A Multimedia Capture System for Wildlife Studies

Kim Arild Steen
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Dalgas Avenue 2, Aarhus, Denmark
KimA.Steen@agrsci.dk

Henrik Karstoft
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Aarhus, Denmark
hka@iha.dk

Ole Green
Department of Engineering,
Aarhus University
Faculty of Science and Technology
Aarhus, Denmark
Ole.Green@agrsci.dk

***Abstract*—This paper presents a system for video and audio recording of wildlife geese in their natural environment. The system enables remote controlled recording, and is designed for an outdoor environment. The recordings lasted 1 month, where 4-5 hours of geese video and audio were succesfully captured. Data recorded using the system is a part of ongoing research to design a method for automatic recognition of animal behaviour and species based on audio and video recordings. Automatic recognition could potentially lead to systems capable of reducing habituation.**

***Keywords*-video; audio; recording; wildlife surveillance; remote**

I. 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 [6][8] and audio processing for recognition of animal vocalizations [1][5][7]. 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 [3][4]; 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 [3] 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.

The structure of this paper is as follows: In Section II the requirements for the system is presented followed by a description of the system in Section III. The implemented infrastructure is presented in Section IV and preliminary results, using the system, are presented in Section V. The discussion in Section VI follows up on the results and experience gained while using the system. The paper ends with a conclusion in Section VII.

II. 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 $20\frac{m}{s}$ [9], 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.

 ISBN: 978-1-61208-174-8

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

III. System description

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

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 I
TABLE OF SYSTEM COMPONENTS USED IN THE SETUP

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 III.

IV. Infrastructure and data description

The main purpose of the system is to record and store large amounts of data. In Figure 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 $44.1kHz$ 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

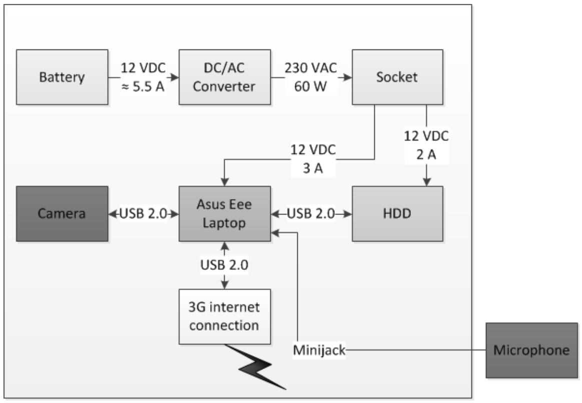

Figure 1. Block diagram of the system

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.

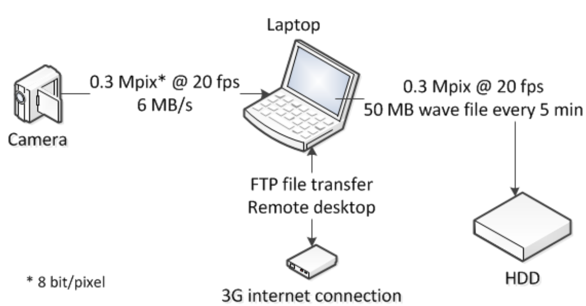

Figure 2. Overview of the infrastructure of the system setup, including description of data

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.

V. RESULTS

The described system was used for recording wild life geese over a period of one month, where the only downtime was due to replacement of batteries. Over 4.5 hours of geese audio and video, capturing landing, eating and fleeing, were succesfully recorded. During the recordings, the weather conditions were diverse, including storms and sunny weather, and the system and the recordings were not affected.

The power supply used for the setup was car batteries, and with two 95 Ah batteries, the system was able to run for approximately three and a half days. This was accomplished by putting the system to stand-by every night.

During the recording, the average CPU load was $45 - 50\%$, with a peak load of 70%. The memory load was constantly on 40%. If processing of the signals were to be implemented, the system has to be upgraded, however this was not the scope of the system at hand.

In Figure 3 an example of the recorded audio is shown. The geese vocalizations are clearly visible in the shown waveform, as they appear with high amplitude in the recordings. Visual inspection of the recorded video also verify this, as the geese were standing close to the microphone. Recordings of the three desired behaviours were present in the recorded data (both video and audio).

The waveform shown in Figure 3 was recorded on a windless day, however some noise is present in the recording. This is due to the amplification of the audio signal in the built-in sound card in the laptop.

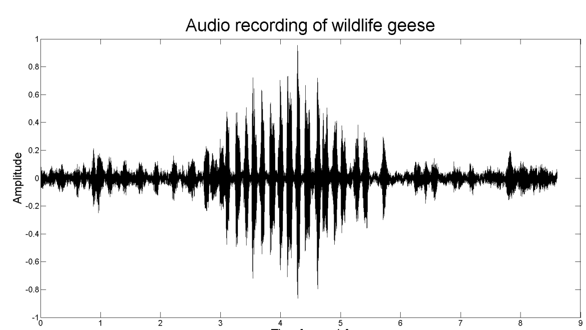

Figure 3. A sample waveform of the recorded audio while geese were foraging. The vocalizations are easily detected, as they appear with high amplitude

VI. Discussion

As mentioned in the results section, the chosen laptop cannot be used if processing of the recorded audio or video were to be done, which will be a part of further development of the system. A laptop with more memory and a faster CPU is desirable. Other limitations with the chosen laptop, are the frame rate and resolution of the images. The camera supports both higher frame rates and higher resolution, and a faster laptop could increase both without over-burdening the CPU.

The overall power consumption was reduced by setting the system to stand-by at nighttime. Another approach could be to trigger the recording, and only record when animals are present. This was not chosen as a loss of usefull data, due to potential trigger errors, could delay the further research. However a suggested modification could be to use a computer vision approach to trigger the recording.

The microphone used in the experiment, was a directional shotgun microphone. Another approach could be parabolic microphones, which are directional and amplify the sound before the digitizing of the audio. This was not used in this recording due to the physical size of the microphone.

Some noise is present in the recordings, even on windless days. This is due to the low quality sound card in the laptop, and an external sound card or a microphone amplifier could provide a better quality recording, however the vocalizations were clear in the recordings. A spectrogram analysis of the recorded vocalizations have shown that all information in the geese vocalizations is within 10 kHz, which means that sample rate could be reduced, which is preferable if more processing were to be performed in the system.

The remote access allowed for adjustments of audio and video recording, however aperture could not be adjusted as this must be done manually. Adjustments were not made during the recordings, as it was hard to verify the image quality on the remote access, which was mainly used to verify the presence of geese.

The recorded data contained examples of the three desired behaviours: land, eat and take off. Automatic behaviour recognition research, based on these recordings, could lead to a system capable of reducing habituation, as scaring mechanism could be targeted towards specific behaviours or species.

Even though the system was used for a specific scenario, it is applicable to other wildlife studies where audio and video material is essential. It is designed to cope with different weather conditions and the remote access makes it possible to verify recordings.

VII. 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.

Acknowledgment

We would like to thank Morten Laursen for contributing to the overall design, and I would also like to thank Esben Rasmussen and Claus Andersen for support in developing the imaging capture software.

References

[1] Seppo Fagerlund. Bird Species Recognition Using Support Vector Machines. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

[2] 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.

[3] Roman Gula, Jörn Theuerkauf, Sophie Rouys, and Andrew Legault. An audio / video surveillance system for wildlife. *European Journal of Wildlife Research*, 56(5):803–807, 2010.

[4] Andrea M. Kleist, Richard A. Lancia, and Phillip D. Doerr. Using Video Surveillance to Estimate Wildlife Use of a Highway Underpass. *Journal of Wildlife Management*, 71(8):2792–2800, 2007.

[5] C Lee, C Chou, C Han, and R Huang. Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. *Pattern Recognition Letters*, 27(2):93–101, January 2006.

[6] Maja Matetić, Slobodan Ribarić, and Ivo Ipšić. Qualitative Modelling and Analysis of Animal Behaviour. *Applied Intelligence*, 21(1):25–44, July 2004.

[7] Vlad M Trifa, Alexander N G 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):2424–31, April 2008.

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

[9] 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*, 205(Pt 21):3347–56, November 2002.