclic graph [45, 104], and is trained in a one-versus-one manner. To avoid over-fitting in the training, a *soft margin* SVM is utilized [20, 26]. The soft margin SVM allows for misclassifications during training, which makes it possible to adjust the generalization properties of the model. As the extracted acoustic features were not linearly separable, a radial basis function kernel has been utilized as kernel in the algorithm. The weights and bias of the SVM are found through optimization. However, the soft margin parameter and the kernel parameter are fixed values, which must be defined before training. To choose the best possible parameters for the algorithm, grid search, which is a standardized method for SVM kernel parameter selection, was utilized [29, 60].

In [P5] MFCC features were used for detection and species recognition. These features have been shown to be useful in both human speech recognition [42, 139] and animal vocalization recognition. However, animals do not perceive sounds equally as humans, which means that MFCC may not be useful for audio based behavior recognition, as this is based on the inter-species communication. In [34] generalized perceptual features are introduced. This feature extraction method is based on the Greenwood function [57], which assumes that sound perception is on a logarithmic scale (like the Mel-scale) but that this scale differs for different species. Greenwood found this to hold true for mammals, however, in [15] the *Greenwood Function Cepstral Coefficients* (GFCC) are used for recognition of Ortolan Bunting songs with good results. In [P2] GFCC features are utilizes for the audio based behavior recognition. The paper includes data from two occurrences of the specific

Table 2.4: *Model performance for each behaviour classification*

	Performance		
Behaviour	Accuracy[a]	Precision [b]	Sensitivity[c]
Flushing	0.93	0.66	0.79
Landing	0.90	0.79	0.91
Foraging	0.91	0.98	0.86

[a] Ratio of correct predictions (both positive and negative) that were correct
[b] Ratio between correct postive and incorrect postive predictions
[c] Ratio of correct classifications versus total number of predictions

behaviors. Performance measures are found via five-fold cross-validation, where data from these occurrences have been mixed and randomized. The GFCC and SVM based algorithm achieves the performance measures shown in table 2.4.

Landing behavior, which is an important behavior to detect[11], achieves fair performance with precision of 0.79. However, some overlap between flushing and landing behavior is present in the feature space (as seen in the paper). Both training and test data were manually chosen in order to isolate the hypothesis of audio based behavior recognition, and not focus on detection of specific vocalizations. This means that the SNR is high for the data utilized in the paper, and it is expected that decreased SNR would result in more overlap between classes. This is the motivation for further work within automatic behavior recognition, which includes audio-visual behavior recognition and array based tracking.

Contribution 3. A novel method for recognition of flocking behavior based on acoustic pattern recognition

2.3.5.2 Audio-visual recognition

The main contribution of sensor fusion is based on the idea that multiple, orthogonal, sources of information achieve better performance compared to using one source alone. An excellent example of this is within audio-visual speech recognition, where a combination of visual and acoustic features provide better results in the case of low SNR [106].

In animal flocks, both the movement and the vocalizations, i.e. the communication within the flock, is often associated with certain behaviors. The hypothesis of [P4] is that fusion of audio and video are suitable for robust multi-modal recognition of animal flocking behavior, and thereby provide improved performance compared to using audio alone.

There are different strategies for fusing audio and video information. In human audio-visual speech recognition research, feature fusion and classifier fusion have been used to fuse the information from the two sources [106]. The most common method used in speech recognition is to perform feature fusion in a multi-stream HMM. However, SVM models for audio based recognition was developed in [P2], and, therefore, multi-stream HMM was not

[11] The birds are more alert while landing and a disruptive stimuli will have a better chance of success

found suitable for this application. Furthermore, in contrast to feature fusion methods, the classifier fusion framework provides a mechanism for capturing the reliability of each modality, and thereby design the algorithm for robust recognition based on knowledge of the individual classifier performance [70]. In [P4], a classifier fusion strategy is implemented, where classifiers for each individual source were designed.

Vision based recognition of bird flock behavior is similar to the task of crowd behavior recognition. Popular approaches in crowd motion estimation are background subtraction, temporal differencing and optical flow [37, 61, 135]. Optical flow estimation is utilized in the paper regarding audio-visual recognition. It is an approximation of the motion in an image sequence and has been used in animal behavior recognition, crowd motion simulations and event detection [37, 40, 86].

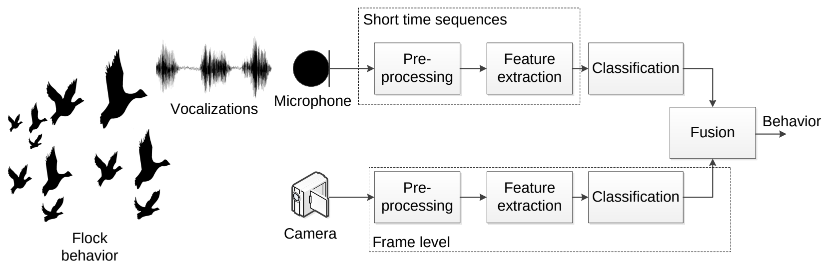


Figure 2.9: *Flow of the proposed framework for fusion of audio and video information for animal behavior recognition. The framework is based on classifier fusion.*

It has not been possible to find any work regarding automatic recognition of animal behavior based on audio-visual information. Most research regarding the link between visual and acoustic information for animal behavior recognition utilize either manual observations or manual inspections of video recordings. In [96] video recordings of chickens were used for manual detection of group behavioral pattern in an experiment to link their vocalizations with the thermal environment. Likewise, [125] uses video recordings for linking dolphin sound to their location and behavior, and in [97] manual observations were used to link vocalizations to the stress level of piglets.

In figure 2.9 the proposed framework, for classifier fusion based audio-visual recognition, is shown. The audio based recognition is based on SVM models as in [P2], however, with a slight change in output values. In the paper regarding audio based recognition, the output from the SVM models, were *hard* outputs, meaning one class or another. In the proposed classifier fusion framework, this has been modified to *soft* outputs. The soft output is a measure of probability of each behavior. This property is also the case for the video based classifier and is utilized in the final step of classifier fusion.

In the video based classifier, the probability of each behavior is estimated based on optical flow estimation for each frame. As behavior is a sequential action, these probability estimates are updates based on Bayes' rule (2.7)

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{2.7}$$

where $P(B \mid A)$ and $P(A)$ is the likelihood and prior, respectively. The denominator $P(B)$, sometimes called evidence [23], ensures that the posterior probability $P(A \mid B)$ is a valid probability measure. In the presented algorithm, B denotes data, which is given by optical flow estimates and A denotes behavior. At each frame, the posterior probability of a behavior given the data, is updated based on prior information of this behavior and the current probability of data given a specific behavior. Thereby, the output of the video based classifier is a vector containing probability estimates for each behavior (like the audio based classifier).

The classifier fusion, given soft outputs from the individual classifiers, have been implemented in the following manner: Given R classifiers, the pattern Z can be assigned to m possible classes $\{\omega_1, \ldots, \omega_m\}$ by

$$\begin{aligned} &assign \quad Z \to \omega_j \quad if \\ P(\omega_j \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) &= \max_k P(\omega_k \mid \mathbf{x}_1, \ldots, \mathbf{x}_R) \end{aligned} \tag{2.8}$$

where $P(\omega_j \mid \mathbf{x}_1, \ldots, \mathbf{x}_R)$ denotes the probability of class ω_j given the vector containing soft outputs, $\mathbf{x}$, from R different classifiers. In the paper, the sum- product- and mean rule for calculating class probabilities are evaluated.

In table 2.5 the performance of the different fusion strategies is compared to using audio and video alone. It is seen that the overall performance is improved using all fusion strategies.

The fusion slightly degrades the performance of the video-based classifier with respect to sensitivity, which is a result of the outputs from the audio based classifier (as presented in the paper). However, even though the sensitivity is degraded both the accuracy and specificity has been improved.

Table 2.5: *Comparison of performance using audio (A), video (V) or classifier fusion, sum (S), product (P) and mean (M), (mean ± s.d.). The four classes are No Geese (NG), Flushing (FL), Landing (L) and Foraging (FO)*

		Performance		
Behavior	**C**	Accuracy[a]	Specificity[b]	Sensitivity[c]
	A	0.97 ± 0.03	0.97 ± 0.03	0.95 ± 0.05
	V	0.96 ± 0.03	0.99 ± 0.004	0.88 ± 0.001
NG	S	0.99 ± 0.002	0.99 ± 0.003	1 ± 0
	P	0.99 ± 0.004	0.99 ± 0.005	1 ± 0
	M	0.99 ± 0.002	0.99 ± 0.003	1 ± 0
	A	0.91 ± 0.05	0.98 ± 0.02	0.71 ± 0.22
	V	0.88 ± 0.08	0.84 ± 0.11	1 ± 0
FL	S	0.98 ± 0.04	1 ± 0	0.92 ± 0.17
	P	0.98 ± 0.04	1 ± 0	0.9 ± 0.17
	M	0.98 ± 0.04	1 ± 0	0.92 ± 0.17
	A	0.88 ± 0.07	0.88 ± 0.1	0.88 ± 0.07
	V	0.89 ± 0.08	0.97 ± 0.06	0.65 ± 0.33
L	S	0.97 ± 0.04	0.97 ± 0.06	0.99 ± 0.007
	P	0.97 ± 0.04	0.97 ± 0.06	0.99 ± 0.01
	M	0.97 ± 0.04	0.97 ± 0.06	0.99 ± 0.007
	A	0.96 ± 0.03	0.97 ± 0.03	0.92 ± 0.08
	V	0.97 ± 0.04	0.99 ± 0.003	0.91 ± 0.17
FO	S	1 ± 0	1 ± 0	1 ± 0
	P	0.99 ± 0.01	0.99 ± 0.01	1 ± 0
	M	1 ± 0	1 ± 0	1 ± 0

[a] Ratio of correct predictions (both positive and negative) that were correct
[b] Ratio of correct negative predictions (the ability to reject)
[c] Ratio of correct positive predictions

The improvement of landing behavior recognition is an important result in the paper, since robust recognition of landing behavior is an important part of an adaptive frightening device. Immediate detection of landing behavior is crucial to scare off bird flocks while they are alert. However, the video-based recognition performs worse than the audio based in the case of landing behavior. This is due to the nature of landing behavior. Not all geese land at the same time and some geese might take off again to find a better location.

This affects the robustness of the optical flow estimates, which are utilized in the video based classifier.

Furthermore, the limited field of view of cameras, and the constant need for maintenance, makes a video based method impractical in real life scenarios. The advantage of using visual information in behavior recognition is the direct link between movement and behavior. However, this property is not unique for camera based systems, as localization through sensor arrays may provide similar results. This is the motivation the work within microphone array based localization and tracking of acoustic sources, which is presented in the next section.

Contribution 4. A theoretical framework for fusion of audio and video in animal behavior recognition

2.3.5.3 Multiple microphone setup

Microphone arrays have been utilized by [132] to localize and track individual birds, in order to investigate behavior. Here microphone arrays were chosen as it was not possible to use camera technology, as the application of interest was antbirds in the jungle[12]. They use an array of arrays, meaning that a single sensor node is comprised of four microphones. This results in robust direction of arrival (DOA) estimates which are used for localization of the birds. The array system is very sophisticated, and several papers have been published on the subject including descriptions of the hardware architecture [52], self-localization algorithms and bird localization algorithm [18], and individual bird recognition based on vocalizations [132]. However, their application is not comparable to the application of an adaptive frightening device, as a frightening device should monitor and frighten bird flocks in large regions of interest. Furthermore, there is a desire to develop a cost-efficient method for wildlife damage management.

Sensor array based localization is mostly based on three types of physical variables: time difference of arrival (TDOA) [79, 120], DOA [68] and received signal strength, e.g. energy based methods [62, 87]. Long baseline microphone arrays, where the distance between single microphones is large, is a cost-efficient method to monitor activity in a large region of interest. In long baseline arrays, the geometry of the array affects both time and energy measurements, due to long travel times and a high degree of attenuation. The

[12] Dense vegetation led minimum visibility

high range between source and sensors leads to low SNR that affects both energy- and phase-based methods (TDOA and DOA). The large separation between sensors also affects the phase-based methods, as turbulence will de-correlate the signals. Furthermore, TDOA and DOA requires very accurate synchronization between microphone pairs. This requires high sampling rates, complicated infrastructure and more computation at each sensor. Energy measurements reduce the need for high sampling rates and data transfers, compared to the TDOA and DOA. Moreover, the cost of computing energy levels is very low. This makes energy based localization (EBL) a cost-efficient method for acoustic surveillance of a large area.

In [118], EBL methods are used for localization and tracking of an amphibious assault vehicles and dragon wagons along a road. They used approximately seven to nine sensors in a region of $100 \times 150 m^2$. In [43] microphone arrays are used for vehicle tracking, including classification of the vehicles based on acoustic features. These applications are not within animal behavior recognition; however, there is a common objective, which is to monitor movement from specific objects within a large region.

Acoustic source localization

In [P7] a tracking algorithm based on energy measurements within a long baseline microphone array is presented. Energy based localization is based on the energy decay model (2.9), where the energy at the *i*th sensor, within a microphone array, is calculated as follows:

$$y_i(t) = \frac{s(t\text{-}\tau_i)}{\|\mathbf{r}_s\text{-}\mathbf{r}_i\|^{\alpha}} + \varepsilon_i(t) \quad i = 1, 2, \ldots, N \tag{2.9}$$

Here $s(t\text{-}\tau_i)$ denotes the source energy over a time interval, given by t. The time required for the signal to propagate from the source to the *i*th sensor is given by τ_i. The vectors $\mathbf{r}_s$ and $\mathbf{r}_i$ denote the unknown source position and the known sensor position, respectively. The noise term ε_i is white Gaussian measurement noise with variance $\sigma^2_{\varepsilon_i}$. The decay factor α may vary due to the environment, however $\alpha = 2$ is a good approximation [62].

This model has two unknowns, namely the source energy $s(t\text{-}\tau)$ and its position $\mathbf{r}_s$. By calculating the energy ratio (2.10) between two microphones (the i*th* and the j*th*), the source energy, which is not of interest in the scope of localization, is removed from the problem. By approximating the additive noise term ε by its mean value μ, the energy ratio K can be calculated:

$$K_{ij} = \left(\frac{y_i\text{-}\mu_i}{y_j\text{-}\mu_j}\right)^{-\frac{1}{\alpha}} = \frac{\|\mathbf{r}_s\text{-}\mathbf{r}_i\|}{\|\mathbf{r}_s\text{-}\mathbf{r}_j\|} \tag{2.10}$$

All possible source locations $\mathbf{r}_s$ that satisfy the above equation is located on a D-dimensional hyper-sphere given by $|\mathbf{r}_s\text{-}\mathbf{c}_{ij}|^2 = \rho_{ij}^2$ where the center $\mathbf{c}_{ij}$ and radius ρ_{ij} associated with the sensor pair *i* and *j* are given by:

$$\mathbf{c}_{ij} = \frac{\mathbf{r}_i\text{-}K_{ij}^2\mathbf{r}_j}{1\text{-}K_{ij}^2} \quad \text{and} \quad \rho_{ij} = \frac{K_{ij}\,\|\mathbf{r}_i\text{-}\mathbf{r}_j\|}{1\text{-}K_{ij}^2} \tag{2.11}$$

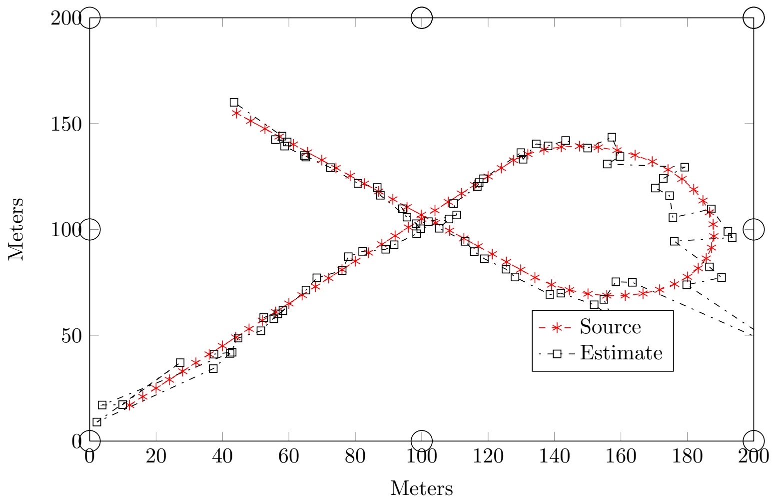


Figure 2.10: *Example of LS based localization in noisy measurements with $SNR = 7 \pm 1$ dB. It is seen that LS estimates are erroneous. The notation ◯ indicate a microphone position*

Based on these hyper-spheres, an unconstrained least-squares (LS) solution can be formulated by considering different pairs of hyper-spheres [87]. In figure 2.10 the LS solution is utilized for localization of a moving acoustic source within a long baseline array. In the simulation, background noise was added to the measurements, and it is seen that the estimate is erroneous in many cases. This makes localization alone unsuitable for tracking moving acoustic sources as background noise will occur.

Acoustic source tracking

The tracking algorithm presented in [P7] is based on the cost reference particle filter (CRPF). The CRPF is a new class of particle filters which is able to estimate the system state from the available observations without a priori knowledge of any probability density function [91]. Hence, it is suitable for real life application, where a statistical model of the system is not always known. In the paper, the system state consists of the position $\mathbf{r}$ of the acoustic source and its velocity in the x- and y-directions, denoted by $\dot{\mathbf{r}}$, resulting a four dimensional state space $\mathbf{x} = [\mathbf{r} \quad \dot{\mathbf{r}}]^T$.

In CRPF, a user-defined cost function measures the quality of the state estimates according to the available observations. As this approach is not based on probabilistic assumptions, the CRPFs yield local representations of the cost function specifically built to facilitate the computation of minimum cost estimates of the state signal [91]. The framework of the CRPF features an incremental cost function $\Delta\mathcal{C}(\mathbf{x} \mid \mathbf{y})$, which assigns a cost on each particle after measurements $(\mathbf{y})$ are obtained. It is a user-defined cost function and can be designed based on the desired application.

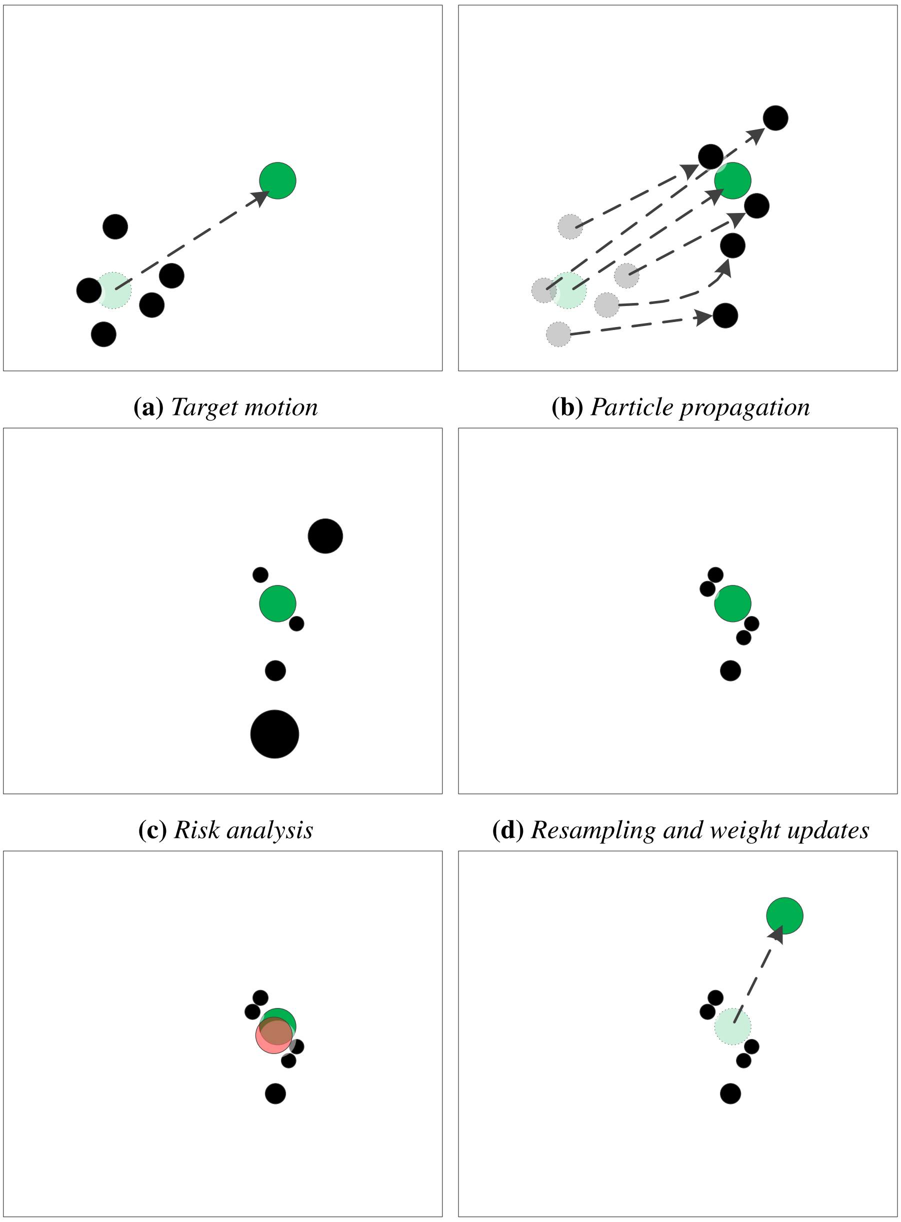
A risk function, $\mathcal{R}(\cdot)$, is used to select the most promising particle trajectories based on measurements. The risk of each particle motion is based on the incremental cost as such

$$\mathcal{R}(\mathbf{x}_t \mid \mathbf{y}_{t+1}) = \Delta C(f_x(\mathbf{x}_t) \mid \mathbf{y}_{t+1}) \tag{2.12}$$

Here $f_x(\mathbf{x}_t)$ is the dynamic model for $\mathbf{x}$. Each particle stores its cost, which is updated by (2.13). Here, a memory term λ can be used to preserve the cost information from previous iterations

$$\mathcal{C}_{t+1} = \lambda\mathcal{C}_t + \Delta\mathcal{C}_{t+1}(\mathbf{x}_{t+1} \mid \mathbf{y}_{t+1}) \tag{2.13}$$

In figure 2.11 the flow of CRPF based tracking is shown. For illustration purposes only five particles (the black dots) are used in the example. The particles are initialized in a bounded interval [91] and all particles have the same weight. During tracking of the target (the green dot), the particles are propagated (figure 2.11b) and the most promising trajectories are chosen (figure 2.11d). This is based on evaluation of the risk function (figure 2.11c), which is illustrated as different sizes of the particles. Here a larger size indicates a



(a) *Target motion* (b) *Particle propagation*

(c) *Risk analysis* (d) *Resampling and weight updates*

(e) *Estimate* (f) *Next iteration*

Figure 2.11: *Illustration of the flow in CRPF based tracking*

larger risk. In figure 2.11d it seen that the larger particles are removed, and the smaller particles are copied. This is the process of resampling in the particle filter. The final state estimate (the red dot) is based on the resampled particles, and calculated as such:

$$\mathbf{x}^{mean} = \sum_{i=1}^{Q} \pi^{(i)} \mathbf{x}^{(i)} \tag{2.14}$$

where Q is the number of particles in the filter and $\pi^{(i)} \propto \left(\mathcal{C}^{(i)}\right)^{-1}$ is a probability mass function, which maps the cost of a particle to an importance weight (as in standard particle filters).

Unlike the standard particle filter, the particles store their weight for the next iteration, as seen in figure 2.11f.

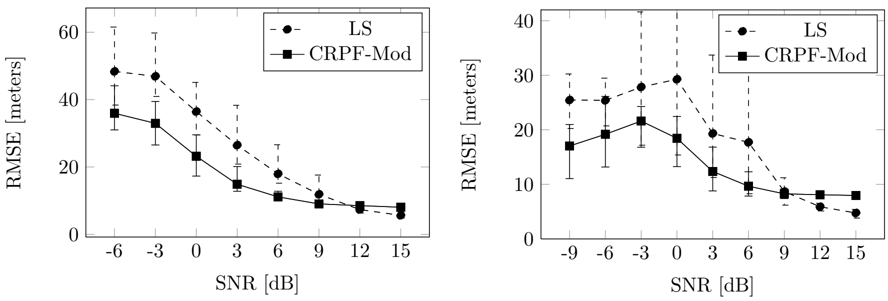
Based on observations of localization errors during simulated noise and wind gusts, a user-defined cost function (2.15) is proposed, which can be used within the CRPF framework for acoustic source tracking. The new cost function is a modification to the non-linear least-squares formulation presented in [62]. The main contribution of the modified cost function is that it is designed to promote particle estimates that lie on or very close to the hyper-spheres, estimated from calculated energy ratios.

$$\hat{J}(\mathbf{r}) = \sum_{m=1}^{M} |\|\mathbf{r}\text{-}\mathbf{c}_m\|\text{-}\rho_m|^p \quad 0 < p \leq 1 \tag{2.15}$$

Here the exponent p is a design parameter, which can be used for choosing to which degree the particles have to fit the hyper-spheres. The variable $\mathbf{r}$ denotes the source position estimate, and m denotes a sensor pair.

In figure 2.12 tracking using LS estimation is compared to CRPF based tracking using the modified cost function (CRPF-Mod). In figure 2.12a different levels of background noise are simulated, and it is seen that CRPF-Mod outperforms LS based tracking when SNR is low. The same result is seen in figure 2.12b, where sudden wind gusts are simulated. When SNR is increases, the advantage of CRPF based tracking is decreased. Here, the LS based method is able to achieve fair estimates, and the modeling noise in the CRPF framework contributes to more erroneous estimates.

Like the standard particle filter, the CRPF is able to estimate non-linear and non-Gaussian states. However, compared to LS estimation it is less computational efficient, due to the number of particle updates in each iteration of the filter. Computational cost is an important design parameter in sensor



(a) *Background noise*

(b) *Wind gusts*

Figure 2.12: *Simulation results of tracking in different levels of background noise and wind gusts. Tracking error is measured by root mean squared error (RMSE), and the plots compare LS estimation versus the CRPF-Mod algorithm*

arrays due to energy consumption. Nonetheless, the increased availability of graphical processing units (GPUs) in mobile devices [31] and research within real time particle filters [59] makes it possible to utilize these algorithms in sensor arrays. Other, more computationally efficient tracking filters are available, including the *Interacting Multiple Model* Kalman filter (IMM-KF) [85]. Unlike, the standard Kalman filter or the Extended Kalman filter, the IMM-KF incorporates multiple dynamic models, and fuse these models to reach a single state estimate. This method has proven very efficient in radar tracking [69, 101], where rapid changes between linear and non-linear movement often happen. The IMM-KF was not utilized in the contribution as it does not allow user-defined cost functions in the same manner as the CRPF framework. However, further work could include research within IMM-KF based tracking.

Contribution 5. Theoretical framework for acoustic source tracking in long baseline microphone arrays

Contribution 6. A modification to cost function for energy based localization, which increases tracking performance in noisy conditions

2.3.6 Decision of action

The proposed component, named **Decision of Action**, decides, based on the previous processing and estimates, whether an action should take place or not. Based on the promising results of using bioacoustics in wildlife damage management systems, the initial action should be a playback of distress or alarm calls when conflict species are detected and recognized. Based on the measured response to the chosen stimuli, the system should choose whether or not to use the same method again immediately or wait, or maybe use another stimuli.

In [P3], the results of using methods from the framework are presented. The main hypothesis of the experiment was that an acoustic based adaptive frightening device would outperform existing methods used for managing wildlife barnacle geese. This includes gas exploders and visual stimuli, which is subject to habituation after a few days or at best a few weeks. The experiment involved one system, and the achieved results are, therefore, of a preliminary nature, as more experiments needs to take place, in order to achieve a statistically sound result. The prototype system, used for the ex-

(a) *The hardware inside the device* **(b)** *The device (right) and a recording setup (left)*

Figure 2.13: *Prototype of an adaptive frightening device. Photo by Kim Arild Steen*

periment, included a computer, a microphone, an amplifier and two speakers (see figure 2.13). The microphone continuously recorded, while the computer performed acoustic pattern recognition to decide whether barnacle geese were landing, foraging or flushing. The system did not include species recognition, as the barnacle geese were the only conflict species within the region at the time of the experiments. The system was set to take action if landing or foraging behavior was observed. Unfortunately, the algorithm for robust conflict species detection were not available at the time of the experiment, which resulted in multiple false detections as wind gusts and other background noise triggered the system. The classifier, implemented for the experiment, had a class of *no geese* included, which was designed to decide whether geese were present or not. However, the trained model proved unfit for the task. Despite this, the system did a very good job at keeping the geese away from the test area.

The main result is shown in figure 2.14, where barnacle goose activity have been measured for three periods of time: an active period, an inactive period and again an active period. The goose activity was measured based on counting goose droppings along three transects[13], starting from the system. First, the system was turned on for three weeks (active period #1). Then it was turned off for almost another three weeks (inactive) to check if the barnacle geese were not interested in the specific area chosen for the experiment. As seen in the figure, this was not the case as goose activity increased in that time. When the system was turned on again for almost three weeks (active period #2), the activity dropped again.

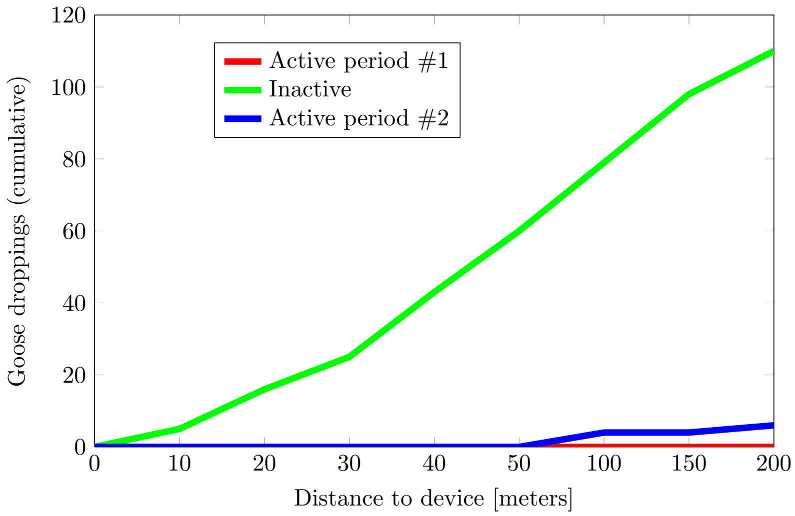


Figure 2.14: *Results from preliminary experiments with an adaptive frightening device*

[13] A transect is a path along which one counts and records occurrences of the phenomena of study. In this case goose droppings.

Obviously, the conclusions, which can be made on the basis of this preliminary field test, are limited by the lack of adequate statistical power. Therefore, the results should be interpreted with caution. However, the system was successful at reducing the presence of barnacle geese for a total period of almost six weeks and thereby preventing damages to the pasture at a radius of up to 200 meters from the system. This is equivalent to an area of 12.6 ha.

Contribution 7. An evaluation of using acoustic behavior recognition to perform wildlife damage mangement of goose flocks

In the spring of 2014 field tests including five systems and more frequently counting of goose dropping, are being performed. These tests involve a commercially available system called *AniMan*® (see figure 2.15), which implements the concepts presented in this chapter. The system is being developed by a small company called Wildlife Communication Technologies[14], which is a spin-off company based on the research conducted during this Ph.D. project.

[14] http://www.wildlifecommunication.dk/

Figure 2.15: $AniMan^{®}$

3

Wildlife-Friendly Farming

This chapter presents the work and contributions regarding wildlife-friendly farming. The motivation and current strategies for wildlife-friendly farming are presented, followed by a presentation of state of art work within automated solutions for detection of wildlife during mowing operations. A theoretical framework for implementing sensor based solutions for wildlife-friendly farming is presented, and the contributions of this thesis are related to the identified components of the framework.

The publications: [P6, P8] are presented in this chapter. These publications include a total of three contributions, which are framed at the end of each section concerning a specific publication. The contributions are mostly focused on digital image processing of thermal images.

3.1 Wildlife Mortality in Mowing Operations

As presented in the introduction, the increased need for high-efficiency agricultural production has resulted in high wildlife mortality rates during mowing operations. Mowing operations take place during the summer, where roe deer fawns and leverets are immobile, vulnerable and hiding in the grass. When a large and noisy agricultural machine is approaching these animals, their natural instinct is not to run, but rather to remain motionless on the ground. This makes it very difficult to see and react to the presence of the animals, and they are often overlooked by farmers. As a result of the increase in both working speed and width, adults of otherwise mobile species, e.g., fox and roe deer, are also at risk of being killed or injured in farming operations as they may be unable to escape the approaching machinery.

Various methods and approaches have been used to reduce wildlife mortality. Delayed mowing date, altered mowing patterns (e.g., mowing from the center outwards [55]) or strategy (e.g., leaving edge strips), longer mowing intervals, reduction of speed or higher cutting height [55] have been suggested to reduce wildlife mortality rates. Likewise, searches with trained dogs prior to mowing may enable the farmer to remove e.g., leverets and fawns to safety, whereas areas with bird nests can be marked and avoided. Alternatively, various scaring devices such as flushing bars [55] or plastic sacks set out on poles before mowing [67] have been reported to reduce wildlife mortality. Altered mowing patterns might work for mobile animals; however, fawns are immobile, and they are not able to run towards safety even if they got the time to do it. Reduction of speed, the use of trained dogs and actions before mowing all results in lower efficiency. Therefore, the development of automated systems, capable of detecting animals in the vegetation could have a positive impact on both agricultural production and wildlife mortality. This chapter present related work regarding the development of an automated system, and the contributions made during this Ph.D. project.

3.2 Automatic Detection of Wildlife

The idea of automatic detection of wildlife in grasslands is not novel, and various attempts have been made to develop such a system. In [58] a manual operated portable system is presented. The system is based on infrared technology and works very well under defined weather conditions (early morning and cloudy days), and it has been applied the patented *Infrared Wild Savior* system since 1999. The disadvantage of the system is its low efficiency, as the maximum search power is around 3 ha[1]/h, when the weather conditions are fit.

In the *WILDRETTER* project[2], principles from [58] were further developed and tested. Here the initial idea was to mount sensors to a mechanical arm next to the mower, which would make it possible to analyze the part of the field that were to be mowed next. The arm could be equipped with multiple sensors including, LDS[3], Infrared Thermal camera, radar and spectral cameras [63].

In [44] a multistatic radar array for detecting wildlife is presented. The method is not sensitive to weather conditions and works at high speeds. The solution is sensitive to the orientation of the wildlife, however, a solution to this is presented in the paper. In [63] the authors do not present further result using the multistatic radar system, but put more focus on thermal imaging systems. However, they conclude that vision systems are not a viable solution when the cameras are mounted on the arm, as image quality is highly affected

[1] ha = $100 \times 100 m^2$
[2] http://www.wildretter.de
[3] Laser Distance Sensor

by the speed and vibrations of the machine. Instead a UAV-based system is utilized [64]. Using this solution, the movement of the tractor does not affect the image quality, and it is possible to manually scan large areas. The authors show that thermal imaging can be used to detect roe deer fawns based on aerial footage. However, the detection is performed manually, and should be automated to increase efficiency. They conclude that the thermal imaging strategy is sensitive to detection of false positives, meaning that objects that are heated by the sun are falsely labeled (manually) as a roe deer fawns. The authors suggest further research within sensor fusion to reduce the number of false detections.

Based on this review, a theoretical framework for automatic detection of wildlife is proposed (see figure 3.1). The framework is based on initial detection of objects in the field, followed by a subsequent recognition of the object (animal or not) to reduce false positives. The final component is the decision of which action to take. As described earlier, once an animal is detected, it may not be able to escape by itself. In [64] trained farmers or hunters move the fawns manually. This is inefficient and may also impose problems to the mother/fawn relationship (human odor etc.). A brief discussion of this is found in Section 3.3.3.

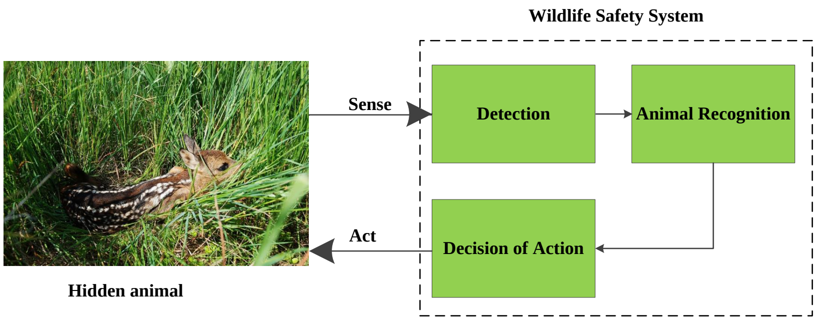


Figure 3.1: *The proposed theoretical framework of a sensor based wildlife safety system. Photo by Gilles San Martin*

In table 3.1, an overview of the published, or submitted, work and their contributions to the framework, is shown. A brief description of each component, together with the rationale for including the component in the framework, will be presented in Section 3.3. The proposed methods, and a short presentation of the results and contributions is also included. The component regarding decision of action has not been investigated during this Ph.D. project; however, a discussion on the subject is included in this chapter.

Table 3.1: *Overview of publications and their contributions within the framework of a wildlife safety system*

		Component		
		Detection	Animal Recognition	Decision of Action[a]
Ref.	[P6]	x		
	[P8]	x	x	

[a] The final component of the proposed framework has not been investigated during this PhD project

3.2.1 Thermal imaging

The work carried out in the Ph.D. project is based on vision systems using thermal camera technologies. This section gives a brief introduction to thermal imaging and the capabilities of this technology within the scope of the Ph.D. project.

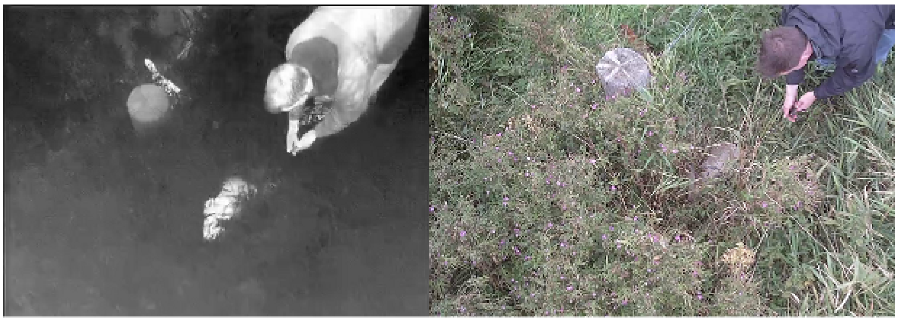


Figure 3.2: *Image of rabbit, human and chicken in both the thermal and visual spectrum*

The main advantage of infrared imaging is that it is invariant to illumination and color balance, which are always changing in outdoor applications. In figure 3.2 another advantage of using thermal imaging is seen. Here it is easy to spot the rabbit (in the center of the image) in the thermal image, but in the visual spectrum the color and fur works as a camouflage in the dense grass. This property can be utilized for detection of wildlife in dense vegetation.

Infrared imaging can be divided into active and passive sensors, where the active sensors require infrared illumination to work. This technology is used in most night vision devices. Thermal cameras are passive sensors, which operates in the mid- (MWIR) and long-wavelength (LWIR) infrared spectrum. In the MWIR and LWIR infrared spectrum (3–14 μm), radiation is emitted by the objects themselves, with a dominating wavelength and intensity depending on the temperature [47].

Thermal imaging is commercially available, and the technology has developed quickly over the last decades. This has resulted in both better and cheaper cameras, and the technology is now being introduced to a wide range of different applications, such as building inspection, medical science, agriculture, fire detection, and surveillance [47].

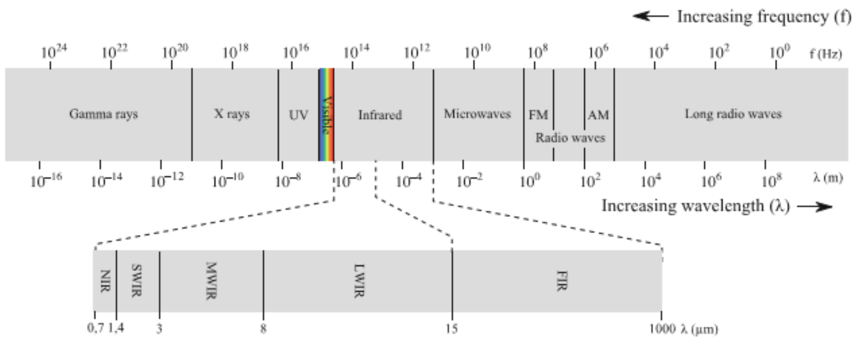


Figure 3.3: *The subdivision of the infrared domain [47]*

3.2.1.1 Thermal radiation

All objects with a temperature above absolute zero (-$273.15° C$) emit infrared radiation. This is often referred to as thermal radiation. The thermal radiation from an object is a function of both temperature and wavelength. In figure 3.3, the wavelengths of the infrared domain is shown, together with the division of the infrared spectrum into several regions.

In figure 3.4 the radiation intensity for different temperature levels is shown. The radiation intensity is described by Planck's wavelength distribution function [117]. It is seen that objects with temperatures from $0° C$ to $37° C$ have their peak within the LWIR band. Due to this, most thermal cameras used for surveillance are designed for these wavelengths, and this has also been utilized in this research. The radiation intensity showed in figure

3.4 is based on black body radiation. Most materials in practical applications are assumed to be so called grey bodies, which have a constant scale factor of the radiation between 0 and 1 [47]. This factor is called the emissivity and is 1 for black bodies. As an example, the emissivity for human skin is very close to 1, whereas it is very low for polished silver (0.02).

A thorough review of thermal cameras and their applications can be found in [47].

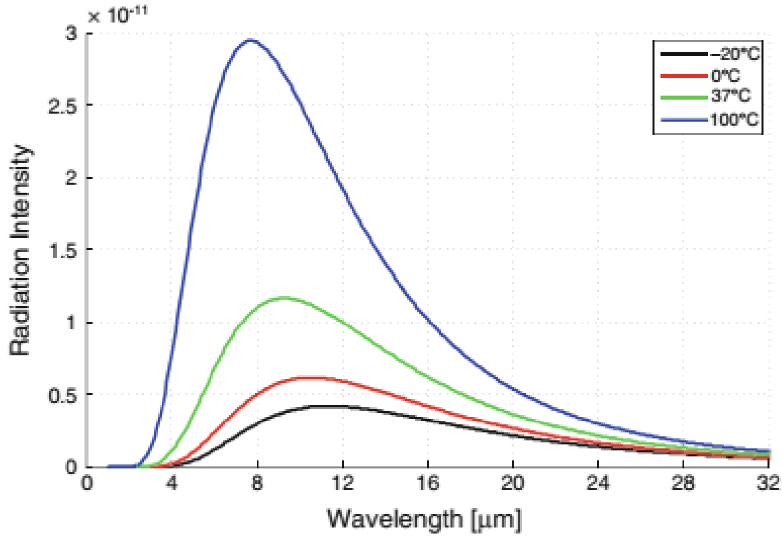


Figure 3.4: *Radiation intensity of black body at four temperatures [47]*

3.3 Research Contributions

Here, the contributions from the published, or submitted work, regarding wildlife-friendly farming are presented. The contributions are presented according to the identified components presented in table 3.1.

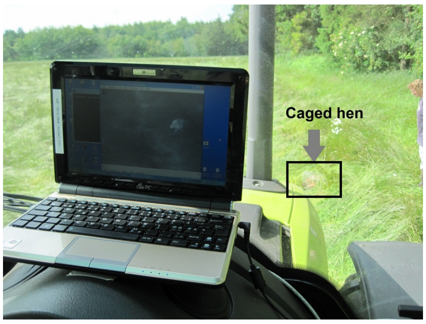
3.3.1 Detection

The output of thermal imaging is a greyscale image, where the intensity is related to the measured temperature. Ideally, the thermal radiation of the animals exceeds the radiation from the background, which makes detection of the animals straightforward. However, during sunlight periods, the temperature difference between animal and background or other objects with high emissivity may be subtle or nonexistent. This was part of the problem with the infrared based solution presented in [58]. Here, weather conditions had to be just right for the system to achieve good performance.

In [38] and [39] thermal imaging is used for person detection. The authors present thermal images of people at different times of day and during summer and winter. Here it is clear that the object of interest (people) does not always appear brighter (higher temperature) than the background. They propose background subtraction techniques, followed by a contour based approach to detect people in the thermal images. Background subtraction is also utilized in [54, 77, 128], however, this approach is not suitable for mobile application with non-fixed cameras.

Another approach is detection of hot spots based on a fixed temperature threshold [32, 78, 116, 112]. In [102] a probabilistic approach for defining the threshold value is presented, however, it is still a fixed value. The detection is usually followed by a subsequent classification of human versus non-human objects based on shape or size. In people detection and recognition, the *Histogram of Oriented Gradients* (HOG), which is a shape-based feature, is a frequently used feature in the classification step [49, 74, 75].

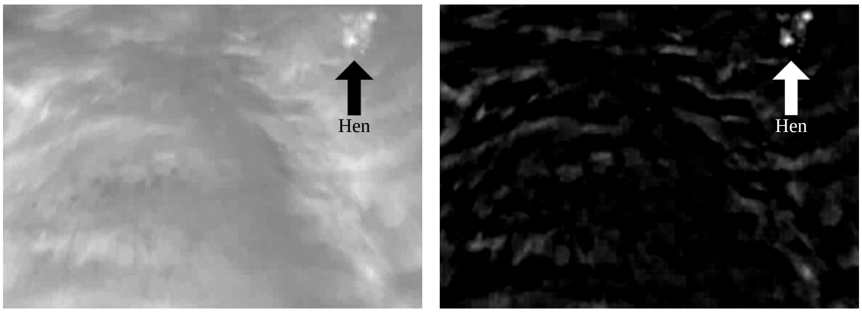
(a) *Camera placement*



(b) *Photo from inside the tractor. A caged hen is emphasized in the image, and it can also be seen on the laptop screen*

Figure 3.5: *Experimental setup for investigation of using tractor mounted thermal cameras to detect animals in grass. Photos by Ole Green*

In [63], a tractor mounted system was discarded as the quality of the thermal images were affected by the motion of the tractor. This problem occurred as the camera was mounted on a mechanical arm next to the mower, looking down in the grass. Hence, the limited frame rate of the thermal camera and vibrations resulted in motion blur, in the thermal image.



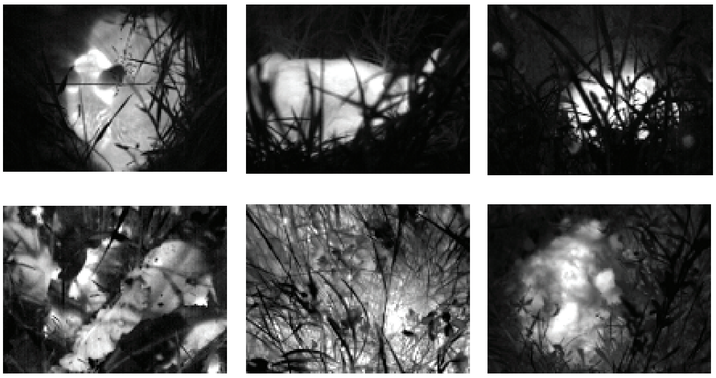
(a) *Thermal image of heated grass patches and the hen*

(b) *Thermal image after filtering. The hen is enhanced and the patches are almost removed*

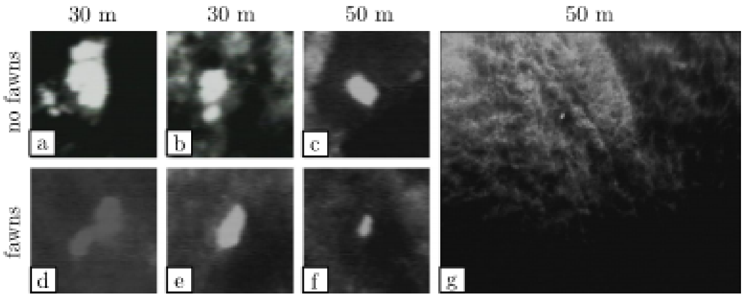
Figure 3.6: *Pre-processing of thermal image to enhance animal versus background*

In [P6] a different tractor mounted solution is investigated. Here, a thermal camera with a resolution of 320×280 pixels and a frame rate of 9 frames per second was used for the experiment. It was placed on top of the hood of the tractor (see figure 3.5a) which gives a different field of view compared to [63]. A hen and a rabbit (see figure 3.5b) were placed in dense grass in front of the tractor, and recordings were performed at different driving speeds. This resulted in eight different test runs which were evaluated in the paper.

The recorded data were utilized for the development of an adaptive detection algorithm based on digital image filtering and an adaptive threshold value. To enhance animal versus background, the Laplacian of Gaussian (LoG) filter is applied before thresholding the thermal image. In figure 3.6 a thermal image from one of the experiments is shown. Here the thermal pixel values of surrounding grass patches are comparable to animal pixels (the hen is marked with the black arrow in figure 3.6a). However, the thermal signatures of grass patches are more diffuse than the thermal signature of the animal, and the LoG is, therefore, able to enhance the animal versus background, as seen in figure 3.6b.



(a) *Thermal images of roe deer fawns versus other hot objects [28]. The top row is fawns and the bottom row is not fawns*



(b) *Thermal images of roe deer fawns versus other hot objects. Images are captures from a UAV [64]*

Figure 3.7: *Examples of thermal images of fawns versus non-fawns. In the thermal image it can be very hard to distinguish between the two*

The subsequent detection of the animals (hot spots) is based on an adaptive thresholding, where the threshold value is based on the maximum pixel value of the current image compared to the mean value of maximum pixel values of previous images (the previous 10 images were used in the paper). The maximum pixel values increase significantly when an animal is present in the image, and this rapid increase in the values can be used to detect the animal. The threshold value is, therefore, adaptively set with respect to the maximum pixel value within the image, when a significant increase in maximum values has been detected. When a significant decrease in maximum values is detected, the threshold value is set to a default value above the current maximum value within the image to avoid false detection. A prerequisite for this algorithm to achieve good performance is that the thermal camera records data at a fixed temperature interval, where the maximum temperatures are set to the expected temperature of the animals.

The proposed algorithm has been tested using two different animals (hen and rabbit) at six different driving speeds. Detection was almost 100% for all but one test run. In this test run, the hen was covered in very dense grass, and the detection algorithm was only able to detect the animal in 26% of the frames (were the animal was present). These frames were the four frames, where the animal was closest to the camera. This would give the farmer under one second to react. The risk of not detecting an animal due to high density of grass is, of course, a disadvantage to the tractor mounted solution. In the study, the field of view of the camera was approximately two meters wide with a distance to the crops of approximately five meters. This is only a smaller part of the potential working width in mowing operations. Therefore, there is a need to increase the field of view, both in width and distance, and, most important, the visibility in very dense grass. This could be achieved by increasing the distance to the crop by different camera positioning, multiple cameras or other lens types. This could potentially increase detection distance, and should be considered in future research.

In [P6], the heated grass patches were removed using image filtering. This was possible as the thermal radiation is more diffuse for these patches. However, as seen in figure 3.7, molehills, rocks and other objects may have a very similar thermal signature to animals, and detection based on thresholding alone is not a robust method, as false detection would occur [28, 64, 65]. Here, classification, like in people detection, could increase performance.

> **Contribution 8.** Investigation of tractor mounted thermal camera system for automatic detection of animals in during grass mowing

> **Contribution 9.** An adaptive detection algorithm for detection of hot spots in thermal images

3.3.2 Recognition of animals

The similarities between animal and non-animal hot spots in thermal images motivates research within recognition of animals. There is little research within automatic detection and recognition of animals in thermal images. However, in people detection, the intial segmentation, or detection of regions of interest, is usually followed by a recognition of human versus non-human [49, 54, 74, 75, 77, 128].

In [28] an algorithm for recognition of roe deer fawns is presented. The algorithm is based on *Normalized Compression Distance* for features extraction and a clustering algorithm for classification. The dataset consists of 103 images, with 26 containing fawns hidden in grass. The same dataset is used in [65], where *Fast-Compression-Distance* is applied in the feature extraction step, and a nearest neighbor classifier is used for classification. In both papers, the features are derived from a dictionary, generated by a compression algorithm. The proposed algorithms perform well on the dataset evaluated in the papers. However, even though the features are scale invariant, they are not rotation invariant, and they rely on absolute temperature measurements.

Another algorithm for identification of deer, to avoid deer vehicle-crashed, is presented in [140]. Here, HOG features are utilized followed by an SVM classifier. Their method relies on occlusion-free side-view images, and performs poorly if these criteria are not met.

Inspired by the work carried out in [64], the experiment and contribution of [P8] is based on top-view images. This reduces problems with vegetation density. Another advantage, which is also part of the hypothesis of this research, is that images in the visual domain could help increase recognition performance in a sensor fusion setup [63]. Almost occlusion-free images would enable conventional cameras to be utilized for sensor fusion.

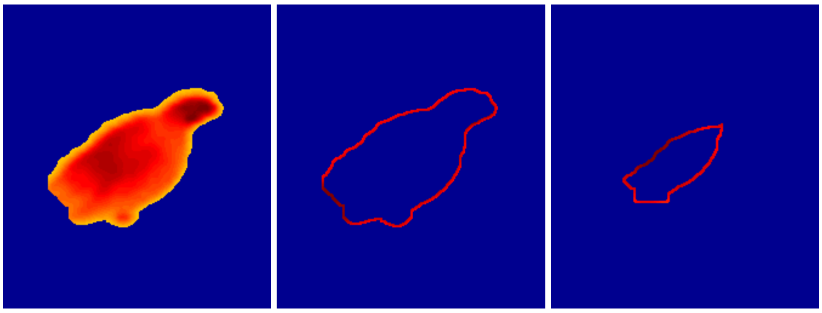
In figure 3.8 the experimental setup used in [P8] is shown. A thermal- and a conventional camera is mounted next to each other, looking down on the ground, where animals are manually placed[4]. The telescopic boom lift could adjust recording heights from 3 to 43 meters, thus simulating a UAV. The reason for not using a commercially available UAV, as the Huginn X1[5], is that these are designed for real time manual operation, hence, real time transmission of compressed thermal video is favored compared to recorded un-compressed data. This was also part of the problem in [64], where data was captured on a ground station based on radio transmitted data. The compression of the thermal images could have removed important detail information, which could be utilized for automated recognition of animals. Therefore, un-compressed data was a priority in [P8].

[4] Thanks to Børnebondegården for lending us the animals (www.børnebondegården.dk)
[5] http://sky-watch.dk/product-line.aspx

Figure 3.8: *A photo from the experiment in [P8]. Photo by Kim Arild Steen*

Even though color images were recorded, it has not been possible to investigate and implement a sensor fusion based recognition of wildlife within the timeframe of the Ph.D. project. Therefore, the contribution of [P8] is focused on recognition based on thermal images and thermal feature extraction. This approach is comparable to the related work in [28, 65, 140] where thermal images were utilized for animal recognition. However, in the context of top-view images, both shape and rotation invariance is important, as the animals are lying flat on the ground in various positions. This is part of the contribution of [P8].

The thermal feature extraction presented in [P8] is based on *Shrinking Thermal Contours*. The main idea of this is to extract thermal pixel values from small slices of the detected hot spots. This is accomplished by the process shown in figure 3.9. The contour of a detected hot spot (fig. 3.9a) is extracted by morphological boundary extraction with a small disk shaped structure element. These contours (fig 3.9b and fig. 3.9c) are used as a binary masks to extract thermal values from the thermal object. For each of the contours, the mean temperature is calculated, which results in a thermal signature for each hot spot. In figure 3.10 examples of these signatures are shown. The x-axis indicates the contour number. Here contour number -1 and 0 are not



(a) *Thermal image of detected object* **(b)** *First contour of the detected object* **(c)** *Seventh contour of the detected object*

Figure 3.9: *The process of thermal contour extraction*

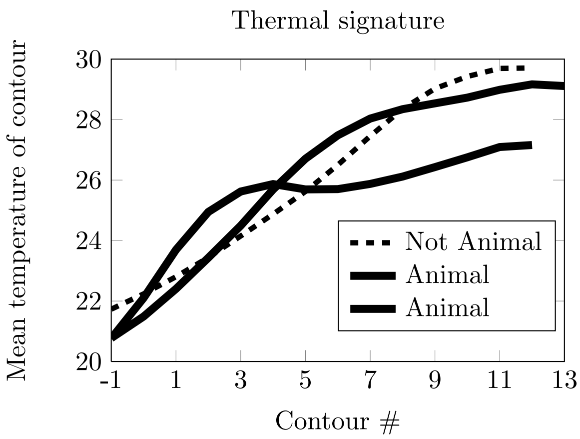


Figure 3.10: *Thermal signatures extracted from shrinking thermal contours. Contour number* -1 *and* 0 *are not part of the object, but used for edge feature extraction*

part of the detected hot spot, but rather thermal pixel values around the hot spot. These contours are used for edge feature extraction.

From these thermal signatures, three features, for each detected hot spot, are extracted: Center-Edge Difference, Variance and Edge. The Center-Edge Difference feature is the difference in mean temperature between the innermost contour and the edge contour (contour #1). Based on visual observations, and as seen in figure 3.10, this difference is smaller for animals, as the thermal signature is more uniform than, as an example, molehills[6].

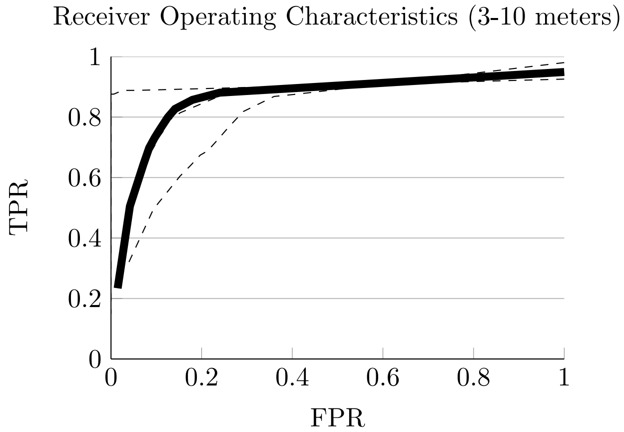
The variance measures the variability in the thermal signature and give a statistical measure of the distribution of thermal pixel values within the detected hot spot. The edge features are based on the transition in temperature from background to the object, which is calculated as the difference between contour #1 and contour #-1. As observed in [P6], regarding automatic detection of hot spots, the changes in temperature are more abrupt when the object is a living creature compared to heated grass patches. This observation is utilized in the proposed feature extraction algorithm. These features were used in the subsequent classification.

The classification is performed using a kNN classifier. In [P8] 80 animal feature vectors and 95 non-animal feature vectors were used as the training data for the classifier. Both resolution and the distance between the animal and the camera [123] affect the quality of the thermal images. Therefore, the algorithm was evaluated at two different height intervals. A total of 3987 frames containing one, two or three animals were evaluated, and the results are shown in figure 3.11. By sweeping the threshold for the kNN classifier, the *Receiver Operating Characteristics* (ROC) for the classifier could be obtained. The area under the curves works as a performance measure, and it is that the algorithms achieve the best performance at height interval 3-10 meters. The captured dataset comprised of five recordings, where the height was increased from 3 to 30 meters. All five datasets included useful data at height interval 10-20 meters; however, only three of the recordings were useful at height interval 3-10 meters (as seen in the figure).

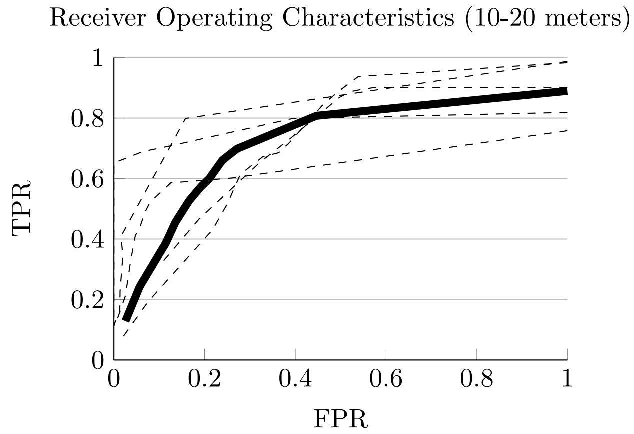
In the figure, it is seen that performance decreases as height increases. This fits well with observations from [64] where the authors were able to manually detect row deer fawns at 30 meters, but had problems at 50 meters with a thermal camera with a resolution of 640×512 pixels. The animals used in [P8] are smaller than roe deer fawns, which results in fewer thermal pixels, compared to the roe deer fawns. This means that a UAV has to fly at a lower height to detect and recognize smaller animals like pheasants and hares.

When the height is increased, the image frame could contain more false positive candidates, which would decrease performance measures. However, empirical experiments with lowered spatial resolution at height interval 3-10 meters indicate that the spatial resolution is also important for performance.

[6] They are heated by the sun from above, and due to its shape, the temperature increase from edge to center is not as uniform as with an animal



(a) *ROC of the three recordings plotted as dashed lines, and mean performance as bold*



(b) *ROC of the five recordings plotted as dashed lines, and mean performance as bold*

Figure 3.11: *Receiver Operating Characteristics for two different height intervals*

The goal of the research within recognition of wildlife in thermal images is to decrease the number of false positives. Sensor fusion techniques, temporal information and other thermal features, derived from the thermal signatures, should be investigated within this context.

Contribution 10. Feature extraction methods for recognition of animals in thermal images

3.3.3 Decision of action

Once an animal has been detected, the next task is to act upon this information. In [64] they move the animals manually prior to mowing. Other approaches include stopping the machine or avoiding the animal. In the case of mobile animals, simply slowing down could reduce the risk of injuring or killing it. However, immobile animals are not able to escape even if the machine is driving slowly. Independent of the solution, the sensor system needs to detect and recognize the animal in time. Here top-view images are more reliable as they are not as affected by very dense vegetation as tractor mounted systems are. These considerations are important in future research regarding wildlife-friendly farming.

4

Discussion and Conclusion

This chapter summarizes and concludes the results achieved in this thesis. The hypothesis and objectives of the thesis defined in Chapter 1 are related to the contributions regarding an adaptive frightening device and wildlife-friendly farming presented in Chapters 2 and 3. The proposed theoretical frameworks, the developed pattern recognition and signal processing algorithms, and the conducted field experiments comprise the results of this thesis.

4.1 Introduction

This thesis introduces various pattern recognition and signal processing algorithms in the context of human-wildlife conflicts. The scientific contributions of the Ph.D. project are mostly focused on how pattern recognition and signal processing methods can be applied in the design of solutions for reducing human-wildlife conflicts. The work has been motivated by limitations in existing devices utilized in wildlife damage management, and the opportunity to investigate the possible effect of using intelligent sensor strategies in the interaction with wildlife.

The purpose of this chapter is to evaluate the outcome of the thesis and assess to what extent the hypothesis and objectives have been met. Section 4.2, summarizes the research contributions made, and evaluates and compare the contributions to state-of-art research within the domain of the contribution. Proposed future work is described and presented in Section 4.3, and the conclusion of the Ph.D. thesis is found in Section 4.4.

4.2 Research Contributions

The research contributions are grouped into two main categories: *adaptive frightening device* and *wildlife-friendly farming*. The contributions to the proposed frameworks are shown in table 4.1. The contributions within an adaptive frightening device are focused on applied pattern recognition and signal processing, whereas the contributions within wildlife-friendly farming are within digital image processing of thermal images.

Table 4.1: *Overview of publications and their contributions to the defined frameworks*

Adaptive Frightening Device					
Ref.	Data Collection	Detection	Species Recognition	Behavior Recognition	Decision of Action
[P1]	x				
[P5]		x	x		
[P2]				x	
[P7]				(x)[a]	
[P4]		x		x	
[P3]					x
Wildife-friendly Farming					
Ref.		Detection		Animal Recognition	Decision of Action[b]
[P6]		x			
[P8]		x		x	

[a] Behavior recognition was not implemented, however acoustic source tracking could be utilized for recognition of behavior
[b] The final component of the proposed framework has not been investigated during this PhD project

4.2.1 Adaptive frightening device

This section summarizes the contributions within the context of an adaptive frightening device.

4.2.1.1 Detection and recognition of conflict species

The contributions [C1] and [C2] presented in [P5] are part of a generic framework for detection and recognition of conflict species. The purpose of the framework is to detect and recognize species of bird flocks to enable corresponding disruptive stimuli. Here, the purpose of the acoustic based recognition differs from other studies, as these are mostly focused on classification of individuals or specific call types. When dealing with specific calls, both the choice of acoustic features and pattern recognition algorithm can be designed specifically to this [19].

An adaptive frightening device needs to handle very different bird species at different geographic locations. Hence, the developed algorithm should be generic and easily adjustable to various species. In [P5] MFCC features were chosen as acoustic features for all conflict species. The focus of the paper is on geese and rooks, however, other species can be included in the framework. MFCC features have yielded good results across a variety of taxa including frogs, crickets, birds, cows and fish [30], and this generic feature of MFCC is attractive within the context of the proposed framework.

In [P5], the robustness to background noise is implemented in the detection step of the algorithm, as background sounds and noise are modeled as a background model. This model is compared to the conflict species model in the detection step. This performs well, when the background noise is non-additive, meaning that the noise occurs at another time window than the conflict species sounds. However, in the case of additive noise, MFCC features are sensitive [46], which affects performance. This was also observed in the publication, as performance decreased when SNR decreased. Here, automatic noise removal techniques could be implemented to increase classification performance. In [19] and [45] background noise is removed via noise level estimation of frequency bands known not to contain any bird vocalization. This technique could be applied in the proposed framework.

Both detection and species recognition are call-independent, meaning that the framework should detect and recognize the conflict species regardless of call type. Here, GMM based recognition has been chosen, as this technique is also frequently used in human speaker recognition, which includes the same functional properties. The contribution of using GMM to classify between species [C2] is not novel, as it has been used in acoustic based recognition of birds, in other contexts [30, 130]. However, these papers do not deal with the soundscape of multiple birds within a flock. The GMM based framework also allows for future work regarding model robustness. This will be discussed in Section 4.3.

4.2.1.2 Behavior recognition

Research regarding a direct link between vocalizations and behavior is limited. In [81] and [97] vocalizations are utilized to measure pig welfare and stress levels, and in [36] examples of utilizing automatic vocalization analysis to monitor thermal stress in the poultry industry is shown. Using acoustic information to monitor behavior has great potential as the available technology is cost-efficient and non-invasive. This is important in wildlife damage management.

Contribution [C3] is a novel method to automatically monitor bird flock behavior based on their interspecies acoustic communication. Within wildlife damage management, this can be utilized to react to the specific behavior and thereby altering the flocking behavior based on stimuli. Audio based behavior recognition could also be beneficial to research within biology and other aspects of agricultural production. The contribution in [P2] show that automatic audio based recognition of vocalization is not limited to single calls and species, but it can be utilized to monitor the behavior of multiple animals by recognizing their interspecies communication.

Another small contribution within audio based behavior recognition is the use of GFCC features. The GFCC features have been utilized in bird vocalization recognition in [15]. However, in [P2] the motivation for using GFCC is based on their ability to map hearing capabilities of animals instead of humans. In human speech recognition, MFCC features are utilized because of the ability to describe human sound perception. Therefore, the use of GFCC features seems promising in the context of interspecies communication.

Contribution [C4], which is presented in [P4], is a framework for audio-visual behavior recognition. Most research regarding the link between visual and acoustic information for animal behavior recognition utilize either manual observations or manual inspections of video recordings. In [96] video recordings of chickens were used for manual detection of group behavioral pattern in an experiment to link their vocalizations with the thermal environment. Likewise, [125] uses video recordings for linking dolphin sound to their location and behavior, and in [97] manual observations were used to link vocalizations to the stress level of piglets. The link between visual information, like movement or posture, and behavior is more straightforward than using audio. Therefore video based recognition is a frequently used method in behavior recognition [17, 36, 80, 84, 127, 138]. However, automated methods for fusing these video based algorithms with audio has not been reported. This makes contribution [C4] a novel approach in automated animal behavior analysis and recognition.

The performance the audio based approach, given by contribution [C3], is affected by low SNR. This was not investigated further in the publication [P2], however, low SNR would cause class overlap in the classification. This could be improved by contribution [C4], where fusion with video data would increase performance. However, the limited field of view of the vision system and the limited capabilities during sunrise and sunset[1] makes an

[1] Goose flocks are especially active in early morning hours

audio-visual approach impractical within the context of cost-efficient wildlife damage management.

The advantage of using visual information in behavior recognition is the direct link between movement and behavior. However, this property is not unique for camera based systems, as localization through sensor arrays may provide similar results. This is the motivation for contributions [C5] and [C6], where a framework for tracking acoustic sources in a long baseline microphone array is proposed. Here the contributions are more general within the context of target tracking. Contribution [C5] is a theoretical framework for tracking maneuvering targets within a long baseline microphone array. The tracking algorithm is based on the CRPF, which is a new class of particle filters. The CRPF has been utilized in target tracking within wireless sensor networks [90], however, CRPF tracking in long baseline arrays is a novel contribution of [P7].

Contribution [C6] is a modification to the cost function within energy based localization. The modified cost function is designed to increase tracking performance in the case of low SNR and sudden wind gusts. Here, the CRPF framework allows for the use of user-defined cost functions, which has been implemented in [P7].

The limitation of the proposed tracking framework is the computational cost of the CRPF compared to LS localization or Kalman filter based tracking. Computational cost is an important design parameter in sensor arrays due to energy consumption. Future work regarding the microphone array based tracking should include more research within computational efficient algorithms. Furthermore automated behavior recognition, based on estimates of source movements, should be developed. This will be further discussed in Section 4.3.

Three different methods for behavior recognition have been investigated during this Ph.D. project. All methods could contribute to a sensor based solution for efficient wildlife damage management. However, audio based behavior recognition could be sensitive to low SNR, and the method would also require labeled behavior data, which is time consuming to collect. The audio-visual fusion is less sensitive to low SNR, but the use of vision systems in this context would be impractical. The microphone array based method seems the most promising, as it is able to cover a large region, and the cooperation between sensors could make the system less sensitive to low SNR. Furthermore, the automated behavior recognition could be based on flock movement rather than soundscape, which could potentially increase behavior recognition performance. Each sensor could utilize the detection and species recognition algorithms from contribution [C1] and [C2], and thereby make it possible to develop species specific tracking.

4.2.1.3 Decision of action

The last part of the proposed framework for an adaptive frightening device is a decision on how to act based on knowledge from detection and recognition algorithms. In [P3], the contribution [C7] is an investigation of the effectiveness of the proposed framework for barnacle geese management. The investigation itself is novel, as it is based on the proposed framework developed during this thesis. Alarm and distress calls have been used in related work [16, 136, 137], however, these systems do not include the ability to automatically recognize and react to behavior.

The investigation showed promising results as the system was able to reduce crop damage to a minimum during almost six weeks[2]. However, the study was only based on one system, and the results lack statistical power. Furthermore, the system experiences many false detections caused by sudden wind gusts and other background noises. Despite this, habituation was not observed during the experiment, which may be because the geese were not in the area at the same time as the false detections. In conclusion, it must be concluded that more field experiments are required to achieve significant practical results using the proposed framework.

4.2.2 Wildlife-friendly farming

This section summarizes the contributions within the context of wildlife-friendly farming.

4.2.2.1 Detection

In [63], a tractor mounted system was discarded as the quality of the thermal images were affected by the motion of the tractor. This problem occurred as the camera was mounted on a mechanical arm next to the mower, looking down in the grass. Here, the limited frame rate of the thermal camera together with high driving speeds and vibrations from the machine resulted in motion blur, in the thermal image. In [P6] a different tractor mounted solution is investigated. Here, the thermal camera is placed on top of the hood of the tractor. This investigation constitute contribution [C8]. This tractor mounted system does not affect the quality of the thermal images. However, as presented in Chapter 3, timely detection of the animals in the grass is important,

[2] Until the geese eventually migrated north

and this could be a problem when the vegetation is dense. Therefore, the field of view of the thermal camera needs to be investigated further.

Contribution [C9] is an adaptive detection algorithm based on digital image filtering and adaptive thresholding. The proposed detection algorithm makes it possible to detect animals while avoiding detection of heated grass patches. This is an important feature, as multiple false detections does not promote efficiency during farming operations. The proposed method is not able to distinguish between animals and other hot objects[3], like molehills and stones. Here a detection and recognition of the animals, using other features than absolute temperature is suggested. Preliminary research within this has been performed during this Ph.D. project [P8].

4.2.2.2 Recognition of animals

In [P8], an algorithm for extraction of thermal features is presented [C10]. These thermal features are used for recognition of animal versus non-animal in top-view thermal images. Top-view images reduce problems with vegetation density, and enables the use of other sensors, such as conventional cameras[4].

There exist limited work regarding recognition of animals in thermal images. In [28] and [65] template based methods are presented. The proposed algorithms in these papers perform well on the limited dataset evaluated in the papers. However, even though the thermal features are scale invariant, they are not rotation invariant, and they rely on absolute temperature measurements. Another algorithm, which identifies deer, to avoid deer vehicle-crashed, is presented in [140]. Here, HOG features are utilized for recognition. Their method relies on occlusion-free side-view images, and performs poorly if these criteria are not met.

The thermal feature extraction technique presented in [P8] is a novel method for extracting scale and rotation invariant thermal features. The features are derived from a thermal signature generated by morphological contour extraction. The method has been tested in a controlled experiment with good results for a selected height interval (3-10 meters). However, additional thermal features could be calculated from the thermal signature, and sensor fusion methods could be utilized to increase performance. Additionally, more experiments including more animals should be conducted.

[3] Heated by the sun

[4] That rely on a hard line of sight measurement

4.3 Future Work

Here proposed future work regarding an adaptive frightening device and wildlife-friendly farming is presented.

4.3.1 Adaptive frightening device

For the purpose of overview, this subsection has been divided into the specific components of the proposed framework for an adaptive frightening device.

4.3.1.1 Detection and species recognition

Both detection and species recognition are based on GMM, which is a statistical model of the different classes (e.g. conflict species versus background). In the detection algorithm, the background model comprises of background sounds, which are not from a defined conflict species. The quality of this model, with respect to classification performance, is dependent on the training data available. It is impossible to sample the entire soundscape of an agricultural field, as this varies from location to location. However, a local background model[5] could be used. This could be accomplished by incremental learning techniques [41, 94], where both background and species models could be incrementally constructed based on the soundscape of the specific location.

Incremental learning is part of semi-supervised learning, where new observations may be chosen as part of training data in an incremental framework. There is much work regarding incremental learning with respect to GMMs using the EM-algorithm, and the theoretical algorithms for implementing this exists. However, a very important step in this is the automated choice of which new observations are included as new training data. This is a difficult task, as the quality of the model relies on the quality of data, and wrong decisions could decrease performance rather than increasing it. Future work should include development of methods to automatically choose this data and investigate the performance over time in real life scenarios.

4.3.1.2 Behavior recognition

A long baseline microphone array is a promising method for automatic flock behavior recognition. It is possible to monitor a large region and behavior recognition could be performed on the basis of motion rather than soundscape. Future work regarding acoustic source tracking in long baseline array

[5] A model representing the common sounds in a specific geographic location

should focus on automated behavior recognition based on tracking results. Here, methods from the research found in [132] and [18] can be used for inspiration.

Computational efficiency is an important design parameter in sensor arrays due to energy consumption. In [P7] the CRPF filter were used for tracking. However, other, more computationally efficient tracking filters are available, including the IMM-KF [85]. Unlike, the standard Kalman filter or the Extended Kalman filter, the IMM-KF incorporates multiple dynamic models, and fuse these models to reach a single state estimate. This method has proven very efficient in radar tracking [69, 101], where rapid changes between linear and non-linear movement often happen. The IMM-KF was not utilized in the contribution of this thesis as it does not allow user-defined cost functions in the same manner as the CRPF framework. However, future work could include incorporating the concepts from contribution [C5] and [C6] within the IMM-KF framework to increase computational efficiency.

4.3.1.3 Decision of action

The results from the field experiments lack statistical power, and more experiments should be conducted. An important part of this is, of course, to investigate the performance of the proposed framework, but also to gain insight in which components are more important to performance than other. Thereby, requirements for such a system could be defined.

4.3.2 Wildlife-friendly farming

An algorithm for recognition of animals in thermal images has been presented. However, more thermal features and other sensors could be utilized to increase performance. An important task in the context of a sensor based wildlife safety system is to achieve robust detection, while simultaneously avoiding false detections. A high number of false detection would decrease efficiency and also introduce a lack in confidence towards the system, which could result in farmers discarding the solution. Therefore, sensor fusion should be investigated further to increase detection performance.

Futhermore, field experiments should be conducted in multiple fields, in various weather conditions. Based on this, the constraints of sensor based solutions could be found, and requirements and guidelines for future research could be defined.

4.4 Conclusion

The hypothesis of this thesis was that *wildlife damage management can be performed in a more ethical, efficient and wildlife-friendly manner, if based on new sensor technology, pattern recognition and automation within tools and methods for wildlife damage management*. To investigate this, three main objectives were defined, which inluded development of sensor based methods for detecting and recognizing wildlife, and investigate the effect of using smart sensing in the context of wildlife damage management.

The research within an adaptive frightening device has been focused on the development of algorithms and methods to introduce the intelligent use of sensor technologies within wildlife damage management. Here, acoustic pattern recognition based on cepstral feature extraction and GMMs have been utilized to develop algorithms for detection and recognition of conflict species. The timely detection and recognition of conflict species enables a system to monitor a large region, and only apply a disruptive stimuli when it is needed. Futhermore, the stimuli can be targeted towards specific species. A framework, which implements an adaptive system has been presented. The framework includes automated flock behavior recognition, which can be accomplished through analysis of soundscape or flock movement. Here, acoustic, audio-visual fusion and array methods have been proposed. These contributions promote ethical and efficient wildlife damage management, as the system is able to monitor possible habituation and react accordingly. The feature of having a wildlife damage management system capable of monitoring its own performance and be species specific is a novel concept, which can be further developed based on the work carried out during this Ph.D. project.

An automatic detection and recognition of animals in mowing operations would promote both efficient and wildlife-friendly farming. The use of thermal cameras allows for robust detection of animals during mowing operations. In this thesis, an algorithm for detection and recognition of wildlife in thermal images have been presented. The recognition of wildlife in thermal images is based on a novel thermal feature extraction method, which captures the thermal signature of detected hot spots. However, using this technology alone will not solve the task of automated detection and recognition of animals, as it is sensitive to false positives, and sensor fusion techniques should be investigated to improve this.

The achieved results and contributions in this thesis are a significant step towards a more efficient sensor based solutions for wildlife-friendly farming and reduction of human-wildlife conflicts within agriculture.

Part II

Publications - Adaptive Frightening Device

5

A Multimedia Capture System for Wildlife Studies

The paper presented in this chapter is a peer-reviewed conference paper and has been presented at *Emerging 2011*.

[P1] Kim Arild Steen, Henrik Karstoft and Ole Green (2011). *A Multimedia Capture System for Wildlife Studies*. Paper presented at The Third International Conference on Emerging Network Intelligence, Lissabon, Portugal.

6

A Vocal-Based Analytical Method for Goose Behaviour Recognition

The paper presented in this chapter has been published in *Sensors*.

[P2] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2012). *A Vocal-Based Analytical Method for Goose Behaviour Recognition*. Sensors 12(3), pp. 3773-3788

7

Audio-Based Detection and Recognition of Conflict Species in Outdoor Environments using Pattern Recognition Methods

The paper presented in this chapter has been published in *Applied Engineering in Agriculture*.

[P5] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2014). *Audio-Based Detection and Recognition of Conflict Species in Outdoor Environments Using Pattern Recognition Methods*. Applied Engineering in Agriculture vol. 30(1), pp. 89-96

8

Audio-Visual Recognition of Goose Flocking Behavior

The paper presented in this chapter has been published in *International Journal of Pattern Recognition and Artificial Intelligence*.

[P4] Kim Arild Steen, Ole Roland Therkildsen, Ole Green and Henrik Karstoft (2013). *Audio-Visual Recognition of Goose Flocking Behavior*. International Journal of Pattern Recognition and Artificial Intelligence. 27(7), pp. 21

9

Acoustic Source Tracking in Long Baseline Microphone Arrays

The paper presented in this chapter has been accepted for publication in *Applied Acoustics*.

[P7] Kim Arild Steen, James H. McClellan, Ole Green and Henrik Karstoft. *Acoustic Source Tracking in Long Baseline Microphone Arrays*. Submitted for publication in Applied Acoustics, March 2014.

10

An Audio Based Adaptive Goose Scaring Device

The paper presented in this chapter has been presented at *CIOSTA (Commission Internationale de l'Organisation Scientifique du Travail en Agriculture)*.

[P3] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green (2013). *An Audio Based Adaptive Goose Scaring Device*. Paper presented at CIOSTA XXXV Conference, Billund, Danmark.

Part III

Publications - Wildlife-friendly Farming

11

Automatic Detection of Animals in Mowing Operations Using Thermal Cameras

The paper presented in this chapter has been published in *Sensors*.

[P6] Kim Arild Steen, Andrés Villa-Henriksen, Ole Roland Therkildsen and Ole Green (2012). *Automatic Detection of Animals in Mowing Operations Using Thermal Cameras*. Sensors 12(6), pp. 7587-7597

12

Detection and Recognition of Wildlife in Thermal Images

The paper presented in this chapter has been submitted to *IEEE International Conference on Image Processing 2014*.

[P8] Kim Arild Steen, Rasmus Nyholm Jørgensen, Ole Green and Henrik Karstoft. *Detection and Recognition of Wildlife in Thermal Images*, Submitted to IEEE International Conference on Image Processing, January 2014

Bibliography

[P1] Kim Arild Steen, Henrik Karstoft, and Ole Green. A multimedia capture system for wildlife studies. In *EMERGING 2011, The Third International Conference on Emerging Network Intelligence*, pages 131–134, 2011.

[P2] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft, and Ole Green. A vocal-based analytical method for goose behaviour recognition. *Sensors*, 12(3):3773–3788, 2012.

[P3] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft, and Ole Green. An audio based adaptive goose scaring device. *CIOSTA XXXV Conference*, 2013.

[P4] Kim Arild Steen, Ole Roland Therkildsen, Ole Green, and Henrik Karstoft. Audio-visual recognition of goose flocking behavior. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(07), 2013.

[P5] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft, and Ole Green. Audio-based detection and recognition of conflict species in outdoor environments using pattern recognition methods. *Applied Engineering in Agriculture*, 30(01):89–96, 2014.

[P6] Kim Arild Steen, Andrés Villa-Henriksen, Ole Roland Therkildsen, and Ole Green. Automatic detection of animals in mowing operations using thermal cameras. *Sensors*, 12(6):7587, 2012.

[P7] Kim Arild Steen, James H. McClellan, Ole Green, and Henrik Karstoft. Acoustic source tracking in long baseline microphone arrays. *Submitted to Applied Acoustics*, 2014.

[P8] Kim Arild Steen, Rasmus Nyholm Jørgensen, Ole Green, and Henrik Karstoft. Detection and recognition of wildlife in thermal images. *Submitted to IEEE International Conference on Image Processing*, 2014.

[P9] Kim A Steen, Ole R Therkildsen, Henrik Karstoft, Ole Green, et al. Video-based detection of goose behaviours. *Agriculture and Engineering for a Healthier Life, Valencia, Spain*, page 6, 2012.

[P10] Kim A Steen, Andrés Villa-Henriksen, Ole R Therkildsen, Henrik Karstoft, Ole Green, et al. Automatic detection of animals using thermal imaging. In *International Conference on Agricultural Engineering*, 2012.

[P11] Maiken Bjerg Moller, Kim Arild Steen, Jesper Gaarsdal, and Torben Gregersen. Acoustic pattern recognition on android devices. In *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013*, pages 279–284. IEEE, 2013.

[P12] Kim Arild Steen. Brug af varmefølsomt kamera ved høst. *Plantekongres 2012:-produktion, plan og miljø*, 2012.

[P13] Kim Arild Steen and MS Laursen. Identification and tracking of sows. In *CIGR/NJF seminar 441–Automation and system technology in plant production*.

[P14] Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft, and Ole Green. Wildlife communication: Electrical and computer engineering. Technical report, Aarhus University, Department of Engineering, 2012.

[15] Kuntoro Adi, Michael T Johnson, and Tomasz S Osiejuk. Acoustic censusing using automatic vocalization classification and identity recognition. *The Journal of the Acoustical Society of America*, 127(2):874–883, 2010.

[16] Elizabeth Aguilera, Richard L Knight, and John L Cummings. An evaluation of two hazing methods for urban canada geese. *Wildlife Society Bulletin*, pages 32–35, 1991.

[17] Peter Ahrendt, Torben Gregersen, and Henrik Karstoft. Development of a real-time computer vision system for tracking loose-housed pigs. *Computers and Electronics in Agriculture*, 76(2):169–174, 2011.

[18] Andreas Mantik Ali, Shadnaz Asgari, Travis C Collier, Michael Allen, Lewis Girod, Ralph E Hudson, Kung Yao, Charles E Taylor, and Daniel T Blumstein. An empirical study of collaborative acoustic source localization. *Journal of Signal Processing Systems*, 57(3):415–436, 2009.

[19] Rolf Bardeli, D Wolff, Frank Kurth, M Koch, K-H Tauchert, and K-H Frommolt. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, 31(12):1524–1534, 2010.

[20] Asa Ben-Hur and Jason Weston. A user's guide to support vector machines. In *Data mining techniques for the life sciences*, pages 223–239. Springer, 2010.

[21] Jeff Beringer, Kurt C VerCauteren, and Joshua J Millspaugh. Evaluation of an animal-activated scarecrow and a monofilament fence for reducing deer use of soybean fields. *Wildlife Society Bulletin*, pages 492–498, 2003.

[22] Frédéric Bimbot, Jean-François Bonastre, Corinne Fredouille, Guillaume Gravier, Ivan Magrin-Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega-García, Dijana Petrovska-Delacrétaz, and Douglas A Reynolds. A tutorial on text-independent speaker verification. *EURASIP journal on applied signal processing*, 2004:430–451, 2004.

[23] Christopher M Bishop. *Pattern recognition and machine learning*. springer New York, 2006.

[24] Mary Bomford and Peter H O'Brien. Sonic deterrents in animal damage control: a review of device tests and effectiveness. *Wildlife Society Bulletin*, pages 411–422, 1990.

[25] Judith C Brown and Paris Smaragdis. Hidden markov and gaussian mixture models for automatic call classification. *The Journal of the Acoustical Society of America*, 125(6):EL221–EL224, 2009.

[26] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[27] William M Campbell, Joseph P Campbell, Terry P Gleason, Douglas A Reynolds, and Wade Shen. Speaker verification using support vector machines and high-level features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(7):2085–2094, 2007.

[28] Daniele Cerra, Martin Israel, and Mihai Datcu. Parameter-free clustering: Application to fawns detection. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, volume 3, pages III–467. IEEE, 2009.

[29] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[30] Jinkui Cheng, Yuehua Sun, and Liqiang Ji. A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines. *Pattern Recognition*, 43(11):3846–3852, 2010.

[31] Kwang-Ting Cheng and Yi-Chu Wang. Using mobile gpu for general-purpose computing–a case study of face recognition on smartphones. In *VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on*, pages 1–4. IEEE, 2011.

[32] Grzegorz Cielniak and Tom Duckett. People recognition by mobile robots. *Journal of Intelligent and Fuzzy Systems*, 15(1):21–27, 2004.

[33] Patrick J Clemins, Michael T Johnson, Kirsten M Leong, and Anne Savage. Automatic classification and speaker identification of african elephant (loxodonta africana) vocalizations. *The Journal of the Acoustical Society of America*, 117(2):956–963, 2005.

[34] Patrick J Clemins, Marek B Trawicki, Kuntoro Adi, Jidong Tao, and Michael T Johnson. Generalized perceptual features for vocalization analysis across multiple species. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages I–I. IEEE, 2006.

[35] Michael R Conover. *Resolving human-wildlife conflicts: the science of wildlife damage management*. CRC press, 2001.

[36] Gerard Corkery, Shane Ward, Colum Kenny, and Phil Hemmingway. Incorporating smart sensing technologies into the poultry industry. *Journal of World's Poultry Research*, 3(4):106–128, 2013.

[37] Nicolas Courty and Thomas Corpetti. Crowd motion capture. *Computer Animation and Virtual Worlds*, 18(4-5):361–370, 2007.

[38] James W Davis and Mark A Keck. A two-stage template approach to person detection in thermal imagery. In *Proc. Workshop Applications of Computer Vision*.

[39] James W Davis and Vinay Sharma. Robust background-subtraction for person detection in thermal imagery. *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.

[40] Marian Stamp Dawkins, Hyoung-joo Lee, Corri D Waitt, and Stephen J Roberts. Optical flow patterns in broiler chicken flocks as automated measures of behaviour and gait. *Applied Animal Behaviour Science*, 119(3):203–209, 2009.

[41] Arnaud Declercq and Justus H Piater. Online learning of gaussian mixture models-a two-level approach. In *Proceedings of the third international conference on computer vision theory and applications*, pages 605–611, 2008.

[42] John R Deller, John G Proakis, and John HL Hansen. *Discrete-time processing of speech signals*. IEEE New York, NY, USA, 2000.

[43] Marco F Duarte and Yu Hen Hu. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 64(7):826–838, 2004.

[44] A Fackelmeier and EM Biebl. A multistatic radar array for detecting wild animals during pasture mowing. In *Radar Conference, 2009. EuRAD 2009. European*, pages 477–480. IEEE, 2009.

[45] Seppo Fagerlund. Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

[46] EJS Fox, JD Roberts, and M Bennamoun. Text-independent speaker identification in birds. In *Proceedings of the Interspeech 2006 and Ninth International Conference on Spoken Language Processing*, pages 2122–2125, 2006.

[47] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine Vision and Applications*, pages 1–18, 2013.

[48] FD Galey, R Terra, R Walker, J Adaska, MA Etchebarne, B Puschner, E Fisher, RH Whitlock, T Rocke, D Willoughby, et al. Type c botulism in dairy cattle from feed contaminated with a dead cat. *Journal of veterinary diagnostic investigation*, 12(3):204–209, 2000.

[49] David Geronimo, Antonio M Lopez, Angel Domingo Sappa, and Thorsten Graf. Survey of pedestrian detection for advanced driver assistance systems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7):1239–1258, 2010.

[50] Jason M Gilsdorf, Scott E Hygnstrom, and Kurt C Vercauteren. Use of frightening devices in wildlife damage management. *Integrated Pest Management Reviews*, 7(1):29–45, 2002.

[51] Jason M Gilsdorf, Scott E Hygnstrom, Kurt C VerCauteren, Greg M Clements, Erin E Blankenship, and Richard M Engeman. Evaluation of a deer-activated bio-acoustic frightening device for reducing deer damage in cornfields. *Wildlife Society Bulletin*, 32(2):515–523, 2004.

[52] Lewis Girod, Martin Lukac, Vlad Trifa, and Deborah Estrin. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 71–84. ACM, 2006.

[53] W Paul Gorenzel and Terrell P Salmon. Tape-recorded calls disperse american crows from urban roosts. *Wildlife Society Bulletin*, pages 334–338, 1993.

[54] Emmanuel Goubet, Joseph Katz, and Fatih Porikli. Pedestrian tracking using thermal infrared imaging. In *Proc. SPIE Conf. Infrared Technology and Applications*, pages 797–808, 2006.

[55] C Green. Reducing mortality of grassland wildlife during haying and wheat-harvesting operations. *Oklahoma State University Forestry Publications*, pages 1–4, 1998.

[56] RE Green and TJ Stowe. The decline of the corncrake crex crex in britain and ireland in relation to habitat change. *Journal of Applied Ecology*, pages 689–695, 1993.

[57] Donald D Greenwood. Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hearing research*, 54(2):164–208, 1991.

[58] Peter Haschberger, Michael Bundschuh, and Volker Tank. Infrared sensor for the detection and protection of wildlife. *Optical Engineering*, 35(3):882–889, 1996.

[59] Gustaf Hendeby, Rickard Karlsson, and Fredrik Gustafsson. Particle filtering: the need for speed. *EURASIP Journal on Advances in Signal Processing*, 2010:1–22, 2010.

[60] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. *Technical Report; Department of Computer Science and Infor-*

mation Engineering, National Taiwan University: Taipei, Taiwan, 2010. Available at http://www.csie.ntu.edu.tw/ cjlin/papers/guide/guide.pdf, 2003.

[61] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A survey on visual surveillance of object motion and behaviors. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(3):334–352, 2004.

[62] Y.H. Hu and D. Li. Energy based collaborative source localization using acoustic micro-sensor array. In *Multimedia Signal Processing, 2002 IEEE Workshop on*, pages 371–375. IEEE, 2002.

[63] M Israel, G Schlagenhauf, A Fackelmeier, P Haschberger, DLR Oberpfaffenhofen, CLAAS Saulgau GmbH, and TU München. Study on wildlife detection during pasture mowing. *http://elib.dlr.de/65977/1/WildretterVDIv4.pdf*.

[64] Martin Israel. A uav-based roe deer fawn detection system. pages 1–5, 2011.

[65] Martin Israel, Stephan Evers, Deutsches Zentrum für Luft, and Oberpfaffenhofen Raumfahrt. Mustererkennung zur detektion von rehkitzen in thermal-bildern. In *Proceedings of the 17. Workshop in Computerised Image Analysis in Agriculture*, pages 20–28, 2011.

[66] G Jahns. Call recognition to identify cow conditions—a call-recogniser translating calls to text. *Computers and electronics in agriculture*, 62(1):54–58, 2008.

[67] Anders Jarnemo. Roe deer capreolus capreolus fawns and mowing-mortality rates and countermeasures. *Wildlife Biology*, 8(3):211–218, 2002.

[68] Lance M Kaplan, Qiang Le, and N Molnar. Maximum likelihood methods for bearings-only target localization. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 5, pages 3001–3004. IEEE, 2001.

[69] T Kirubarajan, Y Bar-Shalom, KR Pattipati, and I Kadar. Ground target tracking with variable structure imm estimator. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(1):26–46, 2000.

[70] Josef Kittler, Mohamad Hatef, Robert PW Duin, and Jiri Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.

[71] L Kittler. Wildverluste durch den einsatz landwirtschaftlicher maschinen nach einer erhebung aus dem jagdjahr 1976/77 in nordrhein-westfalen. *Zeitschrift für Jagdwissenschaft*, 25(1):22–32, 1979.

[72] Ann E Koehler, Rex E Marsh, and Terrell P Salmon. Frightening methods and devices/stimuli to prevent mammal damage–a review. In *Proceedings of the 14th Verbrate Pest Conference*, pages 168–173. University of California, 1990.

[73] Joseph A Kogan and Daniel Margoliash. Automated recognition of bird song elements from contionous recordings using dynamic time warping and hidden Markov models: A comparative study. *America*, 103(4):2185–2196, 1998.

[74] Stephen J Krotosky and Mohan M Trivedi. On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *Intelligent Transportation Systems, IEEE Transactions on*, 8(4):619–629, 2007.

[75] Stephen J Krotosky and Mohan M Trivedi. Person surveillance using visual and infrared imagery. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(8):1096–1105, 2008.

[76] Chang-Hsing Lee, Chih-Hsun Chou, Chin-Chuan Han, and Ren-Zhuang Huang. Automatic recognition of animal vocalizations using averaged mfcc and linear discriminant analysis. *Pattern Recognition Letters*, 27(2):93–101, 2006.

[77] Alex Leykin, Yang Ran, and Riad Hammoud. Thermal-visible video fusion for moving target tracking and pedestrian classification. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[78] Jianfu Li, Weiguo Gong, Weihong Li, and Xiaoying Liu. Robust pedestrian detection in thermal infrared imagery using the wavelet transform. *Infrared Physics & Technology*, 53(4):267–273, 2010.

[79] Yun Li, KC Ho, and Mihail Popescu. A microphone array system for automatic fall detection. *Biomedical Engineering, IEEE Transactions on*, 59(5):1291–1301, 2012.

[80] Derek R Magee and Roger D Boyle. Detecting lameness using 're-sampling condensation'and 'multi-stream cyclic hidden markov models'. *Image and Vision Computing*, 20(8):581–594, 2002.

[81] Gerhard Manteuffel, Birger Puppe, and Peter C Schön. Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, 88(1):163–182, 2004.

[82] Gerhard Manteuffel and Peter Christian Schön. Measuring pig welfare by automatic monitoring of stress calls. *Agrartechische Berichte*, 29:110–118, 2002.

[83] Paula Martiskainen, Mikko Järvinen, Jukka-Pekka Skön, Jarkko Tiirikainen, Mikko Kolehmainen, and Jaakko Mononen. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied animal behaviour science*, 119(1):32–38, 2009.

[84] Maja Matetic and Slobodan Ribaric. Qualitative modelling of object behaviour in the dynamic vision system using hidden markov models. In *Electrotechnical Conference, 2002. MELECON 2002. 11th Mediterranean*, pages 570–574. IEEE, 2002.

[85] Efim Mazor, Amir Averbuch, Yakov Bar-Shalom, and Joshua Dayan. Interacting multiple model methods in target tracking: a survey. *Aerospace and Electronic Systems, IEEE Transactions on*, 34(1):103–123, 1998.

[86] Gérard Medioni, Isaac Cohen, François Brémond, Somboon Hongeng, and Ramakant Nevatia. Event detection and analysis from video streams. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):873–889, 2001.

[87] Chartchai Meesookho, Urbashi Mitra, Shrikanth Narayanan, and Senior Member. On Energy-Based Acoustic Source Localization for Sensor Networks. *IEEE Transactions on Signal Processing*, 56(1):365–377, 2008.

[88] Terry A Messmer. The emergence of human-wildlife conflict management : turning challenges into opportunities. *International Biodeterioration & Biodegradation*, 45:97–102, 2000.

[89] Alexander Mielke and Klaus Zuberbühler. A method for automated individual, species and call type recognition in free-ranging animals. *Animal Behaviour*, 86(2):475–482, 2013.

[90] Joaquín Míguez and Antonio Artes-Rodriguez. Particle filtering algorithms for tracking a maneuvering target using a network of wireless dynamic sensors. *EURASIP Journal on Applied Signal Processing*, 2006:1–16, 2006.

[91] Joaquín Míguez, Mónica F Bugallo, and Petar M Djurić. A new class of particle filters for random dynamic systems with unknown statistics. *EURASIP Journal on Applied Signal Processing*, 2004:2278–2294, 2004.

[92] Zvezdelin B Milanov. Effect of mowing fodder plants on small game populations in central bulgaria. In *Proceedings of the International Union of Game Biologist XXII Congress: The Game and the Man*, pages 394–397, 1996.

[93] Thomas B Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2):90–126, 2006.

[94] Pedro J Moreno and Shivani Agarwal. An experimental study of em-based algorithms for semi-supervised learning in audio classification. In *ICML 2003 Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, pages 19–25, 2003.

[95] Donald F Mott and Shirley K Timbrook. Alleviating nuisance canada goose problems with acoustical stimuli. In *Proceedings of the 13th Verbrate Pest Conference*, pages 301–304. University of Nebraska, 1988.

[96] Daniella Jorge de Moura, Irenilza de Alencar Nääs, Elaine Cangussu de Souza Alves, Thayla Morandi Ridolfi de Carvalho, Marcos Martinez do Vale, and Karla Andrea Oliveira de Lima. Noise analysis to evaluate chick thermal comfort. *Scientia Agricola*, 65(4):438–443, 2008.

[97] DJ Moura, WT Silva, IA Naas, YA Tolón, KAO Lima, and MM Vale. Real time computer stress monitoring of piglets using vocalization analysis. *Computers and Electronics in Agriculture*, 64(1):11–18, 2008.

[98] Roger Müller and Lars Schrader. A new method to measure behavioural activity levels in dairy cows. *Applied Animal Behaviour Science*, 83(4):247–258, 2003.

[99] ES Nadimi, Henning Tangen Søgaard, and Thomas Bak. Zigbee-based wireless sensor networks for classifying the behaviour of a herd of animals using classification trees. *Biosystems engineering*, 100(2):167–176, 2008.

[100] L. Naef-Daenzer. Scaring of carrion crows (corvus corone corone) by species-specific distress calls and suspended bodies of dead crows. In *Proc. 9th Bird Control Seminar*, pages 91–93. Bowling Green State University, 1983.

[101] VPS Naidu, G Girija, and N Santhakumar. Three model imm-ekf for tracking targets executing evasive maneuvers. In *AIAA Aerospace Sciences Meeting and Exhibit*, volume 45, pages 1–22, 2007.

[102] Harsh Nanda and Larry Davis. Probabilistic template based pedestrian detection in infrared videos. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 1, pages 15–20. IEEE, 2002.

[103] DL Nolte. Behavioral approaches for limiting depredation by wild ungulates. *Grazing behavior of livestock and wildlife*, pages 60–69, 1999.

[104] John C Platt, Nello Cristianini, and John Shawe-taylor. Large margin dags for multiclass classification. In *Advances in Neural Information Processing Systems*, volume 12, pages 547–553. Cambridge, MA: MIT Press, 2000.

[105] Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[106] Gerasimos Potamianos, Chalapathy Neti, Juergen Luettin, and Iain Matthews. Audio-visual automatic speech recognition: An overview. In *Issues in visual and audio-visual speech processing*, pages 1–23. MIT Press Cambridge, MA, 2004.

[107] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286, 1989.

[108] TROY M Radtke and CHARLES D Dieter. Canada goose crop damage abatement in south dakota. *Human-Wildlife Interactions*, 5(2):315–320, 2011.

[109] Javier Ramirez, Juan Manuel Górriz, and José Carlos Segura. Voice activity detection. fundamentals and speech recognition system robustness. *Robust Speech Recognition and Understanding*, 6(9):1–22, 2007.

[110] Douglas K Reiter, Mark W Brunson, and Robert H Schmidt. Public attitudes toward wildlife damage management and policy. *The Wildlife Society Bulletin*, 27:746–758, 1999.

[111] Douglas A Reynolds, Thomas F Quatieri, and Robert B Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41, 2000.

[112] Piotr Rudol and Patrick Doherty. Human body detection and geolocalization for uav search and rescue missions using color and thermal imagery. In *Aerospace Conference, 2008 IEEE*, pages 1–8. IEEE, 2008.

[113] Stuart Jonathan Russell, Peter Norvig, John F Canny, Jitendra M Malik, and Douglas D Edwards. *Artificial intelligence: a modern approach*, volume 2. Prentice hall Englewood Cliffs, 1995.

[114] AL Schaefer, N Cook, SV Tessaro, D Deregt, G Desroches, PL Dubeski, AKW Tong, and DL Godson. Early detection and prediction of infection using infrared thermography. *Canadian journal of animal science*, 84(1):73–80, 2004.

[115] Peter-Christian Schön, Birger Puppe, and Gerhard Manteuffel. Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (sus scrofa). *The Journal of the Acoustical Society of America*, 110(3):1425–1431, 2001.

[116] Juan Serrano-Cuerda, María T Lopez, and Antonio Fernández-Caballero. Robust human detection and tracking in intelligent environments by information fusion of color and infrared video. In *Intelligent Environments (IE), 2011 7th International Conference on*, pages 354–357. IEEE, 2011.

[117] Raymond A Serway, Robert J Beichner, and John W Jewett. Physics for scientists and engineers with modern physics, 6th ed. 2004.

[118] Xiaohong Sheng and Yu-Hen Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *Signal Processing, IEEE Transactions on*, 53(1):44–53, 2005.

[119] DC Slaughter, DK Giles, and D Downey. Autonomous robotic weed control systems: A review. *Computers and electronics in agriculture*, 61(1):63–78, 2008.

[120] Julius Smith and Jonathan Abel. Closed-form least-squares source location estimation from range-difference measurements. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(12):1661–1669, 1987.

[121] John V Stafford. Implementing precision agriculture in the 21st century. *Journal of Agricultural Engineering Research*, 76(3):267–275, 2000.

[122] Gwen R Stevens, Jamie Rogue, Richard Weber, and Larry Clark. Evaluation of a radar-activated, demand-performance bird hazing system. *International biodeterioration & biodegradation*, 45(3):129–137, 2000.

[123] FLIR Commercial Vision Systems. Thermal imaging: how far can you see with it? In *Technical Note*, pages 1–4.

[124] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 4th edition, 2008.

[125] Rebecca E Thomas, Kurt M Fristrup, and Peter L Tyack. Linking the sounds of dolphins to their locations and behavior using video and multichannel acoustic recordings. *The Journal of the Acoustical Society of America*, 112(4):1692–1701, 2002.

[126] RD Thompson, CV Grant, EW Pearson, and GW Corner. Differential heart rate response of starlings to sound stimuli of biological origin. *The Journal of Wildlife Management*, pages 888–893, 1968.

[127] RD Tillett, CM Onyango, and JA Marchant. Using model-based image processing to track animal movements. *Computers and electronics in agriculture*, 17(2):249–261, 1997.

[128] Helene Torresan, Benoit Turgeon, Clemente Ibarra-Castanedo, Patrick Hebert, and Xavier P Maldague. Advanced surveillance systems: combining video and thermal imagery for pedestrian detection. In *Defense and Security*, pages 506–515. International Society for Optics and Photonics, 2004.

[129] Vlad M Trifa, Alexander NG Kirschel, Charles E Taylor, and Edgar E Vallejo. Automated species recognition of antbirds in a mexican rainforest using hidden markov models. *The Journal of the Acoustical Society of America*, 123(4):2424–2431, 2008.

[130] Hemant Tyagi, Rajesh M Hegde, Hema A Murthy, and Anil Prabhakar. Automatic identification of bird calls using spectral ensemble average voice prints. In *Proceedings of the Thirteenth European Signal Processing Conference*, pages 1–5, 2006.

[131] Eugene David Ungar and Steven Mark Rutter. Classifying cattle jaw movements: comparing iger behaviour recorder and acoustic techniques. *Applied animal behaviour science*, 98(1):11–27, 2006.

[132] Edgar E Vallejo and Charles E Taylor. Adaptive sensor arrays for acoustic monitoring of bird behavior and diversity: preliminary results on source identification using support vector machines. *Artificial Life and Robotics*, 14(4):485–488, 2009.

[133] J van Eerbeek. Effectivity of dutch goose management during the breeding season. Master's thesis, Master thesis Animal Ecology & Evolution, University of Groningen, 2013.

[134] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

[135] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern recognition*, 36(3):585–601, 2003.

[136] Philip Whitford. Successful use of alarm and alert calls to reduce emerging crop damage by resident canada geese near horicon marsh, wisconsin. In *Proc. 23rd Vertebr. Pest Conf.*, pages 74–79. Univ. of Calif., Davis, 2008.

[137] Philip C. Whitford and Luke A. Streng. Efficacy of recorded alarm and alert calls for canada goose dispersal. In *Journal of Wildlife Management*, pages 1–21, 2003.

[138] Xinwei Xue and Thomas C Henderson. Video-based animal behavior analysis from multiple cameras. In *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*, pages 335–340. IEEE, 2006.

[139] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, 2009.

[140] Debao Zhou, Jingzhou Wang, and Shufang Wang. Countour based hog deer detection in thermal images for traffic safety. In *Proceedings of the International Conference on Image Processing,Computer Vision, and Pattern Recognition*, pages 1–6, 2012.