Ecological Informatics 30 (2015) 20–28

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Two-dimensional thermal video analysis of offshore bird and bat flight

Shari Matzner [a,*], Valerie I. Cullinan [a], Corey A. Duberstein [b]

[a] Pacific Northwest National Laboratory, Marine Sciences Laboratory, 1529 W. Sequim Bay Rd., Sequim, WA 98382, USA
[b] Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA

ARTICLE INFO

Article history:
Received 1 April 2015
Received in revised form 3 September 2015
Accepted 3 September 2015
Available online 11 September 2015

Editor: Friedrich Recknagel

Keywords:
Thermal imaging
Biological assessment
Image processing
Object tracking

ABSTRACT

Thermal infrared video can provide essential information about bird and bat activity for risk assessment studies, but the analysis of recorded video can be time-consuming and may not extract all of the available information. Automated processing makes continuous monitoring over extended periods of time feasible, and maximizes the information provided by video. This is especially important for collecting data in remote locations that are difficult for human observers to access, such as proposed offshore wind turbine sites. We developed new processing algorithms for single camera thermal video that automate the extraction of two-dimensional bird and bat flight tracks, and that characterize the extracted tracks to support animal identification and behavior inference. The algorithms consist of video peak store followed by background masking and perceptual grouping to extract flight tracks. The extracted tracks are automatically quantified in terms that could then be used to infer animal taxonomy and possibly behavior, as described in the companion article from Cullinan, et al. ["Classification of birds and bats using flight tracks." *Ecological Informatics*, 27:55–63]. The developed automated processing was evaluated using six video clips containing a total of 184 flight tracks. The detection rate was 81% and the false positive rate was 17%. In addition to describing the details of the algorithms, we suggest models for interpreting thermal imaging information.

1. Introduction

The risks to birds and bats posed by wind turbines, particularly offshore installations, are not well understood. Point counts by observers do not provide the level of detail and temporal resolution needed to fully characterize animal activity, and offshore locations are challenging to survey. Video recording from surface-mounted or aerial platforms has the potential to provide the required information but human analysis of video is time-consuming and difficult to validate. Automated processing can make continuous, long-term observations using video feasible and cost-effective, thereby providing the required information for accurate risk assessment.

It is difficult, if not impossible, for onsite human observers to quantify animal movement patterns through an area over the course of multiple diurnal-nocturnal cycles and during varying weather conditions. This is especially true for offshore wind turbine sites, which may be located in open ocean up to 50 km offshore, an area with limited accessibility and significant objective danger. Observations are necessary during daylight and also at night because many seabirds are known to actively forage at all times (Suryan et al., 2006), and both landbirds and shorebirds migrate over water at night (Liechiti et al., 1995; Lindeboom et al., 2011). A remote sensing solution can provide continuous coverage over extended periods of time if 1) an appropriate sensor that can capture the required information is used, and 2) the sensor data are efficiently processed in a way that distills the essential information while minimizing data storage and transmission requirements.

Suggested methods for assessing the impact on birds and bats from offshore wind energy development identify the primary risks as collision mortality and habitat loss through displacement (Desholm et al., 2004; Robinson Willmott et al., 2013). Metrics defined to estimate these risks include relative abundance, annual occurrence in hours, migration traffic rate, mean flight altitude, and time spent feeding in the area. At a minimum, animal abundance and passage rates are needed (Kunz et al., 2007). Concern is focused on threatened and endangered species, and on migratory seabirds and passerines. A remote sensing solution will enable accurate passage counts and capture enough information to make inferences about the taxa of observed animals and their behavior.

Thermal infrared video can provide information about animal passage rates and activity patterns during both day and night. Thermal imaging has been used to study nocturnal bird migration traffic (Gauthreaux and Livingston, 2006; Zehnder et al., 2001), bat behavior in terrestrial settings (Betke et al., 2008; Cryan et al., 2014; Hristov et al., 2008), and avian interactions with offshore wind turbines (Desholm et al., 2006). Of all of these studies, only the bat studies used automated processing to count the number of animals; the other studies using thermal imaging relied on manual analysis.

* Corresponding author. Tel.: +1 360 681 4577.
E-mail addresses: Shari.Matzner@pnnl.gov (S. Matzner), Valerie.Cullinan@pnnl.gov (V.I. Cullinan), Corey.Duberstein@pnnl.gov (C.A. Duberstein).

http://dx.doi.org/10.1016/j.ecoinf.2015.09.001

We have developed algorithms to automatically extract the two-dimensional flight tracks of birds and bats from single camera thermal video. This work was originally inspired by the work of Gauthreaux and Livingston (2006), where a thermal camera, a customized radar system, and a video peak store device were used to record the flight tracks of nocturnally migrating birds.

Automated processing of thermal video was developed by Betke et al. (2007) for censusing bats. The challenge was to track very large numbers of animals in the field of view at the same time. The approach was to use a detection algorithm applied to each video frame to detect objects, combined with a tracking algorithm that tracked the detected objects and an assignment algorithm that assigned detected objects to tracks. The Betke tracking algorithm was sophisticated—and computationally expensive—because it needed to track and count hundreds of individual animals at the same time. Our approach is simpler, less computationally and memory intensive, yet sufficient for tracking tens of animals in the field of view at the same time.

Three-dimensional tracking of bats using stereoscopic methods was demonstrated by Theriault et al. (2010). Three-dimensional flight trajectories provide more information about animal behavior and flight characteristics than two-dimensional tracks, and can be valuable for studying micro-avoidance behaviors around wind turbines. Stereoscopic tracking requires at least two cameras, and the cameras must be calibrated (Theriault et al., 2014). The effective field of view of each camera is reduced to the area of overlap between the two views. Post-processing is required for 3D trajectory reconstruction. The added costs of stereoscopic methods must be considered in the context of study objectives. For obtaining counts and general activity patterns, two-dimensional track analysis is sufficient.

The Thermal Animal Detection System (TADS) developed by Desholm and Bertelsen (2003) was designed to monitor bird collisions with offshore wind turbines. This remote sensing system used a detection threshold to trigger video capture to limit the amount of recorded data to the times when birds were present in the field of view. The trigger was confounded by moving clouds and the reflective surface of the sea, each of which limited the effectiveness of the system. Our detection approach attempts to address these types of environmental noise by using a localized low-pass filter and by constraining detections to objects moving at a minimum rate.

Our objective was to develop automated processing of thermal video to extract useful information about animal activity for risk assessment and monitoring applications. The technical contributions of this work are:

- software that automatically extracts animal flight tracks from thermal video in near real-time and reduces the raw video data to concise flight track information
- the automatic quantitative characterization of flight tracks in the form of summary statistics of animal size, flight speed, and thermal intensity for each track, and
- equations for translating track statistics into physical units using knowledge of the species expected to occur in the area observed with the video.

2. Methods

2.1. Thermal imaging

The basic components of a thermal camera are shown in Fig. 1. A thermal infrared camera works similarly to an optical camera in that a lens focuses energy onto an array of receptors to produce an image. A thermal image is an intensity image, where the intensity value of each pixel is related to the amount of thermal energy incident on an element in the receptor array. A thermal image contains no chromatic (color) information.

The specifications of a thermal imaging camera include the wavelength measured, thermal sensitivity, receptor array size, field of view angles, and the supported frame rates. The specifications for three commercially available cameras are given in Table 1; the leftmost and center columns describe research-grade uncooled cameras and the rightmost column describes a more sensitive cooled camera. Cooled cameras have higher thermal sensitivity than uncooled cameras, and are more expensive to operate and maintain. A shorter wavelength, 3–5 μm, provides better resolution but can be affected by water vapor in the atmosphere. A longer wavelength, 8–12 μm, is more reliable in humid conditions but provides less shape detail. A detailed treatment of atmospheric and metrological effects on thermal imaging can be found on the FLIR webpage (www.flir.com).

The spatial resolution of the camera is determined by the number of receptor elements, which gives the size in pixels of a single frame of video. The field of view, measured in angle degrees, is determined by the focusing lens; some cameras have a zoom capability that provides a range of field of view angles. The spatial area, X by Y meters, of a camera's field of view is

$$X = 2R\tan\frac{\alpha_H}{2}, \quad Y = 2R\tan\frac{\alpha_V}{2}\text{meters}, \tag{1}$$

where α_H and α_V are the horizontal and vertical field of view angles, respectively, and R is the range in meters from the camera. A convenient measure of area coverage is the diagonal size of the field of view,

$$D = \sqrt{X^2 + Y^2}\ \text{meters}. \tag{2}$$

The diagonal spatial resolution in pixels per meters is

$$r = \frac{\sqrt{N_H^2 + N_V^2}}{D}\ \text{pixels per meter}. \tag{3}$$

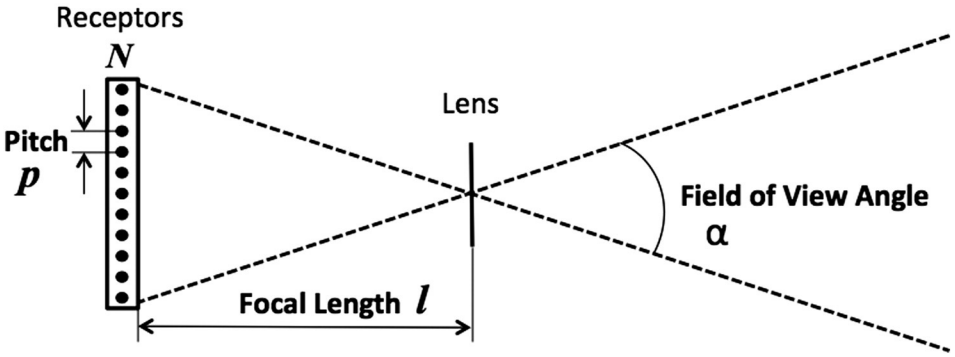

Fig. 1. The thermal imaging camera characteristics that determine the spatial resolution and field of view.

Table 1
Specifications of commercial thermal infrared video cameras.

	Electrophysics PV320 $	FLIR A655sc $$	Axsys FieldPro 5X $$$
Spectral range (wavelength)	7–14 μm	7.5–14 μm	3–5 μm
Thermal sensitivity	0.08 °C	0.05 °C	0.02 °C
Thermal accuracy	±2%	±2 °C or 2%	–
Receptor array, horizontal size	320	640	320
Receptor array, vertical size	240	480	256
Pixel pitch (mm)	0.048	0.017	0.030
Field of view, horizontal (deg)	25	15	6
Field of view, vertical (deg)	19	11	4
Lens focal length (mm)	35	44	100
Frame rate (Hz)	30	50	30
Diagonal field of view at 100 m (m)	55	31	12
Resolution at 100 m (m/pixel)	0.139	0.039	0.030

where N_H and N_V are the number of pixels (receptors) in the horizontal and vertical dimensions, respectively. Note that both the spatial coverage and the spatial resolution depend on the distance R of a target from the camera. Therefore, the position of the camera in relation to the area of interest should be based on the camera specifications and the desired resolution and coverage. For offshore wind turbines, the area of concern is approximately 20 to 250 m above the water surface—the area that is swept by the blades of a typical tower-mounted turbine.

The size and speed of the target animals will influence the choice of camera resolution, field of view angle, and frame rate. In general, birds fly between 9 and 22 m/s (approximately 10 and 50 mph); larger birds fly faster on average than smaller birds, and flight speed will vary depending on the bird's behavior (Alerstam et al., 2007). The flight speed and the camera's field of view angle will determine how long an animal is in the field of view.

The size of the animal can be used to determine the maximum range that results in a given number of pixels on target,

$$\boldsymbol{R} = (\ell/\boldsymbol{p})(\boldsymbol{N}/2\sqrt{\tan^2\alpha_h/2 + \tan^2\alpha_v/2} \tag{4}$$

where

R	is range (distance) in meters,
ℓ	is target animal length in meters,
p	is the pixels on target along the same dimension as ℓ,
N	is the number of pixels along the diagonal of the array, and
α_H, α_V	are the horizontal and vertical view angles of the camera lens.

The frame rate is the rate at which the receptors in the phased array are scanned to form an image, specified in frames per second (fps). A standard NTSC (television) video signal is approximately 30 fps; some thermal cameras can record at higher frame rates. Higher frame rates can reduce blur when imaging fast-moving objects like birds and bats in flight.

In summary, the capabilities of thermal cameras vary, as do their costs. There will be a trade-off between imaging a wide area to maximize spatial coverage and using a high resolution to better identify individual animals, as indicated in the bottom two rows of Table 1. There is also a trade-off between a longer wavelength for minimizing atmospheric effects and a shorter wavelength for increasing the detail of detected targets. These trade-offs must be evaluated in light of the characteristics of the birds and bats expected to be present, the expected environmental conditions, and the goals of the study.

2.2. Automated processing for detection and identification

We have developed algorithms for processing thermal video to automatically detect moving animals and to extract information about them that is relevant for risk assessment. First, we will introduce the terminology used to describe the video data and processing concepts, and then we will present the details of the algorithms. An image is a two-dimensional grid of pixels. A pixel has an x,y location in the image and an intensity value. A video is a sequence of images called frames, as shown in Fig. 2. A single video frame shows what was in the camera's field of view (FOV) at a particular instant in time. When a warm object, such as an animal, is in the thermal camera's FOV, it appears as a bright blob in the image. If an object is moving, then the position of the bright blob in the image changes from frame to frame. A track is an ordered sequence of these positions, and is the unit of information that is extracted from the video.

The automated processing consists of three stages. The first stage, called video peak store (VPS), combines consecutive video frames into a single image that contains the sequence of bright blob positions. The second stage, called perceptual grouping (PG), groups blobs into tracks in the VPS image. The third stage of the processing calculates statistics that characterize each detected track.

2.2.1. Video peak store

Video peak store is a technique that has been used for studying temporal phenomena in the physical sciences (Cadmus, 1990), and more recently for studying bird flight (Zehnder et al., 2001; Gauthreaux and Livingston, 2006) using a commercial hardware device. The VPS processing described here was implemented in software. The VPS image is formed from a sequence of frames referred to as a "window" (a window of time), and the number of frames in the sequence is the window size. The window size should be chosen based on the expected time required for an animal of interest to travel across the camera's FOV, as discussed in the previous section, so that complete tracks are captured in the VPS image. The VPS image P is created by setting each pixel value to the maximum intensity value of that pixel over all of the frames in the window,

$$P(x,y) = \max_{k_0 \le k \le k_0+K}[V(x,y,k)], \tag{5}$$

where $V(x,y,k)$ is the video intensity at pixel location x,y in frame k, k_0 is the first frame in the VPS window, and K is the number of frames in the window (see Fig. 3).

The VPS image contains spatial information; a companion data structure, the VPS frame matrix, stores the frame number of each peak value stored in the VPS image. The VPS frame matrix is

$$F(x,y) + \arg\max_k[V(x,y,k)] \tag{6}$$

The background thermal energy is estimated by the mean intensity value of each pixel location over a longer window of time than that used for the VPS window. Note that the mean is calculated independently for each pixel; this is equivalent to a localized low-pass filter. The background intensity is used to generate a mask image that selects bright pixels in the VPS image that are likely to be associated with an animal,

$$M(x,y) = \begin{cases} 1 : P(x,y) > \mu(x,y) + 3\sigma(x,y) \\ 0 : \text{ otherwise} \end{cases}, \tag{7}$$

where $\mu(x,y)$ and $\sigma(x,y)$ are the mean and standard deviation of the intensity of pixel location x, y over the background estimation window. Subtracting the background makes the flight tracks of animals stand out in the resulting VPS image, as shown in Fig. 2.

Our software implementation of the VPS processing stage outputs the VPS image, the VPS frame matrix, and the VPS mask for each VPS window. The VPS windows overlap by 50% (i.e., the first frame of a VPS window is the center frame of the previous VPS window.) The VPS window data are stored in a binary file for processing by the next

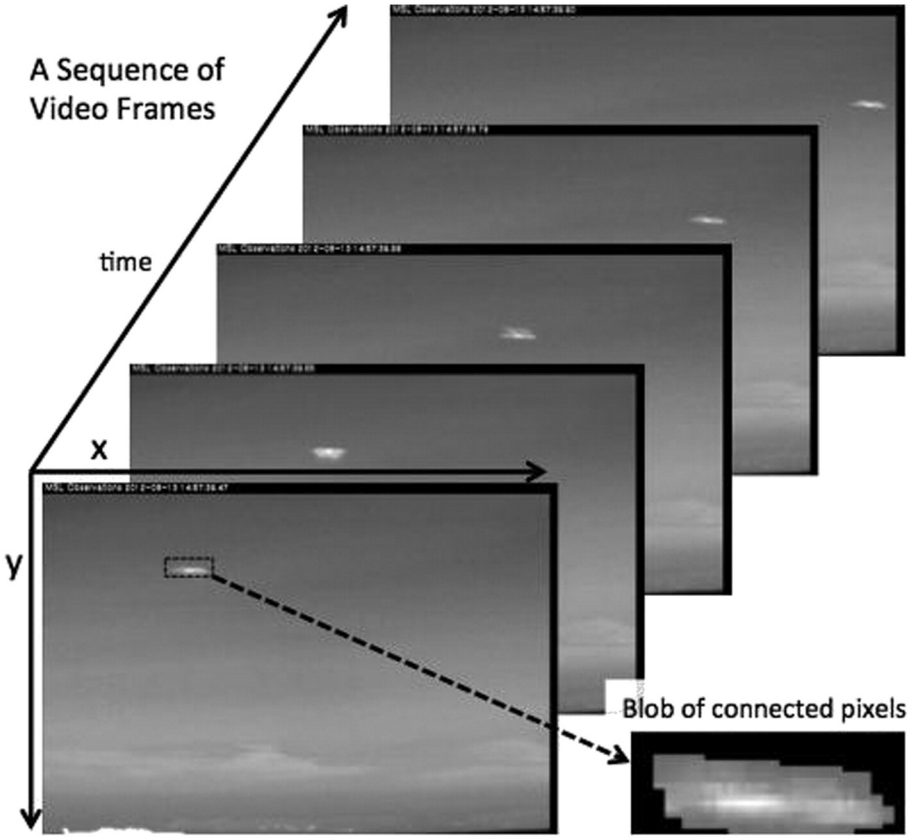

Fig. 2. A video is a sequence of images called frames. Each frame is a two-dimensional grid of pixels. A set of connected pixels is called a "blob." In this case, the blob is a gull flying across the camera's FOV from left to right.

stage, PG. The VPS processing stage also outputs the VPS image multiplied by the VPS mask for each window as a portable network graphic (PNG) file. These images can be used to visually verify the processing, and could be used as a mechanism for triggering additional sensors or to shut down a turbine when an animal is present.

2.2.2. Perceptual grouping

The PG stage automatically identifies tracks in the VPS image. Tracks appear as a sequence of blobs that form a curve (track) as shown in Fig. 4. PG is based on the way a human observer perceives relationships between regions in an image (Treisman, 1982). For example, regions that are similar in size, shape, and intensity are perceived to belong together. And regions that are connected spatially or that fall along a line are perceived to be related (Park et al., 2011).

The first step in the PG algorithm is to assemble the VPS image pixels into connected components called blobs that correspond to an animal's thermal image at a particular instant in time. The pixels in a blob are connected both spatially and temporally. The mask image $M(x, y)$ (7) is used to select the bright pixels in the VPS image. These bright pixels are the "seeds" for growing the connected pixel blobs that correspond to an animal. A blob is grown from an initial seed pixel by adding all connected pixels with the same peak frame number as the seed pixel, using the frame matrix $F(x, y)$ (6) to get the frame numbers. A blob is represented as a data structure that contains the list of member pixels, as well as attributes including the frame number, the bounding box, the complex hull, the mean intensity, and the centroid. The latter two attributes are calculated from the VPS image $P(x, y)$ (5).

The second step in the PG algorithm is to connect the blobs into tracks. Blobs are assigned to tracks in order by frame number, starting with the earliest frame. A blob is assigned to the track that is nearest to it spatially and temporally. If two different tracks are equally near, then other criteria such as the size and intensity of the track blobs are considered to break the tie. A track is considered complete when no new blobs have been added in some user-defined number of frames.

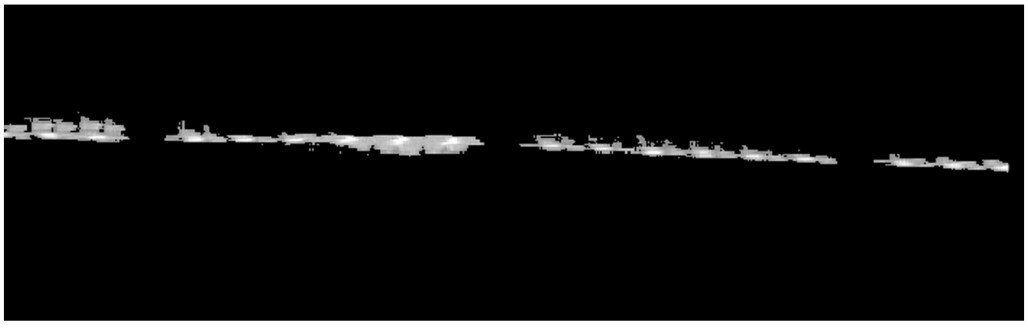

Fig. 3. A VPS image formed by combining the sequence of frames in Fig. 1 into a single image that retains the peak value of each pixel location. The background is removed by thresholding the pixel values.

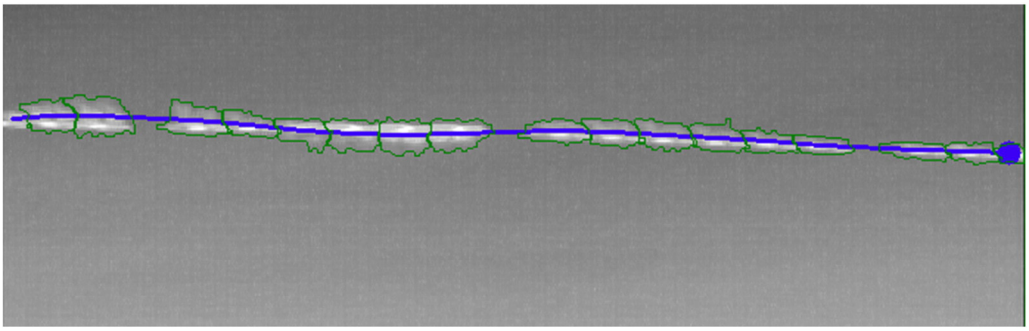

Fig. 4. A track is a connected sequence of blobs. Each blob is outlined in green and represents a snapshot of the animal at one instant in time. The purple line is the track of the animal moving across the FOV of the camera and the dot indicates where the track exits the FOV.

The end result is a set of tracks, where a track is composed of an ordered sequence of blobs.

2.2.3. Track features

The final stage of processing is the calculation of track features. The objective of this stage is to extract all of the available information from the video. To achieve this, the tracks are characterized as thoroughly as possible within the limitations of the physics of the thermal imaging process. The track features can be post-processed by subject matter experts to make inferences about the type of animals that generated the tracks and their behavior (Cullinan et al., 2015). The complete list of variables calculated for each track is given in Table 2. The physical interpretation of these features relating to animal attributes is described in the remainder of this section.

The track features are related to the following animal attributes, either directly or indirectly:

- altitude
- flight speed
- size
- shape
- temperature
- wingbeat frequency
- direction of travel
- flight track pattern.

The precision and accuracy with which these attributes can be estimated from thermal video depends on the thermal imaging camera used, the animal's physical characteristics, and the animal's distance from the camera.

2.2.3.1. Altitude. An animal's altitude has a direct bearing on the risk from wind turbines and can provide a clue about the taxonomical family of the animal. In thermal video, spatial information is limited to the two dimensions of the video frame; an animal can be located in terms of the x- and y-coordinates of the video frame, but the distance of the animal from the camera cannot be determined from the video alone. Knowledge of the camera parameters and the size and speed of animals expected to be present can be used to estimate the third dimension—distance—and from that, altitude using Eq. (4).

2.2.3.2. Speed. The absolute speed of the animal's motion cannot be calculated from the video alone; only the time that elapses as the target moves across the FOV is known. The absolute speed depends on the actual distance traveled, which is not known and depends on the distance between the animal and the camera as well as the direction of travel relative to the camera. For example, suppose a bird traveled in a straight line across the FOV, entering from the left and exiting to the right perpendicular to the camera's line of sight, as shown in Fig. 4, over the course of 2 s. The horizontal distance traveled by the bird depends on R, which is its distance from the camera,

$$d = 2R\tan\frac{\alpha_H}{2}\text{meters}. \tag{8}$$

If the camera had a horizontal view angle $\alpha_H = 24°$ and the bird was at $R = 50$ m away, then its speed would be 8.4 m/s (18.6 mph). If the

Table 2
Track Features.

Variable	Units	Description
ID	None	Unique identifier
Start time	Seconds	Elapsed video time when track starts
Start frame	Frame number	Video frame number of first blob in track
Start x,y	Image co-ordinates	Centroid location of the first blob in track
End time	Seconds	Elapsed video time when track ends
End frame	Frame number	Video frame number of last blob in track
End x,y	Image co-ordinates	Centroid location of the last blob in track
Frame span	Frames	Number of frames from start frame to end frame
Frames visible	Frames	Number of blobs in track
Sinuosity	None	Ratio of actual track length to straight line distance between start and end points
Total distance	Pixels	Length of curve through centroids of blobs in track
Speed mean, min, max, std dev[a]	Pixels/s	Speed is calculated between each successive pair of blobs in track
Blob size mean, min, max, std dev, median, Q1[b], Q3[c], MAD[d]	Pixels	Number of pixels contained in a blob
Blob width mean, min, max, std dev, median, Q1, Q3, MAD	Pixels	Horizontal extent of blobs
Blob height mean, min, max, std dev, median, Q1, Q3, MAD	Pixels	Vertical extent of blobs
Blob intensity mean, min, max, std dev, median, Q1, Q3, MAD	None	Average intensity value of pixels in each blob

[a] Standard deviation.
[b] Median absolute deviation.
[c] First quartile.
[d] Third quartile.

Table 3
Parameters used for evaluation.

Parameter	Value
VPS Window size	300 frames (10 s)
Minimum number of pixels in a blob	9
Minimum number of blobs in a track	6
Maximum distance between consecutive blobs in a track	0.10 × the size of the video frame in pixels
Maximum number of frames between consecutive blobs in a track	10

bird was at $R = 100$ m away, then its speed would be 16.8 m/s (37.2 mph). Both speeds fall within the range of possible bird flight speeds; knowledge of expected species and their average flight speed could be used to infer the possible distance range.

2.2.3.3. Size. The size of the animal in pixels can be quantified, but converting pixels to physical size cannot be done without knowing the animal's distance from the camera because the physical area represented by a pixel depends on the distance from the camera. On the other hand, if the approximate target size were known, then the target size in pixels could be used to estimate the distance of the target from the camera similar to the method discussed above for using speed to estimate distance.

2.2.3.4. Shape. The surface temperature of different parts of the animal will determine how the animal's shape appears in a thermal image. The general shape could possibly be discerned, if there were a sufficient number of pixels on target. Body parts such as the neck, beak, wings, and tail are smaller than a resolution cell (pixel) for most birds except at ranges less than about 10 m for research-grade cameras.

2.2.3.5. Temperature. The temperature of an animal can be estimated from pixel intensity, if the camera has been calibrated and the gain applied to the signal is known. Calibration is a non-trivial process, and the resulting information may not be useful for classifying targets because of the similarity in body temperature between bird species and between birds and active bats (Hock, 1951; Robinson et al., 1976; (Prinzinger et al., 1991).

2.2.3.6. Wingbeat frequency. It may be possible to estimate the wingbeat frequency from thermal video, using a method similar to that of Li and Song (2013). The maximum wingbeat frequency that can be extracted unambiguously is determined by the frame rate. For 30 fps video, the maximum frequency is 15 bps. Gulls (Laridae) and other seabirds typically exhibit wingbeat frequencies of 5 bps or less; puffins (Alcidae) have been observed to exhibit frequencies as high as 9 bps (Pennycuick, 1990). These frequencies are well within the limits of 30 fps video. The other factor that determines whether wingbeat frequency could be extracted from the video is the ability to discern the position of the wings. This factor depends on the camera resolution, the size and wingspan of the animal, the aspect angle of the animal, and its distance from the camera.

2.2.3.7. Direction of travel. An animal's general direction of travel is available from the thermal video, although there is ambiguity in the target's motion toward or away from the camera. The change in blob size over the course of the flight track could be used to estimate motion toward or away from the camera. An animal's flight direction can be calculated in world coordinates directly from the flight track, if the camera's geo-referenced position and look direction are known.

2.2.3.8. Flight pattern. The flight track can be characterized in terms of its pattern, e.g., straight, curved, swerving, etc. The pattern can be used to infer behavior, e.g., foraging or migrating, and to infer the taxonomy of an animal, e.g., diving bird or raptor. For an example of animal identification from flight track pattern, see Cullinan et al. (2015).

In summary, the information that can be extracted directly from thermal video is the flight track pattern, the general direction of travel, and, under certain conditions, the wingbeat frequency. Shape details are generally not available because the scale of bird and bat anatomical features is smaller than the camera resolution. Spatial information such as altitude and size are limited by the lack of distance information. However, knowledge of expected species body length, wingspan, and flight speed could be used in combination with knowledge of the camera parameters to estimate distance or altitude using thermal video from a single camera.

2.3. Evaluation of automated processing

The automated processing was evaluated for detection rate, false positive rate, tracking accuracy, data volume reduction, and processing time. The first two metrics are important for using the automated processing to obtain counts. The tracking accuracy is important for using the extracted track data to infer species identification and behavior. Data reduction and processing time metrics indicate the suitability of the automated processing for autonomous remote sensing applications.

The algorithms described in Section 3 were implemented in C++ using the open source computer vision library, OpenCV (Bradsky, 2000). The processing was performed on a Mac Pro with a 2.66 GHz quad-core Intel Xeon processor and 8 GB of RAM. The software parameter settings used for evaluation are given in Table 3.

The data used for the evaluation consisted of six 5-min video clips that were selected from a larger collection recorded near the Pacific Northwest National Laboratory's Marine Sciences Laboratory in Sequim, WA, during July and August 2012. The camera used was the Axsys FieldPro 5X (see Table 1 for specifications.) The video was recorded in advanced systems format files and then converted to MPEG-4 (MP4). The evaluation clips were chosen to represent a range of environmental conditions and animal activity (see Table 4.)

The detection rate and false positive rate were evaluated by comparing the output of the automated processing to annotations made by an experienced field ecologist. The annotator listed each track, its start

Table 4
Video clips used for evaluation.

Video clip	Date and time (PDT)	Conditions	Number of tracks[a]	Types of animals[b]	Animal distance from camera (m)
V1	11-Jul 20:29	40 min before sunset, no clouds	34	Swallows	~150
V2	16-Jul 21:45	After civil twilight, no clouds	11	Bats	~150
V3	17-Jul 19:50	Some clouds	10	Gulls, swallows, terns	~50–300
V4	18-Jul 19:30	Moderate cloud cover	21	Gulls, swallows, terns	~50–300
V5	18-Jul 20:36	30 min before sunset, heavy cloud cover	10	Gulls, others	~50–300
V6	13-Aug 14:57	High clouds	98	Gulls, terns, swallows	~50–300
Total			184		

[a] The number of tracks was calculated as the number of original annotated tracks corrected by subsequent review to include tracks detected by the software that were not included in the original annotations.
[b] The animal types were identified in the field during recording and then verified by viewing the recorded video.

Table 5
Detection and false alarm rates.

Video clip	Total tracks[a]	Total detections[b]	True detections[c]	Annotator missed[d]	Detection rate[e]	False positives[f]	False splits[g]	False positive rate[h]
V1	34	34	30	1	88.2	0	4	11.8
V2	11	7	7	1	63.6	0	0	0.0
V3	10	6	5	0	50.0	1	0	16.7
V4	21	19	15	6	71.4	0	4	21.1
V5	10	5	5	0	50.0	0	0	0.0
V6	98	109	87	14	88.8	5	17	20.2
Total	184	180	149	22	81.0%	6	25	17.2%

[a] Total tracks are the number of human-verified tracks, including the tracks counted in "annotator missed."
[b] Total detections are the number of tracks reported by the software.
[c] True detections were detected tracks verified by a human annotator.
[d] Annotator missed tracks were tracks detected by the software but not detected by the original annotator, then verified after processing.
[e] Detection rate is the percentage of total tracks detected by the software.
[f] False positives were tracks that do not correspond to a verified track.
[g] False splits were single tracks that were erroneously split into multiple tracks by the software.
[h] False positive rate is the percentage of total reported tracks that were not animal flight tracks or were not unique (false splits).

and end times within the video, and the type of animal. The type of animal was determined from the field notes that were taken during recording and from the appearance and flight pattern of the animal in the video, combined with knowledge of the species that occur at the recording location. The output of the automated processing was then compared to the expert annotations. Two different evaluators verified each disparity between the annotations and the automated output by reviewing the video.

The tracking accuracy was evaluated in terms of track continuity, a measure of the correctness and completeness of each reported track. For the tracks that were detected, each track was labeled as complete, partial, false split, or switched. A complete track was one in which all of the blobs belonging to that track were correctly assigned to the track, from entry into the FOV to exit from the FOV. A partial track was not complete, but was accurate in that the assigned blobs were correct. A false split was a single track that was reported as two or more tracks; in other words, during the tracking process, the algorithm incorrectly started a new track rather than assigning the next blob to the correct existing track. A switched track occurred when a track was formed incorrectly by combining the blobs of multiple animals that were in the FOV simultaneously, switching targets. The labels were determined by a visual inspection of the output images with the tracks marked, as shown in Fig. 4.

The data reduction was evaluated by comparing the size in bytes of the video per unit time to the size in bytes of the saved images and track statistics per unit time, for a given number of tracks per unit time. The size in bytes of video per unit time, or video data rate, is

$$V = H \times W \times D \times F \text{ bytes per second} \tag{9}$$

where H and W are the height and width of the video frames in pixels, D is the number of bytes per pixel, and F is the frame rate. The rate for the video used in this research was $704 \times 480 \times 2 \times 30 = 20.3$ MBps. That is the rate for the raw uncompressed video. Using MPEG-4 video compression prior to processing, our video rate was reduced to approximately 240 KBps or by about two orders of magnitude (the exact compression rate is scene dependent.)

3. Results and discussion

The detection rate varied from video to video, with the best rate achieved being 88.8% and the lowest being 50%, and with an overall detection rate of 81% (Table 5). The detection rate was worse on the videos recorded during cloudy conditions because the clouds reduced the contrast between birds and background. The missed detections tended to be birds flying at a distance from the camera, such that they were small in terms of the number of pixels contained in their image blobs, and exhibited low contrast relative to the background. The false positive rate was calculated from two different types of errors: incorrectly tracking a non-animal object or video artifact and incorrectly splitting a single track into multiple tracks. Both these types of error result in an inflated passage count. The first type of error was infrequent. The second type of error may indicate a "bug" in the processing code; this will be addressed in future work.

The tracking accuracy results are given in Table 6. Most of the detected tracks were complete in that the animal was tracked correctly from the frame in which it entered the FOV to the frame in which it exited the FOV. Partial tracks were due to low contrast, due to an animal's distance from the camera and the resulting attenuation of the animal's radiated thermal energy, and also due to the presence of clouds. Video V6 contained several VPS windows with multiple birds in view and there were four instances of incorrectly switching from one bird to another in one track. The other tracking error, false splits, was discussed in the previous section.

The data reduction and processing time results are given in Table 7. The effective data reduction depends on the number of tracks, or the average number of tracks per unit time. The size in bytes of the extracted information is approximately

$$I = N \times K + C \text{ kilobytes} \tag{10}$$

where N is the number of tracks, K is the kilobytes per track, and C is the file header size that is independent of the number of tracks. For the current implementation of the software, $K \approx 0.33$ KB/track and $C \approx 0.5$ KB.

For example, video clip V6 contained a high rate of animal activity with 109 detected tracks in 5 min (including false positives). The video size was 71,250 KB, or 300 s at a rate of approximately 240 KB/s. The track file size was approximately 34 KB. The effective data reduction was therefore about 2000:1. Even greater reductions are realized when animal activity is sparse.

Table 6
Tracking accuracy.

Video clip	True detections	Complete	Partial	False splits	Switched
V1	30	29	1	4	0
V2	7	6	1	0	0
V3	5	5	0	0	0
V4	15	10	5	4	0
V5	5	5	0	0	0
V6	87	75	12	17	4
Total	149	130	19	25	4
Percent of true detections		87%	13%	–	3%

Table 7
Data reduction and processing time.

Video clip	Video size (KB)	Number of detected tracks	Output size (KB)		Data reduction[a]		Processing time (min:s)
			VPS images	Features text file	Retain images + text	Retain text only	
V1	71,232	34	19,252	16	27.0%	0.022%	5:28
V2	71,054	7	5284	3	7.4%	0.004%	5:21
V3	70,985	6	6164	2	8.7%	0.003%	5:15
V4	70,949	19	11,803	6	16.6%	0.008%	5:17
V5	70,956	5	3292	2	4.6%	0.003%	5:25
V6	71,256	109	29,172	34	41.0%	0.048%	5:42
Total	426,431	180	74,967	63	17.6%	0.015%	32:28

[a] Retained file size as a percentage of original video size.

The average processing time was 5:25. This is longer than the actual video duration by 25 s. Because no attempt was made to optimize the code, it is reasonable to assume that the processing implementation could be improved to achieve real-time speed without significant effort. A newer computing platform, with solid-state drive technology and a faster CPU, could also be employed to achieve real-time performance.

The basic information provided by thermal video, for environmental risk assessment and monitoring, is the number of animals observed in the camera's FOV during the recording period. The number can be used to estimate average passages per unit area as

$$P_a = \frac{C}{\frac{1}{3}R^3\alpha_H\alpha_V} \text{ animals per cubic meter,} \tag{11}$$

and average passages per unit time as

$$P_t = \frac{C}{T} \text{ animals per unit of time,} \tag{12}$$

where C is the number of animals observed in the video, R is the maximum detection range from Eq. (4), and T is the length in seconds of the video. The number of observations, C, can be automatically extracted from the video using the algorithms described in this paper. The total number of passages can be estimated from the automated detections, accounting for the detection rate and false positive rate. For example, if the detection rate is estimated to be 80%, the false positive rate is 15%, and 100 tracks were detected, then the total number of animals present is estimated to be 100–15 = 85 true detections divided by 0.80 = 106 estimated passages. This level of information may be all that is needed when only one type of animal is expected to occur. A finer characterization of animal presence may be required when more than one animal type may occur, and the appropriate level of classification depends on the application.

4. Conclusion

In general, research-grade cameras will be sufficient to detect medium to large birds at ranges of about 200 m. To detect small birds and bats, a longer focal length and narrower FOV will be more effective, and bats will need to be within 100 m or closer to the camera. The taxonomic classification of animals requires close-range observations, if the classification is based on shape alone. However, other information available from thermal video, such as flight pattern and wing beat frequency, can be used to identify taxonomic groups. Environmental conditions—high humidity, rain, and clouds—may reduce the effectiveness of thermal video for observing birds and bats.

The algorithms and models we developed enable automated passage counts from thermal video. The accuracy of the passage counts in terms of the detection rate and false alarm rate was evaluated using a small, annotated data set. The results were promising: 81% of human observed tracks were detected. Missed detections were mainly due to weak thermal intensity; i.e., the radiated thermal energy of the undetected birds was not greater than three standard deviations above the mean background intensity. This may have been due to the distance of the birds from the camera. Lowering the detection threshold could improve the detection rate at the cost of a higher false alarm rate. The false alarms were few, although overcounting occurred when a single track was erroneously counted as multiple tracks. Further testing is needed to characterize the detection rate as a function of thermal intensity relative to the background and as a function of the size of the animal in pixels. These measures would be camera-independent and could be translated into detection rate as a function of distance for a particular camera, based on the camera's resolution and thermal sensitivity.

The algorithms automatically generated information useful for making inferences about the taxa of observed animals and their behaviors. Inferences can be made using knowledge about animal characteristics such as body length, wingspan, flight speed, flight patterns, and behavior. An example was provided in Section 3.3, where animal flight speed was used to estimate the range of the animal from the camera. With a sky-looking camera, range could provide valuable information about the altitude at which animal passage occurs. When a study is focused on a particular species that is protected, that species' characteristics can be combined with the camera parameters to model how the species might appear in the collected imagery, in terms of size and speed in pixels. The models can then be used to assign a probability to each observed track as to whether it belongs to the species of interest. Integrating complementary sensors such as radar or a second camera for stereovision would increase the usefulness of thermal imagery by providing distance information that could be used to convert observed size and speed from pixels into physical units.

In conclusion, the general approach described here for remotely observing bird and bat activity can be used to better understand the potential impacts to these animals from offshore wind farms and other types of development.

Acknowledgment

We would like to thank Sidney Gauthreaux for generously sharing his thermal video data with us. We would also like to thank Daniel Virden, Josh Myers, Adam Maxwell, and Kate Hall for recording hours of thermal video and detailing observations in the field.

This work was funded by the Wind and Water Power Technologies Office within the U.S. Department of Energy—Office of Energy Efficiency and Renewable Energy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ecoinf.2015.09.001.

References

Alerstam, T., Rosén, M., Bäckman, J., Ericson, P.G.P., Hellgren, O., 2007. Flight Speeds among bird species: allometric and phylogenetic effects. PLoS Biol. 5 (8), e197.

Betke, M., Hirsh, D.E., Bagchi, A., Hristov, N.I., Makris, N.C., Kunz, T.H., 2007. Tracking large variable numbers of objects in clutter. Computer Vision and Pattern Recognition. CVPR'07. IEEE Conference on. IEEE, pp. 1–8.

Betke, M., Hirsh, D.E., Makris, N.C., McCracken, G.F., Procopio, M., Hristov, N.I., Tang, S., Bagchi, A., Reichard, J.D., Horn, J.W., et al., 2008. Thermal imaging reveals significantly smaller Brazilian free-tailed bat colonies than previously estimated. J. Mammal. 89, 18–24.

Bradsky, G., 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Cadmus Jr., R.R., 1990. A video technique to facilitate the visualization of physical phenomena. Am. J. Phys. 58 (4), 397–399.

Cryan, P.M., Gorresen, P.M., Hein, C.D., Schirmacher, M.R., Diehl, R.H., Huso, M.M., Hayman, D.T.S., Fricker, P.D., Bonaccorso, F.J., Johnson, D.H., Heist, K., Dalton, D.C., 2014. Behavior of bats at wind turbines. Proc. Natl. Acad. Sci. 111 (42), 15126–15131.

Cullinan, V.I., Matzner, S., Duberstein, C.A., 2015. Classification of birds and bats using flight tracks. Ecol. Inform. 27, 55–63.

Desholm, M., Bertelsen, J., 2003. Thermal Animal Detection System (TADS). Technical Report 440. National Environmental Research Institute, Denmark.

Desholm, M., Fox, A., Beasley, P., 2004. Best practice guidance for the use of remote techniques for observing bird behavior in relation to offshore wind farms. *Technical Report*, Collaborative Offshore Wind Research into the Environment (COWRIE).

Desholm, M., Fox, A.D., Beasley, P.D.L., Kahlert, J., 2006. Remote techniques for counting and estimating the number of bird–wind turbine collisions at sea: a review. Ibis 148, 76–89.

FLIR, R. Seeing through Fog and Rain with a Thermal Imaging Camera Retrieved from http://www.flir.com/uploadedfiles/Eurasia/MMC/Tech_Notes/TN_0001_EN.pdf.

Gauthreaux, S.A., Livingston, J.W., 2006. Monitoring bird migration with a fixed-beam radar and a thermal-imaging camera. J. Field Ornithol. 77, 319–328.

Hock, R.J., 1951. The metabolic rates and body temperatures of bats. Biol. Bull. 101, 289–299.

Hristov, N.I., Betke, M., Kunz, T.H., 2008. Applications of thermal infrared imaging for research in aeroecology. Integr. Comp. Biol. 48, 50–59.

Kunz, T.H., Arnett, E.B., Cooper, B.M., Erickson, W.P., Larkin, R.P., Mabee, T., Morrison, M.L., Strickland, M.D., Szewczak, J.M., 2007. Assessing impacts of wind–energy development on nocturnally active birds and bats: a guidance document. J. Wildl. Manag. 71.

Li, W., Song, D., 2013. Automatic bird species detection using periodicity of salient extremities. Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, pp. 5775–5780.

Liechiti, F., Bruderer, B., Paproth, H., 1995. Quantification of nocturnal bird migration by moonwatching: comparison with radar and infrared observations. J. Field Ornithol. 66, 457–468.

Lindeboom, H., Kouwenhoven, H., Bergman, M., Bouma, S., Brasseur, S., Daan, R., Fijn, R., De Haan, D., Dirksen, S., Van Hal, R., et al., 2011. Short-term ecological effects of an offshore wind farm in the Dutch coastal zone; a compilation. Environ. Res. Lett. 6 (3), 035101.

Park, M., Brocklehurst, K., Collins, R.T., Liu, Y., 2011. Translation–symmetry-based perceptual grouping with applications to urban scenes. Computer Vision–ACCV 2010. Springer Berlin Heidelberg, pp. 329–342.

Pennycuick, C.J., 1990. Predicting wingbeat frequency and wavelength of birds. J. Exp. Biol. 150, 171–185.

Prinzinger, R., Preßmar, A., Schleucher, E., 1991. Body temperature in birds. Comp. Biochem. Physiol. A Physiol. 99, 499–506.

Robinson Willmott, J.C., Forcey, G., Kent, A., 2013. The relative vulnerability of migratory bird species to offshore wind energy projects on the Atlantic outer continental shelf: an assessment method and database. Final Report to the U.S. Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs (OCS Study BOEM 2013-207. 275 pp.).

Robinson, D.E., Campbell, G.S., King, J.R., 1976. An evaluation of heat exchange in small birds. J. Comp. Physiol. B 105, 153–166.

Suryan, R.M., Sato, F., Balogh, G.R., Hyrenbach, K.D., Sievert, P.R., Ozaki, K., 2006. Foraging destinations and marine habitat use of short-tailed albatrosses: a multi-scale approach using first-passage time analysis. Deep-Sea Res. II Top. Stud. Oceanogr. 53 (3), 370–386.

Theriault, D., Wu, Z., Hristov, N., Swartz, S., Breuer, K., Kunz, T., Betke, M., 2010. Reconstruction and analysis of 3D trajectories of Brazilian free-tailed bats in flight. Tech. Rep. CS Department, Boston University.

Theriault, D.H., Fuller, N.W., Jackson, B.E., Bluhm, E., Evangelista, D., Wu, Z., Betke, M., Hedrick, T.L., 2014. A protocol and calibration method for accurate multi-camera field videography. J. Exp. Biol. 217 (11), 1843–1848.

Treisman, A., 1982. Perceptual grouping and attention in visual search for features and for objects. J. Exp. Psychol. Hum. Percept. Perform. 8 (2), 194.

Zehnder, S., Akesson, S., Liechiti, F., Bruderer, B., 2001. Nocturnal autumn bird migration at Falsterbo, South Sweden. J. Avian Biol. 32, 239–248.