WILDLIFE COMMUNICATION

AARHUS
UNIVERSITY
DEPARTMENT OF ENGINEERING

DATA SHEET

Title: Wildlife Communication
Subtitle: Electrical and Computer Engineering
Series title and no.: Technical report ECE-TR-7

Authors: Kim Arild Steen[1], Ole Roland Therkildsen[2], Henrik Karstoft[1] and Ole Green[1]
[1]Department of Engineering, Electrical and Computer Engineering, Aarhus University
[2]Department of Bioscience, Aarhus University

Abstract: This report contains a progress report for the ph.d. project titled "Wildlife Communication". The project focuses on investigating how signal processing and pattern recognition can be used to improve wildlife management in agriculture. Wildlife management systems used today experience habituation from wild animals which makes them ineffective. An intelligent wildlife management system could monitor its own effectiveness and alter its scaring strategy based on this.

Keywords: digital signal processing, pattern recognition, wildlife management, audio processing, video processing

Supervisors: Ole Green and Henrik Karstoft

Cover image: Kim Arild Steen

WILDLIFE COMMUNICATION

Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft, and Ole Green

Aarhus University, Department of Engineering

Abstract

This report contains a progress report for the ph.d. project titled "Wildlife Communication". The project focuses on investigating how signal processing and pattern recognition can be used to improve wildlife management in agriculture. Wildlife management systems used today experience habituation from wild animals which makes them ineffective. An intelligent wildlife management system could monitor its own effectiveness and alter its scaring strategy based on this.

Table of Contents

Chapter 7   Summary   29

Bibliography   31

1

Introduction

This chapter is a short introduction to the PhD work and the progress report.

1.1 Purpose of the PhD

The purpose of the PhD is to develop methods and algorithms for automation in agriculture. The main focus is wildlife management systems, where new technology could automate or improve existing methods.

The main contribution of the PhD work is the development of pattern recognition algorithms, which are capable of detecting and recognizing wildlife in an agricultural setting. This includes audio and video based systems, who are capable of measuring the presence and behaviour of wildlife.

The expected result of the PhD work is a proof of concept solution, which can be used in further development and fullscale tests.

1.2 Hypothesis and objectives

The hypothesis of the PhD is

Wildlife management can be performed in a more ethical and wildlife friendly manner, if based on new sensor technology, signal processing and automation within tools and methods for wildlife management

Based on the sensor technology, proposed to be used in this project, and the hypothesis, the objectives of the PhD work are:

- Develop algorithm for automatic recognition of wildlife presence and behaviour based on acoustic measurements

- Develop algorithm for automatic recognition of wildlife presence and behaviour based on video recordings

- Develop methods for wildlife communication based on the algorithms above

- Investigate the effect of the communication, with respect to wildlife management

1.3   Structure of the progress report

The structure of the progress report is as follows:

Chapter 1: Short introduction to the PhD

Chapter 2: An introduction to current wildlife management strategies

Chapter 3: An introduction to wildlife communication and audio and video based behaviour recognition

Chapter 4: *A Multimedia Capture System for Wildlife Studies*, part of a conference paper from *The Third International Conference on Emerging Network Intelligence*, Lisbon, Portugal. The paper presents the hardware and software setup used to record wildlife geese.

Chapter 5: *A Vocal based Analytical Method for Goose Behaviour Recognition*, part of a journal paper from *Sensors*, presenting an algorithm for automatic recognition of goose behaviours based on their vocalizations.

Chapter 6: *Automatic Detection of Animals in Mowing Operations using Thermal Cameras*, a journal paper, as submitted[1] to *Sensors*, presenting a novel approach for automatic detection of wildlife animals during mowing operations.

Chapter 7: Summary and future plans of the PhD

---

[1]Abstract and keywords excluded

2

Wildlife management

Human-wildlife conflicts are increasing [1] resulting damages to both the animals as well as human activities. There are many areas in which particular groups of animals are unwanted (e.g. airports, agricultural fields, cities and sport resorts) and a wide range of devices to detect and deter animals, causing conflict, are used in wildlife damage management, although their effectiveness is often highly variable [2].

2.1 Wild geese

In the last couple of years, the population of many species of wild geese have grown exponentially [3]. This growth in population affects agriculture as human-wildlife conflicts are inevitable, as the habitat availability decreases. Also, in Svalbard, grazing Pink-footed and Barnacle Geese can cause changes in plant community structure and increase carbon dioxide release [3].

Highly attractive buffer areas with improved food quality/quantity and reduced disturbance levels, and in [4] it is reported, that a total of 500.000 €was paid for preventing and compensating crop damage caused by protected birds[1] in Sweden.

Present scaring devices are often activated electronically, through detection of motion and/or body heat (e.g. infrared sensors, [2]). In most cases scaring devices are non-specific, so they can be activated by any animal, not only when individuals of the target species enters the area. This increases the risk of habituation, which is often the major limitation on the use of scaring devices [5]. Although random or animal-activated scaring devices may reduce habituation and prolong the protection period over non-random devices [5].

The use of distress calls as scaring stimuli have shown good results, however habituation have been reported [2]. In spite of this, there are several advantages of bioacoustics as alarm and distress calls are meaningful to animals at low intensities, which makes it not necessary to produce loud noises to frigthen the animals. This has an econonmic advantage, as it can be expensive to produce loud noises[2]. Furthermore loud noises are disturbing to humans and other animals.

---

[1]Species which cannot be hunted during the periods when they cause damage

[2]Amplifiers are needed to be able to produce loud sound. Furthermore, many speakers as the animals populate wide areas

2.2   Roe deers

A more deadly human-wildlife conflict exist in harvesting and mowing operations, where roe deer fawns are killed or deadly injured every year. They perceive the farming machinery as predators and therefore lay low and still in the dense vegetation to avoid being found, which makes it hard for the farmer to detect the animals during farming operations.

It is not only roe deer fawns that experience the deadly confrontation with farming machinery, as both speed and working widths have increased during the last decades, making it difficult even for mobile animals, as the roe deer, to escape the machinery in time.

However it is not only an animal welfare issue, as fodder may be contaminated with carcasses of the dead animals which may impose health hazards for live stock from infection by the bacteria *Clostridium botulinum*, causing botulism. Another issue, which has only received little attention, is the mental stress imposed on the farmers, who face an injured animal during farming operations, and needs to to perform mercy killings without professional expertise or the correct tools.

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 [6, 7]) or strategy (e.g. leaving edge strips), longer mowing intervals, reduction of speed or higher cutting height [6, 8] 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 [6] or plastic sacks set out on poles before mowing [9] have been reported to reduce wildlife mortality.

However, these methods 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 [10].

2.3   Hares

Like roe deers, hares are also being killed in farming operations. Because of their size, they are even harder to detect for the farmer, resulting in thousands of dead brown hares every year [11, 9]. Some studies suggest that the mortality during harvesting operations is of minor importance, or have no significance on recruitment in hares [12, 13], however there are still risks of fodder contamination.

3

Wildlife communication

The task of wildlife communication is about providing a stimuli capable of altering animal behaviour. This requires a system capable of measuring the behaviour of animals and responding to this. This approach fits within the framework of an intellingent agent. An intelligent agent is a system or program that is able to get input from an enviroment through sensors and react through actuators (figure 3.1). In wildlife communication, the sensors could be based on many known technologies including: accelerometers,wirelss sensor networks, gps, microphones and cameras [14, 15, 16, 17, 18, 19]. This project concerns wildlife animals, and therefore I have limited the sensors to microphones and cameras, which does not need to be attached to the animals, as this is not possible.

The actuators could also include many known technologies: light sources, speakers, electronic shock devices etc., however the use of a sound source is the main focus within this PhD, as the use of recorded distress calls have proven effective in wildlife management systems [2].

The main workload of the PhD is the development of the model, which enables the correct reaction to the inputs from the sensors. Both within audio and video processing concerning behaviour recognition, the prefered framework for such models are pattern recognition algorithms, which is also the chosen framework for this PhD.
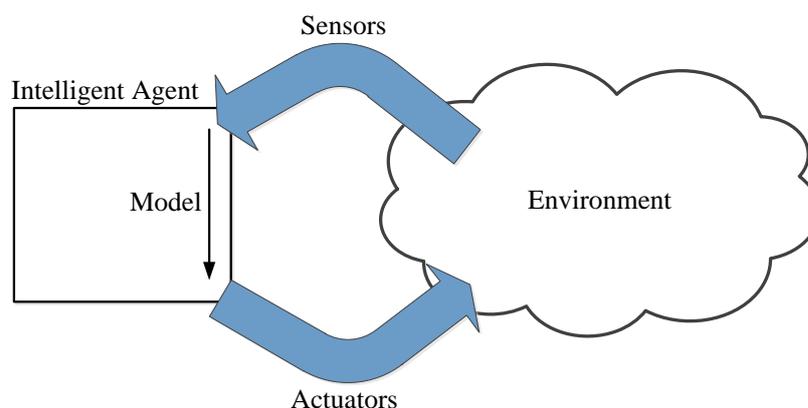
Figure 3.1: An intelligent agent reacts to the environment via actuators based on input from sensors. The intelligence within the agent is based on the program or model in the agent

3.1 Pattern recognition

The program or model within an intelligent often consist of some sort of recognition based on the inputs from the sensors. Pattern recognition algorithms is a class of algorithms that try to make sense of data given to the algorithm. This data could be text, financial data, DNA-strings, audio etc.. In pattern recognition, the data, independent of the origin, is represented in a space (vector space, pixels, dissimilarity space) where math can be applied (figure 3.2). The task of an intelligent agent is to transform the sensor input to a decision, which involves a recognition or classification of the meaning of the data.
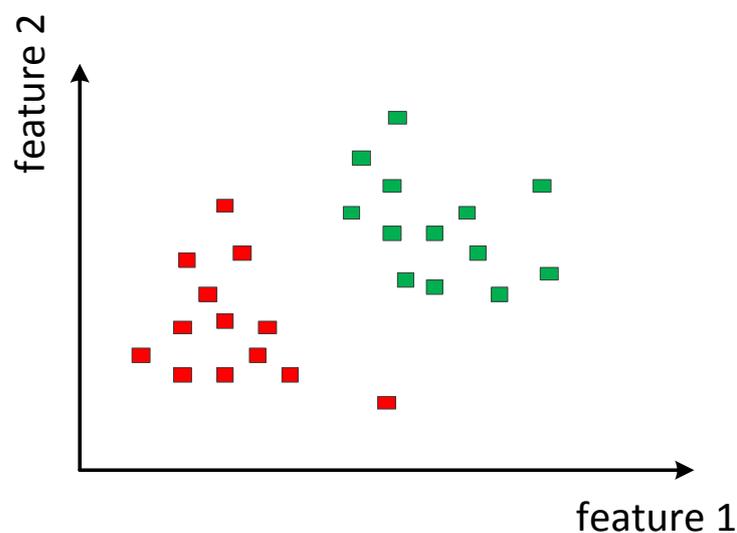
Figure 3.2: In pattern recognition, the data is represented in a feature space making it possible to apply math to a classification problem. The colours indicate two different classes, which could be two different characters in the OCR application

Pattern recognition has been applied to many realworld problems, where computer programs or machines tries to make sense of realworld data, with the most popular being: OCR (Optical Character Recognition), Face Recognition and ASR (Automatic Speech Recognition). These applications includes different algorithms within the framework of pattern recognition.

3.2 Animal behaviour recognition

A certain behaviour is a mixture of responses to internal and external stimuli, and a full description of behaviour would include internal as well as external responses [20]. Therefore, the task of measuring behaviour can be a difficult one.

One way of describing animal behaviour is through ethograms, which is a list of natural behaviours for a particular animal. An ethogram can contain a description of the type of behaviour

(food related, social, aggressive etc.), the behaviour (eat, sleep, play, fight, flee etc.) and a written description of the behaviour. This description can include visual, auditory or other features which are specific to the type of behaviour. The ethogram can drive the task of choosing recording equipment for measuring the animal behaviour.

Another way of measuring animal behaviour is through expert knowledge as in [21]. Here an expert in animal behaviour observes, and characterises, a video sequence with a characteristic behaviour as a part of the supervised learning. In this approach, the expert base the behaviour on the spatial and temporal data provided by the video recording.

In both cases visual and/or auditory information can form the basis for behaviour analysis and recognition.

3.2.1 Acoustic based recognition of behaviour

Most research within automatic recognition within bioacoustics focus on species recognition or individual animal recognition [22, 23, 24]. Not much research has been conducted within automatic behaviour recognition based on animal vocalization. However in [25] a system capable of recognizing pig welfare based on their vocalizations is presented, and in [18] they use automatic recognition of individual bird as an input to behaviour research.

The research within animal vocalization recognition is highly influenced by research within automatic speech recognition. Both the same features and pattern recognition algorithms are applied to animal vocalization recognition.

In human speech recognition, the features proven most succesful are based on the hearing capabilities of humans [26]. The feature extraction is carried out using Mel-Frequency Cepstral Coefficient (MFCC), which are derived from the mel-scale, which is a non-linear frequency mapping adjusted to human hearing capabilities. A mel is a unit of measure of perceived pitch or frequency of a tone. In [27] an approximation is given by

$$F_{mel} = \frac{1000}{log(2)} log \left(1 + \frac{F_{Hz}}{1000} \right) \tag{3.1}$$
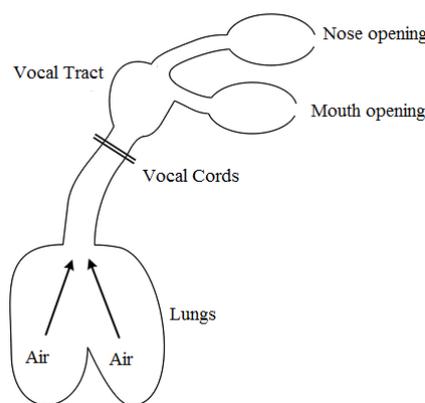
Figure 3.3: A simplified model of human speech production (illustration is based on similar illustration in [28])

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 [26]. The number

of filters in the filter bank depend on the application, and various implementations of MFCC feature extraction have been used in speech recognition tasks [29]. Most research within recognition of animal vocalization are based on this approach [30, 22, 23, 24, 31, 32], however also Linear Predictive Coding features are also popular. These are derived from human speech production (figure 3.3), where the speech production is parameterized by filter and it's coefficients [33, 34, 25, 35].

In acoustic based recognition, the pattern recognition algorithm of choice is Hidden Markov Models (HMM), which are able to model both the stochastic and temporal variablity in animal vocalizations [36, 37, 24, 31, 35], however Gaussian Mixture Models (GMM), Dynamic Time Warping (DTW), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have also been used to some extent in litterature [36, 30, 23, 31, 32, 18]. So far the work in this PhD has focused on SVMs, which is presented in chapter 4.

3.2.2 Visual recognition of behaviour

Visual based behaviour recognition follow the same principles as acoustic based recognition; namely a feature extraction process and a classification based on a pattern recognition algorithm. As visual interpretation of behaviour is based on the movement or change of posture of the animals, these are often used as features for video based behaviour recognition.

For detection of movement, visual tracking algorithms are applied. Several methods for visual tracking exist in litterature, and it is often application specific. However, a common framework for motion tracking are the Kalman filter [38] and Bayesian approaches [19]. These are based on detection of the animals, and the tracking is performed by a prediction and update scheme, where the motion of the animals are governed by a dynamic model.

However other tracking related methods are also applied to behaviour recognition, such as background subtraction and optical flow. In background subtraction, the movement of the animals are found by subtracting the static background and thereby emphasize the motion of the animals. In optical flow, the flow of pixel intensities are used to describe the apparent motion within the video, and this method is widely used when the goal is to track a flock of animals or a crowds [39, 40, 41].

When single animal behaviour are to be recognized, the shape or posture of the animal has been used as features for recognition. In [19] the shape of cow legs are used to detect lameness, and in [20] the shape of a pig i used to track it's movement within the pen.

The recognition of the specific behaviours is often carried out through pattern recognition algorithms such as HMMs, ANNs, Decision trees or Rule-Based schemes [42, 38, 43]. The choice of algorithm is dependent on how the behaviour is measured through feature extraction (e.g. Kalman filtering or optical flow), and is therefore application specific, however HMMs have been extensively used behaviour recognition research [42, 44, 19].

4

A Multimedia Capture System for Wildlife Studies

The following chapter is part of a conference paper, presented at *The Third International Conference on Emerging Network Intelligence*, Lisbon, Portugal. The paper presents the hardware and software setup used to record wildlife geese [45].

4.1 Introduction

In modern society, we often experience unwanted encounters between groups of animals and human activities, such as in agricultural fields or at airports. This can be a costly affair and often inflicts damages both to the animals as well as humans. In the case of agricultural fields, visual and acoustic stimuli may be used as mechanisms for scaring away unwanted animals. However, these methods often have limited success rates, as the animals habituate to the stimulus [2].

Recently, computer technology has been applied for characterising animal behaviour using computer vision for tracking animal trajectories [21][46] and audio processing for recognition of animal vocalizations [23][35][33]. These approaches may lead to systems capable of recognizing specific species and behaviours, and scare off the animals before they inflict damage or get hurt. In [2] different approaches to scaring off animals is reviewed, such as guard animals, gas exploders and distress calls, with the latter showing good results.

In the process of linking wildlife animals vocalizations with specific behaviour, a system for recording video and audio data in a wildlife setting is presented. Wildlife surveillance systems have been previously described [47][48]; however, these systems are designed for specific scenarios. Likewise, the present system is specific to the context of video and audio recording of wildlife birds foraging in agricultural fields. The system described in [47] support both audio and video recording in a harsh environment (humid environment), however the data recorded is used for manual inspection and not research regarding automatic recognition.

The system has been used for video and audio recording of wildlife geese foraging in agricultural fields. The main purpose of this system is to record and store images and audio of geese as they land, eat and flee. The data provided by the system will be used in further research regarding automatic recognition of geese behaviour. Geese were chosen in this study, as they inflict much damage in agriculture, and they are very vocal.

4.2 System requirements

Agricultural fields are wide open spaces, implying windy conditions during wildlife recordings. Wind reduction is therefore necessary to preserve the quality of the audio recordings, and can be accomplished for instance through use of a casing. Furthermore, the remote location of agricultural fields reduces access to power grids. Consequently the consideration of a power source and power consumption is important, as the system requires a standalone power source.

Barnacle geese are highly mobile with flight speed up to $20\frac{m}{s}$[49], and the video recording equipment needs to provide adequate frame rates to capture their movements. It is not desirable to reduce the image quality or add computations by adding compression, as this could degrade performance of later image processing and potentially cause fluctating frame rates, as compression time could be affected by information in the images.

As it is impossible to pinpoint in advance where the geese will land and eat, video and audio recordings need to be inspected during the study. Consequently remote access is an important system requirement, to avoid frightening the geese during inspection.

A further system requirement is a minimum uptime long enough to capture video and audio as the geese return to the location, to avoid interference caused by installation of the system. An uptime of 36 hours has been chosen, as the geese are likely to return to the same location because of the availability of food, however it is not certain that they will return the same day. Therefore, a harddrive with a large enough capacitiy, must be chosen to ensure no loss of data.

To summarize the most important requirements for a multimedia capture device for wildlife studies:

- Reduction of wind noise in the audio recording equipment

- Standalone power source (limitations to power consumption)

- Adequate frame rates ($20 - 30$ frames per second (fps)), due to the mobile animals

- Remote access, to monitor the recording without scaring off the animals

- Minimum uptime of 36 hours

- High harddrive capacity (> 1 TB), due to the long uptime and no compression

4.3 System description

The requirements, specified above, led to the system setup, described in this section and illustrated in Figure 4.1.

The power source needs to be standalone, and two soultions were considered: car batteries and solar panels. The power produced by a solar panel is dependent on the weather, and, as sunlight is not garanteed on the west coast of Denmark, risks downtime. Another risk of solar panels is to scare off or interfere with the animals' behaviour, because of their shiny surface. Car batteries, were therefore chosen as the power source, as they are reliable, however they eventually run out and need to be replaced and charged. To aviod unnecassary power consumption, the system is set to stand-by during the night and automatically restarted the next morning.

The lifetime of average car batteries are highly reduced when they are drained, which would be the case in the system setup. Deep cycling batteries are therefore preferable, as they are designed to cope with this kind of treatment.

The system is a work in progress, and it was chosen to use DC/AC converters, as a part of the power source, for more flexibility. This ensures easy expansion if other equipment were to be used at a later stage, however it also introduces a loss in efficiency. The chosen converter has an efficiency of 90%.

The overall power consumption of the system is 60 W, and with a 90% efficiency. The minimum uptime must be 36 hours, which requires batteries of approximately 200 Ah (two 95 Ah were chosen), however this is derived from the worst case power consumption scenario and without the planned stand-by hours.

To preserve quality in the audio recording a directional shotgun microphone, with wind reduction filter was chosen. For the connection, a 10 m long multiple shielded audio extension cable is being used, which enables different placement of the microphone.

The high frame rates are provided by the chosen camera, which enables $20 - 30$ fps depending on the resolution of the image. The camera uses a global shutter, which reduces blurring caused by movements. It is powered via the USB connection, which is also used for data transfer. The recorded images are not compressed, which requires a high capacity harddrive. The SSD technology would be preferable, because of the low power consumption, however due to dollar/GB, this was not chosen for the system.

For remote connection, a 3G connection was chosen. Due to lack of coverage, this solution can potentially lead to loss of connection, however the location chosen for the recording had good 3G coverage. A lack of coverage would not be vital for the recordings, however remote access would be affected. A list of the specific items used in this setup can be seen in table 3.1.

System components	
Component	**Details**
Battery	12 V Deep Cycling
DC/AC converter	Sine wave converter
uEye Camera UI-1245LE-C	Lens: 6 mm 640 x 480
Harddrive	3 TB
3G connection	5 MB
Sennheiser MKE 400 Microphone	Shotgun
Asus Eee Laptop	1.6 Ghz 1 Gb memory

Table 4.1: Table of system components used in the setup
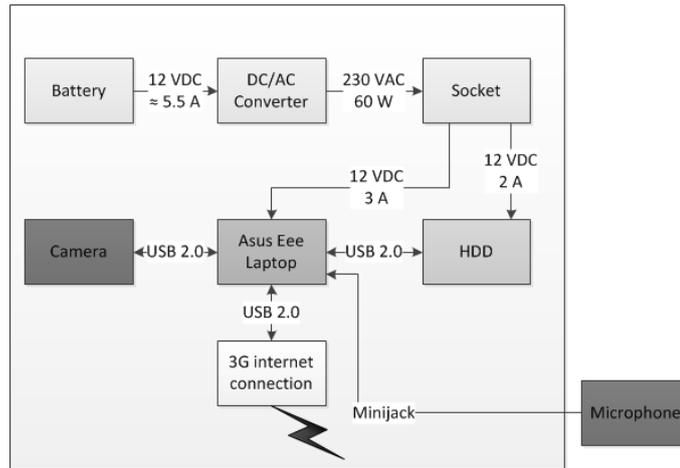
Figure 4.1: Block diagram of the system

4.4 Infrastructure and data description

The main purpose of the system is to record and store large amounts of data. In Figure 4.2 an overview of the data flow and connections are shown. With a frame rate of 20 fps, an image is captured from the camera and stored on the external harddrive. Meanwhile an audio file is saved on the harddrive every 5 minutes. This is accomplished by a loop-recording software, which increments filenames and records while storing the files. The audio recordings were done with a sample rate of $44.1 kHz$ and 16 bit resolution, which is the default settings of the loop-recording software.

The images captured from the camera, are stored as the raw bayer pattern. This reduced the file size (from 900 kb to 300 kb) and the CPU load as image encoding is not being done. The demosaic and encoding of the captured images is done offline in the analysis fase of the research.

The USB 2.0 protocol used for data transfer offers a theoretical maximum rate at 60 MB/s. The image capture requires 6 MB/s, which lies within the specifications. The audio file is saved every 5 minutes, and does not affect the ongoing audio recording. This means that a transfer rate of approximately $\frac{50}{(60 \cdot 5)} \approx 0.2$ MB/s would be sufficient for storing the audio.

The 3G internet connection is used for remote access and uploading files to an FTP-server. The purpose of the file transfer is to monitor the video recording, and as the images are not encoded it is not possible to view the images on the surveillance system laptop. The newest captured image is being uploaded every hour, and accessed from another laptop in the laboratory.

The dataflow and software considerations are summarized here:

- An image is captured every 1/20 second and saved on the external hdd

- Every five minutes an audio file (.wav file) is saved on the external hdd using loop recorder (see www.looprecorder.de)

- Every hour a batch script uploads the newest image to an ftp server

- At sunset, the system is set to stand-by, and at sunrise the system wakes up again (3G connection is automically started to enable remote access)

With a frame rate of 20 fps and chosen audio encoding (.wav files), the system records a data rate of approximately 22 GB/hour.
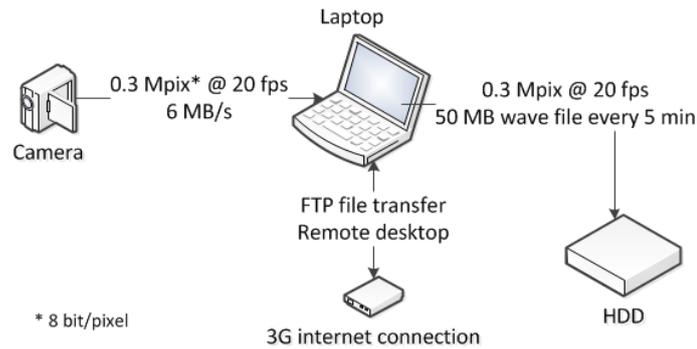
Figure 4.2: Overview of the infrastructure of the system setup, including description of data

4.5 Conclusion

Based on the described system setup, it was possible to record geese in order to analyze the link between their vocalizations and behaviour. The geese quickly grew accustomed to the setup, and only two days after the installation of the system, the geese landed and foraged.

Data provided by the described system is a part of ongoing research to automatically recognize animal behaviour based on audio and video recordings. The results of this research are to be tested using a modfication of the descibed system, where both audio and video processing will be a part of the system.

A Vocal based Analytical Method for Goose Behaviour Recognition

The following chapter is part of a journal paper, published in *Sensors*. The paper presents an algorithm for automatic recognition of goose behaviours based on their vocalizations [50].

5.1 Introduction

In many parts of the world, damage caused by wildlife creates significant economic challenges to human communities. Since human-wildlife conflicts are increasing [1] the development of cost-effective methods for reducing damage or conflict levels is important in wildlife management. A wide range of devices to detect and deter animals causing conflict are used in wildlife damage management, although their effectiveness is often highly variable [2]. Present scaring devices are often activated electronically, through detection of motion and/or body heat (e.g. infrared sensors, [2]). In most cases scaring devices are non-specific, so they can be activated by any animal, not only when individuals of the target species enters the area. This increases the risk of habituation, which is often the major limitation on the use of scaring devices [5]. Although random or animal-activated scaring devices may reduce habituation and prolong the protection period over non-random devices [5], to our knowledge no cost-effective concept circumventing the problems of habituation has yet been developed.

For our purpose, we identified three relevant behaviours (landing, foraging and flushing), which are all accompanied by distinct vocalisations easily identified by the human ear. The vocalisations allow us to identify a flock of geese 1) attempting to land, 2) foraging or 3) being flushed. By using vocalisation recognition, we are then able to automatically detect a flock of geese attempting to land and to assess the effect of a scaring (see figure 5.1). Thereby, the concept allows us to monitor potential habituation (i.e. the situation, when geese no longer respond to scaring) and, accordingly, change our scaring strategy.

Typical methods used within animal behaviour research are based on attached tracking devices, like GPS [14] or other wireless transmitters in a wireless sensor network [15, 16], or accelerometers, measuring the movement of specific parts of the animal body [17]. Acoustic information has also been used in chewing behaviour recognition of cows [51], however these methods also rely on attaching a device on the animals. These methods are not suitable when the purpose of the animal behaviour recognition, is to utilize the results in a wildlife management system,

as it is not possible to attach these devices on the animals. Vallejo [18] uses vocalisations for source identification, based on a microphone array and thereby recognise bird behaviour, however the link between a specific vocalisation and behaviour is not found. Recognition of vocalisation, however does provide a method for behaviour recognition without the need to attach any devices on the free-living animals.

Recently, audio processing and pattern recognition methods have been used for recognition of animal vocalisations [52, 53, 34, 31] and behaviour [54, 32, 25, 55], in a controlled experiments or on single animals. This research within automatic vocalisation recognition has been highly influenced by methods conducted within human speech and speaker recognition. This includes feature extraction techniques, focused on cepstral features [56, 35] and pattern recognition algorithms such as Hidden Markov Models (HMMs) [36, 33], Gaussian Mixture Models (GMMs) [36] and Support Vector Machines (SVMs) [23, 18, 57].

The Mel Frequency Cepstral Coefficients (MFCC) have proven to be good features within human speech recognition, as they model the human perception of sound, and is therefore also widely used within animal vocalisation recogntion. However, animal sound perception may be different than human sound perception, and other features may be more suitable. In this paper, Greenwood Function Cepstral Coefficient (GFCC) features are used as features, to describe the vocalisations, as they, like MFCC, model the preception of sound, but can be adjusted to the hearing capabilities of different species [37].

The SVM is a supervised learning algorithm which can be used in both linear and non-linear pattern recognition problems [58]. The models are based on a structural risk minimisation principle, which improves the generalisation ability of the classifier [59]. Since the introduction of the model in the 1990s [60], the SVM has become a popular method of choice for many applications, including behaviour recognition, speaker identification and object recognition [61, 57, 62]. In our research, the SVM was used in a multiclass classification task to classify one of three behaviours, based on their vocalisations. The models were trained with labeled data, which were extracted from the recordings.

This paper presents a new concept for detection of animal behaviour based on its vocalisation. Methodologies developed for speech recognition have been adjusted and used to distinguish between three specific behaviours. The analytical method, described in this paper, is part of ongoing research regarding a system capable of detecting behaviour of conflict species, such as barnacle goose (*Branta leucopsis*), and adjust its scaring stimuli based on the detected behaviour in order to avoid habituation.
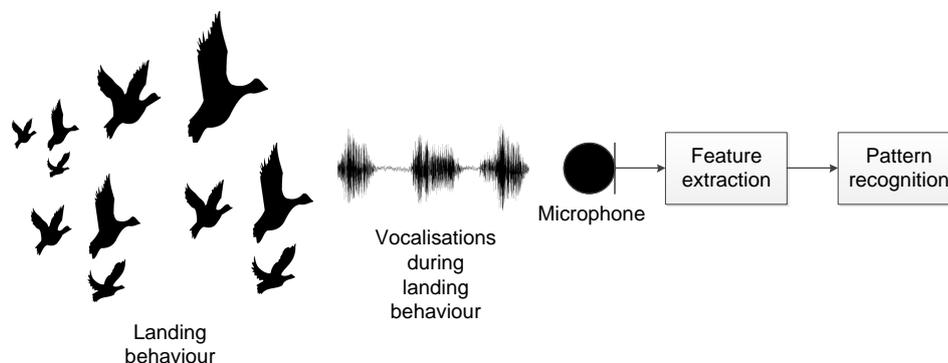
Figure 5.1: Concept of classification of landing behaviour, based on recorded vocalisations

5.2 Materials and methods

5.2.1 Acoustic feature extraction

The features used to describe animal vocalisation, in a recognition setting, are inspired by the research done within human speech and speaker recognition [36, 35]. Here cepstral coefficients, such as the MFCC, are among the most popular [63, 26].

The MFCC features are derived from the mel-scale, which is a non-linear frequency mapping adjusted to human hearing capabilities. A mel is a unit of measure of perceived pitch or frequency of a tone. In [27] an approximation is given by

$$F_{mel} = \frac{1000}{log(2)} log \left(1 + \frac{F_{Hz}}{1000} \right) \tag{5.1}$$

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 [26]. The number of filters in the filter bank depend on the application, and various implementations of MFCC feature extraction have been used in speech recognition tasks [29]. The bandwidth of these applications differ, and as barnacle geese vocalisations contain most of their spectral information in the $500 - 6000Hz$ band [64], it is comparable to the bandwidth used by Davis in their novel paper from 1980, where 20 filters are used. Therefore, 20 filters are used in the feature extraction of geese vocalisations.

These features have been shown to be useful in human speech recognition [26, 65], however animals do not perceive sounds equally as humans, which means that MFCC may not be useful for animal vocalisation feature extraction. In [37] generalized perceptual features are introduced. The feature extraction is based on the Greenwood function [66], 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 Adi use GFCC for recognition of ortolan bunting (*Emberiza Hortulana*) songs in [67]. The frequency warping function looks similar to the mel-scale warping, and the perceived frequency mapping is calculated as

$$F_p = \frac{1}{a} log_{10} \left(\frac{F_{Hz}}{A} + k \right) \tag{5.2}$$

Here the constants a, A, and k are species specific, however the constants a and A can be derived from knowing k. [68] shows that k can be approximated by a value of 0.88, which has been used in this research as well. The constants a and A can then be derived by knowing the hearing frequency range for the specific species (f_{min}, f_{max}), see (5.3) and (5.4).

$$A = \frac{f_{min}}{1 - k} \tag{5.3}$$

$$a = log_{10} \left(\frac{f_{max}}{A} + k \right) \tag{5.4}$$

The calculation of GFCC is illustrated in figure 5.2, where the incoming signal has a duration of 46 ms (2048 samples), as cepstral coefficients are derived from short-time analysis. 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 the formula

$$c_n = \sum_{k=0}^{K-1} S_k cos \left(n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right) \quad n = 0, \ldots, K - 1 \tag{5.5}$$

Here c_n is the nth cepstral coefficients and S_k is the spectral log-energy of the kth band. In this research 20 critical band filters were used, which gives a feature vector of dimension 21, as the 0th order cepstral coefficient is included (see [69]). The filters were hamming shaped, however both hanning and triangle shaped filters are often used in MFCC feature extraction [29].
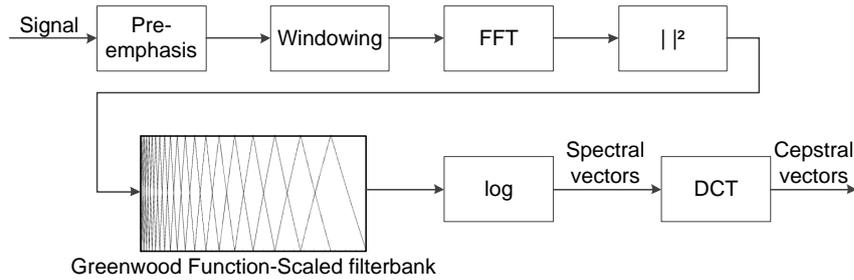
Figure 5.2: Block diagram of the acoustic feature extraction performed on the recorded vocalisations. A total of 21 features were extracted and six features were chosen based on feature selection techniques

As SVM models are based on maximizing the margin, the performance of the classifier will decrease if classes have severe overlaps. In the context of this paper, this could be the case if cepstral features does not describe the actual vocalisation, but the random background noise. These features will not provide information about the behaviour, and they could potentially cause class overlaps. Therefore feature selection has utilised to reduce the class overlap.

In this research, the feature selection selects the subset of cepstral coefficients which have the best discriminant capabilities. The feature selection is performed using the *branch and bound* algorithm, which finds the optimal subset of features given that the selection criterion is monotonic [59]. In this research, the sum of squared euclidian distances between features, have been used as the criterion. Using this strategy, six cepstral coefficients were chosen (cepstral coefficient number 16, 15, 5, 4, 3 and 1) and used for training and classification.

5.2.2 Behaviour classification

The classification of behaviour is based on the methods described in the two previous sections, and a flow describing the procedure of the behaviour classification in this research, is shown in figure 5.3. The vocalisations are divided into short-time sequences, and feature extraction is performed, as shown in figure 5.2. The data is divided into training and test data; whereas the SVM models are trained and utilized for behaviour classification. The behaviour classification is based on the entire audio sequence (100 ms is used in this research).

The acoustic feature extraction was performed in MATLAB R2010b, using the Voicebox toolbox [69]. The training and evaluation of the SVMs was performed using LibSVM, which is an open-source SVM toolbox supporting multiple programming languages [70].

The extracted features for the three behaviours were divided into a training data set and a test data set. There were two strategies for evaluation of classifier performance. One was to use
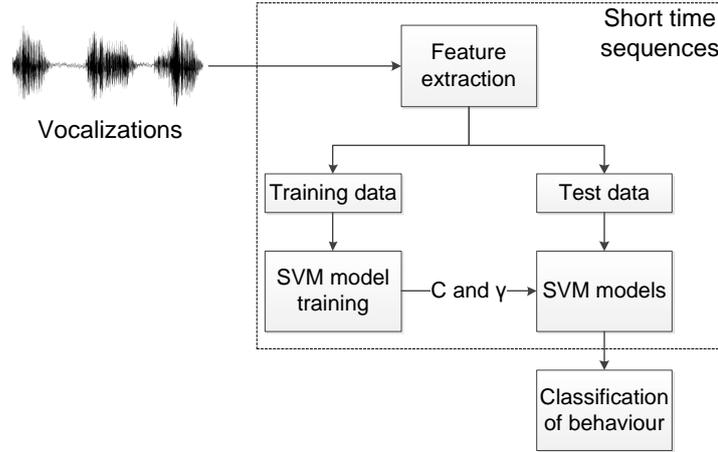
Figure 5.3: The flow of behaviour classification. The audio data is divided into short time sequences and feature extraction, modeling and classification is performed

data from day 1 as training data and data from day 2 as test data. This test strategy covers the generalisation capabilities of the classifer, as a good performance will indicate good performance on unseen data. The second test mixes all data and perform a 5-fold crossvalidation, using $4/5$ as training data and the remaining $1/5$ as test data. This measures the overall performance of the classifier. In the case of using day 1 as training data, the data was divided accordingly (day 1/day 2): flushing (44/56 %), foraging (60/40 %) and landing (62/38 %), due to the distribution of the behaviours in the two days. The two strategies are named Test A and Test B, respectively.

Before training the models, the data was normalised such that all feature vectors had zero mean and unit variance (5.6), to prevent certain features from dominating classification results due to large numerical values [59].

$$F'_{i,j} = \frac{F_{i,j} - \mu_j}{\sigma_j} \tag{5.6}$$

The training of the models consists of finding values for C and γ (as RBF kernel was chosen). This is done with a grid search, where every combination of C and γ is tested, within a predefined range or until a termination criteria is met. The evaluation of C and γ values are conducted using a five-fold cross validation scheme [71], where the C and γ with the average best cross validation rate is chosen. The grid search is done for all three SVM models, with iterative values of $2^{-10}, 2^{-9}, \ldots, 2^9, 2^{10}$ [71]. As more data for foraging and landing behaviour is available, the C values are scaled according to (5.7) and (5.8), to compensate for this [72]

$$C_1 = \frac{N}{2 \cdot N_1} \tag{5.7}$$

$$C_2 = \frac{N}{2 \cdot N_2} \tag{5.8}$$

Here N is the total number of feature vectors in the training data and N_1 and N_2 are the number of feature vectors for class one and two.

A total of three SVM models were trained, in a one-versus-one setup. The classification scheme is seen in figure 5.4, where a directional graph [73, 23] is used in the classification of behaviour. First the SVM model, modeling the hyperplane between *flushing* and *landing* behaviour, is evaluated and further evaluation steps are based on this result. The classification results are presented in a confusion matrix in the results section (see table 5.1), which gives the number of correct positive predictions (as bold numbers) and correct negative predictions, where the classifier rejects a behaviour correctly. Positive predictions or negative predictions, which are incorrect, are also given in the table. The performance of the models are given by three measures: accuracy, precision and sensitivity.
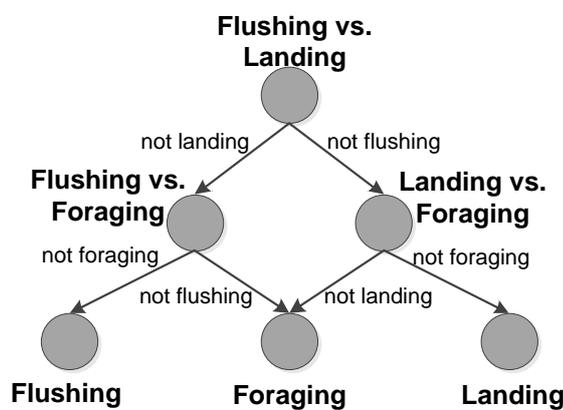
Figure 5.4: One-versus-one classification in a directional graph, where the direction is based on the SVM model results. The binary classification in each node, will result in classification of a single behaviour

5.3 Results

The GFCC feature extraction makes it possible to discriminate between the vocalisations of the described behaviours. This is visualised in figure 5.5, where the three first principal components of the selected features, are shown. The principal components are derived via principal component analysis (PCA)[74], and are the linear combination of the selected features which preserves the most variance in a smaller dimensional space. In figure 5.5, it is seen that foraging behaviour seems easiest to discriminate.

This observation is also supported in table 5.2, where the overall performance of the classification is described via statistical measurements. The results in table 5.2 are derived from the confusion matrix shown in table 5.1, and it is seen that the overall classification performance for foraging behaviour is higher than the other two, which is visualised in figure 5.5. However the overall classification performance is high, with accuracy measures over 90%. Some variability in precision and sensitivity for Test A and B is present.

The results from Test A show that the SVM models are capable of classifying unseen data, from another day, with high accuracy and precision. In this test the ratio between training and test
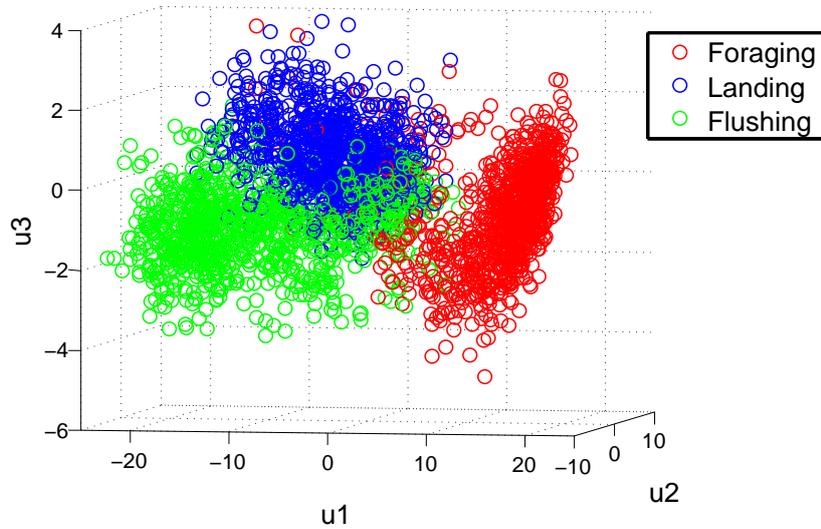
Figure 5.5: Plot of the first three principal components of the extracted features after feature selection has taken place. It can be seen from the plot, that it is possible to discriminate between the three behaviours, however the vocalisations for landing and fleeing have some similarities.

Table 5.1: Confusion matrix obtained from the classification of the three behaviours, using SVM with a six dimensional feature vector and RBF kernel function. The bold numbers indicate correct classification. The samples are 100 ms audio sequences. The notation A/B refers to the notation Test A and Test B, described in the section 5.2.2. A: Classification where data has been divided based on date, B: Classification where data has been mixed

Observed behaviour	Predicted behaviour			
	Flushing	Landing	Foraging	**Total**
Flushing	**129/44**	5/10	16/2	150/56
Landing	28/10	**219/144**	53/4	300/158
Foraging	5/13	14/28	**581/261**	600/302
Estimate	162/67	238/182	650/267	

data was close to $50/50$. The results in Test B show the overall performance of the classifier. In this test, the precision was a bit lower for flushing and landing behaviour. This is expected because the vocalisations of the two behaviours are quite similar, which makes it harder for the classifier to give precise results when these behaviours are present in the audio data.

5.4 Conclusion

It is possible to distinguish between landing, foraging and flushing behaviour based on acoustic information. Landing and flushing behaviours have similarities in their vocalisations, however the accuracy for classification was over 90% for all behaviours.

The SVM modeling has proven robust, with generalisation capabilities, as results from the two test strategies are comparable. The use of GFCC as features shows promising results, however

Table 5.2: Model performance for each behaviour classification. The same notation A/B as in table 5.1 is used in this table.

	Performance		
Behaviour	Accuracy[a]	Precision [b]	Sensitivity[c]
Flushing	0.95/0.93	0.80/0.66	0.86/0.79
Landing	0.91/0.90	0.92/0.79	0.73/0.91
Foraging	0.92/0.91	0.89/0.98	0.97/0.86

[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 (ratio between the bold numbers and total samples)

another choice of constants might prove more useful for this specific classification task.

Automatic behaviour recognition could improve automatic scaring devices, as it makes it possible to evaluate performance and alter strategies. In this paper it is shown that acoustic information can be used in the task of automatic recognition of landing, foraging and flushing behaviour.

<div style="text-align: right">**6**</div>

Automatic Detection of Animals in Mowing Operations using Thermal Cameras

The following chapter is part of a journal paper submitted to *Sensors*. The paper presents a method for automatic detection of wildlife during mowing operations. The methlogies are based on thermal imaging and digital image processing.

6.1 Introduction

During the last decades, strong competition in the agricultural sector has resulted in the development of high-efficiency farm equipment. This acceleration has also included efficiency improvement of moving techniques, which means that for instance grass cutting involves working speeds exceeding 15 km/h and working widths of more than 14 m. 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.

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 (*Crex crex*) [75, 8]. In particular, the nests of ground nesting bird species like grey partridge (*Perdix perdix*) or pheasant (*Phasianus colchicus*) are vulnerable to farming operations in their breeding habitat both as a result of the nests being destroyed [76] or the incubating female being killed or injured [11]. In mammals, the natural instinct of e.g. leverets of brown hare (*Lepus europaeus*) and fawns of roe deer (*Capreolus capreolus*) to lay low and still in the vegetation to avoid predators increase their risk of being killed or injured in farming operations [11]. As a result of the increase in both working speed and width, adults of otherwise mobile species, e.g. fox (*Vulpes vulpes*) 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.

Relatively few attempts have been made to assess the extent to which farming operations may negatively affect wildlife populations. In Germany, [11] estimated that at least 84,000 roe deer fawns, 153,000 brown hares, 11,000 wild rabbits (*Oryctolagus cuniculus*), 249,000 pheasants and 69,000 grey partridges were killed in farming operations. This corresponded to 14.5, 13.4, 1.1,

22.9 and 21.9% of the annual hunting bag, respectively. In Sweden, [9] estimated that fawn mortality caused by mowing ranged from 25-44% of the yearly recruitment during a three year study. In [77] the estimated leveret losses range from 17-44% in forage and grass fields, whereas losses were much lower in arable crops, ranging from 2-4% in spring barley (Hordeum spp.) and winter wheat (Triticum spp.), respectively. In Bulgaria, [76] estimated leveret mortality to be 27% in fodder plant biotopes. In France, [12] found that harvesting operations were of minor importance in adult hares, whereas [13] found no relationship between the juvenile proportion and grass leys or whole-crop suggesting that farming operations had no significance on recruitment in Danish hares. The above examples show that mortality resulting from farming operations may be significant, although highly variable depending on the species, age class and habitat type.

Besides the potential effects on wildlife populations, fodder contaminated with carcasses of animals may impose a health hazard for live stock from infection by the bacteria *Clostridium botulinum* causing botulism [78]. This may lead to commercial loss, which can be substantial.

Moreover, an aspect, which 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 [6, 7]) or strategy (e.g. leaving edge strips), longer mowing intervals, reduction of speed or higher cutting height [6, 8] 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 [6] or plastic sacks set out on poles before mowing [9] 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 [10]. In [79] a UAV-based 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 still performed manually.

Here we present a novel approach based on thermal imaging, which has been widely used to detect human activity [80, 81, 82, 83], whereas in animals, thermal imaging has been used to estimate cervid population densities in various habitats [84, 85], to detect and census mammals [86], for aerial surveys of mammals [87, 88], to study nighttime behaviour in grey partridges [89] and to detect migrating birds around offshore wind turbines [90]. These examples illustrate the wide range of applications of thermal imaging; however, most often the detection of both human and animal activity has been semi-automated, and therefore based on subsequent manual inspection of recorded images. In our study, we assessed the suitability of thermal imaging in combination with digital image processing to automatically detect animals present in the crop during mowing operations as part of a wildlife-friendly farming system.

6.2 Materials and methods

6.2.1 Study area

The experiment took place on the 27th of June 2011 in west Jutland, Denmark (WGS84: North $56°4.355'$, $East 8°23.053'$). The weather was partly sunny with temperatures ranging between 22-24°C.

6.2.2 Study animals

For our purpose, we used a domestic rabbit (*Oryctolagus cuniculus domesticus*) and a domestic chicken (*Gallus domesticus*) as study animals. The rabbit was chosen to resemble a leveret, whereas the chicken resembles a partridge or a pheasant. Specifically, we wanted to investigate whether the insulative property of feathers, which minimize the thermal differential between them and the environment [86], would hamper the detection a bird in the crop. The study animals were kept in a cage during the experiments.

6.2.3 Infrared thermography

Infrared thermography is based on the measurement of the radiation mode of heat transfer of a body in the infrared spectrum, which is a function of the temperature of the body [91]. All matters with a temperature above 0 K emit radiation [92]. In the infrared wavelength spectrum, mid-wave infrared (MWIR: 2.5-7 μm) and long-wave infrared (LWIR: 7-14 μm) are the most interesting wavelengths for imaging [93].

6.2.4 Equipment

An uncooled Forward Looking Infrared (FLIR) thermal camera was used for the recordings. The camera works in the Long-Wave Infrared Band (LWIR), which is preferred for animal detection, since the emitted radiation radiation from objects at ambient temperatures (300 K) peaks in LWIR [94]. The robustness of the camera minimizes the risk of vibrations during mowing operations. The camera was mounted (Fig. 6.1 a) to ensure good coverage of the area right in front of the tractor. The distance to the crop was approximately 4.75 m at an angle of approximately $75°$ perpendicular to the ground. The camera had a field of view of $25°$, giving a working width of 2.1 m in the centre of the video frame. The recorded video was stored on a laptop, which was also used to control the settings for the camera. The box in Fig. 6.1 b indicates the caged chicken used in the experiment. The tractor used in the experiment was a Claas Axos 320.

6.2.5 Data collection

The cages with the study animals were placed in the grass to imitate a natural situation, i.e. the grass was ready for mowing. They were kept in the same place throughout the experiment, except for one case, where the chicken was placed in dense grass cover. The tractor was driven at different speeds (rabbit: 4, 8, 12 and 15 km/h; chicken: 5, 10 and 15 km/h) using the same track. The camera temperature range was set at $10 - 35°C$.

6.2.6 Digital image processing

We used digital image processing techniques in order to automatically detect our study animals on the basis of the video recordings. Ideally, the thermal radiation of the study animals exceeds

Figure 6.1: Illustration of test setup. Figure 1a) shows the placement of the camera, and figure 1b) shows the inside the tractor, where the caged chicken is visible on the laptop screen

the radiation from the background, which makes the animal appear brighter on the video images. However, during sunlight periods, the thermal differences between the animal and the background may become smaller. In this case filtering techniques can be applied to enhance the appearance of the animals. For this purpose, we used the Laplacian of Gaussian (LoG), also known as the Mexican hat function, filter (1) for pre-processing to enhance the appearance.

$$\nabla^2 h(r) = -\left[\frac{r^2 - \sigma^2}{\sigma^4}\right] exp\left(-\frac{r2}{2\sigma^2}\right) \tag{6.1}$$

Here $r^2 = x^2 + y^2$, which are the x- and y-size of the filtering mask. The standard deviation controls the degree of blurring in the image. The size of the filter and blurring is based on the size of the animal, i.e. the number of pixels. The filter suppress the diffuse patches in the background, whereas the animal (the chicken) is enhanced (Fig. 6.2).
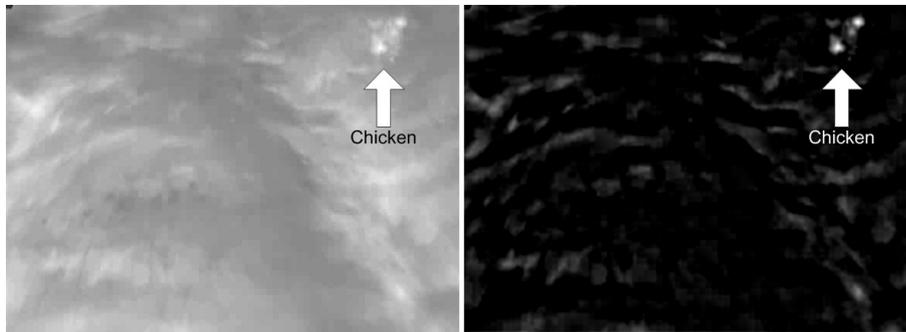
Figure 6.2: Illustration of the effect of pre-processing. Figure 3a) is the original image of a chicken, and figure 3b) is the filtered image. The filtering enhances the chicken so it can be discriminated from the background

On the basis of the pre-processed image, it is possible to identify the animal using adaptive thresholding [95], 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 (10 images have been used in test). This is because maximum values increase significantly when an animal is present in the image (Fig. 4), and this rapid increase in the values can be used to detect the animal in the video. 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.

Maximum value in each frame
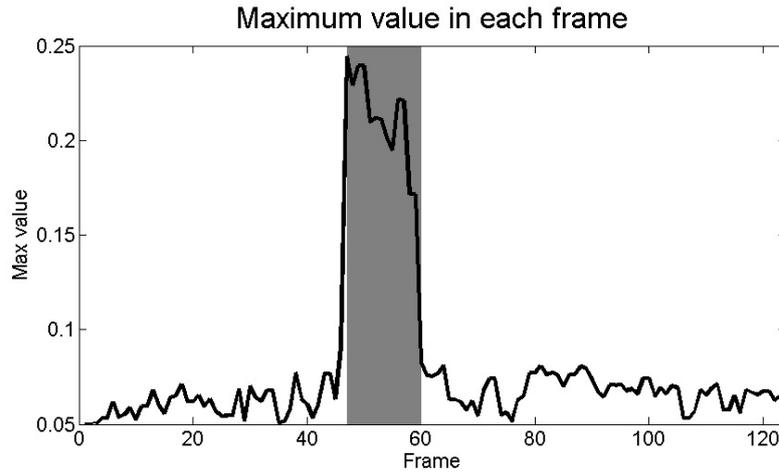
Figure 6.3: Plot of maximum values in the frames after pre-processing. The frames where an animal (the rabbit in this case) is present is marked with dark-grey. It is therefore possible to detect the presence of the animal on the basis of the maximum values in the frame

To ensure robustness to false detections, it is required that an animal have to be detected within a given region in three consecutive frames (inter frame consistency), as the frame rate governs the maximum distance the animal is capable of moving from frame to frame. In Fig. 6.4, a schematic presentation of the image processing algorithm for automatic animal detection is shown. We used this approach to test the suitability of using thermal imaging to automatically detect our study animals on the recorded videos. For comparison, we manually labelled the frames containing an animal and its position through visual inspection of the individual frames in the video recordings.
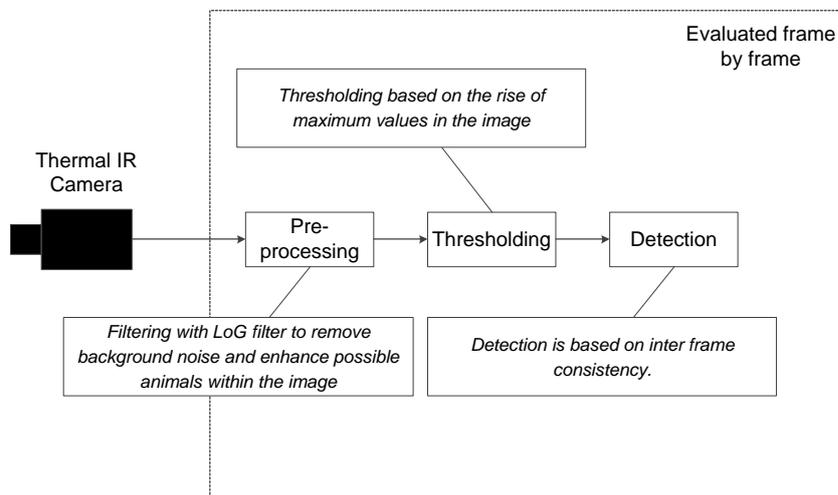
Figure 6.4: Flow of the image processing for the automatic detection of animals using thermal imaging

6.3   Results

The results of the automatic detection of our study animals at different driving speeds are expressed as the number of frames with true positive and false positive detections (Table 6.1).

The detection rate was almost 100% at all driving speeds (i.e. true positives were obtained for most frames with an animal present), whereas in one case, the chicken at 15 km/h, one frame was erroneously classified as containing an animal (false positive).

In one test scenario, where the chicken was covered in dense grass, the system was not able to detect it until it was very close to the camera. In this case, the manual labelling was also challenging and was only possible on the basis of the frames recorded closest to the camera.

Table 6.1: Number of true and false positives for rabbits and chickens at different driving speeds

	True positives	False positives	Number of frames with animal present [a]	Number of frames in video recording
Rabbit at 4 km/h	21	0	22	156
Rabbit at 8 km/h	13	0	13	124
Rabbit at 12 km/h	7	0	7	133
Rabbit at 15 km/h	5	0	5	128
Chicken at 5 km/h [b]	4	0	15	193
Chicken at 5 km/h	20	0	21	206
Chicken at 10 km/h	11	0	11	150
Chicken at 15 km/h	6	1	7	130

[a] Animal detected by means of visual inspection

[b] Dense grass cover

The choice of parameters could affect the detection of the animals in the first couple of frames of animal presence. The choice of parameters is based on the size of the animal, which is small when it first enters the frame, as the tractor was driving towards the animal. The effect of pre-processing therefore makes it difficult to detect the animal in this scenario as it would not be enhanced. This could explain why some the algorithm fails to detect all the frames in some of the recorded scenarios.

6.4   Conclusion

Thermal imaging and digital imaging processing may be an important tool for the improvement of wildlife-friendly farming practices and offers as such a potential for reducing wildlife mortality in agriculture.

We conclude that the use of thermal imaging for automated detection of animals during mowing operations holds potential. Under most circumstances, detection rates were close to 100%, although dense crops may hamper the detection of animals.

7

Summary

During the first half of my PhD work I have recorded audio and video data of wildlife geese during landing, foraging and flushing behaviour. The system and results of this has been presented in the conference paper *A Multimedia Capture System for Wildlife Studies* [45]. Based on the data, I have developed a pattern recognition algorithm for automatic detection of goose behaviour. This work is presented in the paper *A Vocal based Analytical Method for Goose Behaviour Recognition* [50]. The algorithm is however based on limited data, as it was not possible to record many occurences of wildlife geese in the short timespan they where active[1]. More data is therefore being recorded this spring (2012) to verify or improve the existing model.

Furhtermore a vision based approach has been developed in fall, 2011. This work is currently being drafted, and will be submitted to a journal in the near future. Based on these algorithms current research is focused on Audio-Visual recognition, which combines the two information streams. This is inspired by the work within Audio-Visual speech recognition, where a combination of the two gives high classification results.

One way of combining the two is through classifier fusion, which has been implemented in a system, which is currently being tested in a real life scenario. This test includes the audio and video based models, which has been developed during the first half of my PhD, and an actuator (speakers) for communicating with the geese. The expected result of this test is to verify the models in a real life scenario, and to investigate the effect of wildlife communication, with respect to wildlife management.

---

[1]The recordings took place in spring 2011, where the weather suddenly got very warm and the geese flew towards Norway

Bibliography

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

[2] J. M. Gilsdorf, S. E. Hygnstrom, and K. C. Vercauteren, "Use of frightening devices in wildlife damage management," *Integrated Pest Management Reviews*, vol. 7, no. 1, pp. 29–45, 2002.

[3] A. D. Fox, B. S. Ebbinge, C. Mitchell, T. Heinicke, T. Aarvak, K. Colhoun, P. Clausen, S. Dereliev, S. Faragó, K. Koffibjerg, H. Kruckenberg, M. J. J. E. Loonen, J. Madsen, J. Mooij, P. Musil, L. Nilsson, S. Pihl, and H. Van Der Jeugd, "Current estimates of goose population sizes in western Europe, a gap analysis and an assessment of trends," in *Expanding Goose Populations and their management, Proceedings of the 12th Meeting of the Goose Specialist Group*, pp. 115–127, 2010.

[4] M. Hake, J. Må nsson, and A. Wiberg, "A working model for preventing crop damage caused by increasing goose populations in Sweden," in *Expanding Goose Populations and their management, Proceedings of the 12th Meeting of the Goose Specialist Group*, pp. 225–233, 2010.

[5] D. Nolte, "Behavioral Approaches For Limiting Depredation by Wild Ungulates," in *Grazing Behavior of Livestock and Wildlife* (K. Launchbaugh, K. Sanders, and J. Mosley, eds.), pp. 60–69, University of Idaho, 1999.

[6] C. Green, "Reducing Mortality of Grassland Wildlife During Haying and Wheat-Harvesting Operations," *Oklahoma State University Forestry Publications*, pp. 1–4, 1998.

[7] G. Tyler, R. Green, and C. Casey, "Survival and behaviour of corncrake crex crex chicks during the mowing of agricultural grassland," *Bird Study*, vol. 45, no. 1, pp. 35–50, 1998.

[8] K. Koffijberg and N. Schaffer, "International Single Species Action Plan for the Conservation of the Corncrake Crex crex International Single Species Action Plan for the Conservation of the Corncrake," *CMS Technical Series No. 14 & AEWA Technical Series, No. 9, Bonn, Germany*, 1976.

[9] A. Jarnemo, "Roe deer Capreolus capreolus fawns and mowing - mortality rates and countermeasures," *Wildlife Biology*, vol. 8, pp. 211–218, 2002.

[10] V. Tank, P. Haschberger, H. Dietl, and W. Lutz, "Infrarotoptischer wildsensor—eine entwicklung zur detektion von wild in wiesen und zur wildrettung bei der frühjahrsmahd," *Zeitschrift für Jagdwissenschaft*, vol. 38, no. 4, pp. 252–261, 1992.

[11] L. Kittler, "Wildverluste durch den einsatz landwirtschaftlicher maschinen nach einer erhebung aus dem jagdjahr 1976/77 in nordrhein-westfalen," *Zeitschrift für Jagdwissenschaft*, vol. 25, no. 1, pp. 22–32, 1979.

[12] E. Marboutin and N. Aebischer, "Does harvesting arable crops influence the behaviour of the european hare lepus europaeus," *Wildlife biology*, vol. 2, no. 2, pp. 83–91, 1996.

[13] T. Wincentz, "Identifying causes for population decline of the brown hare (lepus europaeus) in agricultural landscapes in denmark," *PhD thesis, Department of Wildlife Ecology and Biodiversity, NERI. National Environmental Research Institute, Aarhus University, Denmark and Department of Population Biology, University of Copenhagen*, p. 194, 2009.

[14] S. M. Rutter, "Revista Brasileira de Zootecnia The integration of GPS , vegetation mapping and GIS," *Revista Brasileira de Zootecnia*, vol. 36, pp. 63–70, 2007.

[15] E. Nadimi, H. Sø gaard, and T. Bak, "ZigBee-based wireless sensor networks for classifying the behaviour of a herd of animals using classification trees," *Biosystems Engineering*, vol. 100, pp. 167–176, 2008.

[16] Y. Guo, P. Corke, G. Poulton, T. Wark, D. Swain, P. Corke, G. Poulton, and T. Wark, "Animal Behaviour Understanding using Wireless Sensor Networks," in *Local Computer Networks*, pp. 607–614, 2006.

[17] R. Müller and L. Schrader, "A new method to measure behavioural activity levels in dairy cows," *Applied Animal Behaviour Science*, vol. 83, pp. 247–258, 2003.

[18] E. E. Vallejo and C. 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*, vol. 14, pp. 485–488, Jan. 2010.

[19] D. Magee, "Detecting lameness using 'Re-sampling Condensation' and 'multi-stream cyclic hidden Markov models'," *Image and Vision Computing*, vol. 20, pp. 581–594, June 2002.

[20] R. Tillett, "Using model-based image processing to track animal movements," *Computers and Electronics in Agriculture*, vol. 17, pp. 249–261, May 1997.

[21] M. Matetić, S. Ribarić, and I. Ipšić, "Qualitative Modelling and Analysis of Animal Behaviour," *Applied Intelligence*, vol. 21, pp. 25–44, July 2004.

[22] P. J. Clemins, M. T. Johnson, K. M. Leong, and A. Savage, "Automatic classification and speaker identification of African elephant (Loxodonta africana) vocalizations," *The Journal of the Acoustical Society of America*, vol. 117, no. 2, p. 956, 2005.

[23] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–9, 2007.

[24] G. Jahns, "Call recognition to identify cow conditions—A call-recogniser translating calls to text," *Computers and Electronics in Agriculture*, vol. 2, pp. 54–58, Nov. 2007.

[25] G. Manteuffel and P. C. Schön, "Measuring pig welfare by automatic monitoring of stress calls," *Bornimer Agrartechnische Berichte*, vol. 29, pp. 110–118, 2002.

[26] J. R. Deller, Jr., J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1st ed., 1993.

[27] C. Fant, *Analysis and synthesis of speech processes.* North-Holland Publishing Comp., 1968.

[28] L. G. Johansen, "Talekodning ved hjælp af lineær prediktiv kodning , LPC," 2007.

[29] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task," in *Proceedings of SPECOM*, vol. 1, pp. 191–194, 2005.

[30] H. A. . M. Hemant Tyagi, Rajesh M . Hegde and A. Prabhakar, "Automatic Identification of Bird Calls Using Spectral Ensemble Average Voice Prints," pp. 1–5.

[31] J. A. Kogan and D. Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," *The Journal of the Acoustical Society of America*, vol. 103, no. 4, pp. 2185–2196, 1998.

[32] G. Manteuffel, "Vocalization of farm animals as a measure of welfare," *Applied Animal Behaviour Science*, vol. 88, pp. 163–182, Sept. 2004.

[33] V. M. Trifa, A. N. G. Kirschel, C. E. Taylor, and E. E. Vallejo, "Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models.," *The Journal of the Acoustical Society of America*, vol. 123, pp. 2424–31, Apr. 2008.

[34] P.-C. Schön, B. Puppe, and G. 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*, vol. 110, no. 3, p. 1425, 2001.

[35] C. Lee, C. Chou, C. Han, and R. Huang, "Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis," *Pattern Recognition Letters*, vol. 27, pp. 93–101, Jan. 2006.

[36] J. C. Brown and P. Smaragdis, "Hidden Markov and Gaussian mixture models for automatic call classification.," *The Journal of the Acoustical Society of America*, vol. 125, pp. EL221–4, June 2009.

[37] P. Clemins, M. Trawicki, K. Adi, and M. Johnson, "Generalized Perceptual Features for Vocalization Analysis Across Multiple Species," *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, pp. I–253–I–256, 2006.

[38] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, pp. 334–352, Aug. 2004.

[39] M. S. Dawkins, H.-j. Lee, C. D. Waitt, and S. J. Roberts, "Optical flow patterns in broiler chicken flocks as automated measures of behaviour and gait," *Applied Animal Behaviour Science*, vol. 119, pp. 203–209, July 2009.

[40] B. N. Courty and T. Corpetti, "Crowd motion capture," *Computer animations and virtual worlds*, vol. 18, no. July, pp. 361–370, 2007.

[41] G. Medioni, I. Cohen, F. Brémond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 8, pp. 873–889, 2001.

[42] M. Matetic and S. Ribaric, "Qualitative modelling of object behaviour in the dynamic vision system using hidden Markov models," in *11th IEEE Mediterranean Electrotechnical Conference*, pp. 570–574, IEEE, 2002.

[43] M. Sugimoto, T. T. Zin, T. Toriu, and S. Nakajima, "Robust Rule-Based Method for Human Activity Recognition," *International Journal of Computer Science and Network Security*, vol. 11, no. 4, pp. 37–43, 2011.

[44] I. L. Macdonald and D. Raubenheimer, "Hidden Markov Models and Animal Behaviour," *Biometrical Journal*, vol. 37, no. 6, pp. 701–712, 1995.

[45] K. A. Steen, H. Karstoft, and O. Green, "A Multimedia Capture System for Wildlife Studies," in *The Third International Conference on Emerging Network Intelligence*, no. c, pp. 131–134, 2011.

[46] D. Tweed and A. Calway, "Tracking multiple animals in wildlife footage," in *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 24–27, IEEE Comput. Soc, 2002.

[47] R. Gula, J. Theuerkauf, S. Rouys, and A. Legault, "An audio / video surveillance system for wildlife," *European Journal of Wildlife Research*, vol. 56, no. 5, pp. 803–807, 2010.

[48] A. M. Kleist, R. A. Lancia, and P. D. Doerr, "Using Video Surveillance to Estimate Wildlife Use of a Highway Underpass," *Journal of Wildlife Management*, vol. 71, no. 8, pp. 2792–2800, 2007.

[49] S. Ward, C. M. Bishop, a. J. Woakes, and P. J. Butler, "Heart rate and the rate of oxygen consumption of flying and walking barnacle geese (Branta leucopsis) and bar-headed geese (Anser indicus).," *The Journal of experimental biology*, vol. 205, pp. 3347–56, Nov. 2002.

[50] K. A. Steen, O. R. Therkildsen, H. Karstoft, and O. Green, "A Vocal-Based Analytical Method for Goose Behaviour Recognition," *Sensors*, vol. 12, pp. 3773–3788, Mar. 2012.

[51] E. David and S. Mark, "Classifying cattle jaw movements : Comparing IGER Behaviour Recorder and acoustic techniques," *Applied Animal Behaviour Science*, vol. 98, pp. 11–27, 2006.

[52] Y. Ren, M. T. Johnson, P. Clemins, M. Darre, S. S. Glaeser, T. S. Osiejuk, and E. Out-Nyarko, "A Framework for Bioacoustic Vocalization Analysis Using Hidden Markov Models," *Algorithms*, vol. 2, pp. 1410–1428, Nov. 2009.

[53] D. Milone, H. Rufiner, J. Galli, E. Laca, and C. Cangiano, "Computational method for segmentation and classification of ingestive sounds in sheep," *Computers and Electronics in Agriculture*, vol. 65, pp. 228–237, Mar. 2009.

[54] R. E. Thomas, K. M. Fristrup, and P. 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*, vol. 112, no. 4, p. 1692, 2002.

[55] D. Moura, W. Silva, I. Naas, Y. Tolon, K. Lima, and M. Vale, "Real time computer stress monitoring of piglets using vocalization analysis," *Computers and Electronics in Agriculture*, vol. 64, pp. 11–18, Nov. 2008.

[56] D. Reby, R. Andre-Obrecht, A. Galinier, J. Farinas, and B. Cargnelutti, "Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (Cervus elaphus) stags," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, p. 4080, 2006.

[57] W. M. Campbell, "Speaker Verification Using Support Vector Machines and High-Level Features," in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 10, pp. 1641–2094, Sept. 2007.

[58] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[59] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*. Academic Press, 4th ed., 2008.

[60] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[61] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, "Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines," *Applied Animal Behaviour Science*, vol. 119, pp. 32–38, June 2009.

[62] V. Blanz, B. Scholkopf, H. Bultho, C. Burges, V. Vapnik, and T. Vetter, "Comparison of view-based object recognition algorithms using realistic 3D models," in *International Conference on Artifiial Neural Networks*, pp. 251–256, 1996.

[63] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D. Petrovska-Delacrétaz, and D. a. Reynolds, "A Tutorial on Text-Independent Speaker Verification," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 4, pp. 430–451, 2004.

[64] L. P. Dmitrieva and G. Gottlieb, "Influence of Auditory Experience on the Development of Brain Stem Auditory-Evoked Potentials in Mallard Duck Embryos and Hatchlings," *Behavioral and neural biology*, vol. 61, pp. 19–28, 1994.

[65] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, pp. 39–58, Jan. 2009.

[66] D. D. Greenwood, "Critical bandwidth and consonance in relation to cochlear frequency-position coordinates.," *Hearing research*, vol. 54, pp. 164–208, Aug. 1991.

[67] K. Adi, M. T. Johnson, and T. S. Osiejuk, "Acoustic censusing using automatic vocalization classification and identity recognition," *Acoustical Society of America*, vol. 127, no. 2, pp. 874–883, 2010.

[68] E. L. LePage, "The mammalian cochlear map is optimally warped," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, p. 896, 2003.

[69] M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," 1997-2011.

[70] C.-c. Chang and C.-j. Lin, "LIBSVM : A Library for Support Vector Machines," 2011.

[71] C.-w. Hsu, C.-c. Chang, and C.-j. Lin, "A Practical Guide to Support Vector Classification," 2010.

[72] A. Ben-Hur and J. Weston, "A users guide to support vector machines," in *Data Mining Techniques for the Life Sciences* (O. Carugo, F. Eisenhaber, and J. M. Walker, eds.), vol. 609 of *Methods in Molecular Biology*, pp. 223–239, Humana Press, 2010.

[73] J. C. Platt, M. Way, and J. Shawe-taylor, "Large Margin DAGs for Multiclass Classification," *MIT Press*, pp. 547–553, 2000.

[74] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[75] R. Green and T. Stowe, "The decline of the corncrake crex crex in britain and ireland in relation to habitat change," *Journal of Applied Ecology*, pp. 689–695, 1993.

[76] Z. Milanov, "Effect of mowing fodder plants on small game populations in central bulgaria," *Proceedings of the International Union of Game Biologist XXII Congress: The Game and the Man. Sofia Bulgaria, September 4-8 1995*, vol. 22, pp. 394–397, 1996.

[77] J. Kaluzinski and Z. Pielowski, "The effect of technical agricultural operations on the hare population," *Ecology and Management of European hare populations*, pp. 205–211, 1976.

[78] F. Galey, R. Terra, R. Walker, J. Adaska, M. Etchebarne, B. Puschner, E. Fisher, R. 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*, vol. 12, no. 3, p. 204, 2000.

[79] M. Israel, "A UAV-BASED ROE DEER FAWN DETECTION SYSTEM," in *Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)* (H. Eisenbeiss, M. Kunz, and H. Ingensand, eds.), vol. XXXVIII, (Zurich, Switzerland), pp. 1–5, ISPRS, 2011.

[80] C. Dai, Y. Zheng, and X. Li, "Pedestrian detection and tracking in infrared imagery using shape and appearance," *Computer Vision and Image Understanding*, vol. 106, pp. 288–299, May 2007.

[81] J. C. Castillo, J. Serrano-Cuerda, A. Fernández-Caballero, and M. T. López, "Segmenting Humans from Mobile Thermal Infrared Imagery.pdf," *Lecture Notes in Computer Science*, no. 5602, pp. 334–343, 2009.

[82] A. Fernández-Caballero, J. C. Castillo, J. Martínez-Cantos, and R. Martínez-Tomás, "Optical flow or image subtraction in human detection from infrared camera on mobile robot," *Robotics and Autonomous Systems*, vol. 58, pp. 1273–1281, Dec. 2010.

[83] H. Sun, C. Wang, B. Wang, and N. El-Sheimy, "Pyramid binary pattern features for real-time pedestrian detection from infrared videos," *Neurocomputing*, vol. 74, pp. 797–804, Feb. 2011.

[84] S. S. Ditchkoff, J. B. Raglin, J. M. Smith, and B. A. Collier, "From the Field. Capture of white-tailed deer fawns using thermal imaging technology," *Wildlife Society Bulletin*, vol. 33, no. 3, pp. 1164–1168, 2005.

[85] R. Gill, M. Thomas, and D. Stocker, "The use of portable thermal imaging for estimating deer population density in forest habitats," *Journal of Applied Ecology*, pp. 1273–1286, 1997.

[86] R. Boonstra, J. Eadie, C. Krebs, and S. Boutin, "Limitations of far infrared thermal imaging in locating birds (limitaciones en el uso de imágenes termales con radiación infraroja extrema para localizar aves)," *Journal of field ornithology*, pp. 192–198, 1995.

[87] K. Havens and E. Sharp, "Using thermal imagery in the aerial survey of animals," *Wildlife Society Bulletin*, pp. 17–23, 1998.

[88] E. Wiggers and S. Beckerman, "Use of thermal infrared sensing to survey white-tailed deer populations," *Wildlife society bulletin*, vol. 21, no. 3, pp. 263–268, 1993.

[89] J. Tillmann, "Fear of the dark: night-time roosting and anti-predation behaviour in the grey partridge (perdix perdix l.)," *Behaviour*, vol. 146, no. 7, pp. 999–1023, 2009.

[90] M. Desholm, "Thermal animal detection system (tads): development of a method for estimating collision frequency of migrating birds at offshore wind turbines," tech. rep., National Environmental Research Institute, 2003.

[91] H. Kaplan, *Practical applications of infrared thermal sensing and imaging equipment*, vol. 75. Society of Photo Optical, 2007.

[92] D. DeWitt and F. Incropera, "Physics of thermal radiation," *Theory and Practice of Radiation Thermometry*, pp. 19–89, 1988.

[93] A. Richards, "Alien vision: exploring the electromagnetic spectrum with imaging technology," vol. 104, pp. 31–48, 2001.

[94] L. St-Laurant, D. Prévost, and X. Maldague, "Thermal imaging for enhanced foreground - background segmentation," in *The 8th International Conference on Quantitative Infrared Thermography*, 2006.

[95] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2nd ed., 2001.

Kim Arild Steen, Ole Roland Therkildsen, Henrik Karstoft and Ole Green: Wildlife Communication, 2012

Department of Engineering
Aarhus University
Edison, Finlandsgade 22
8200 Aarhus N
Denmark

Tel.: +45 4189 3000