

Article

Feasibility Analyses of Real-Time Detection of Wildlife Using UAV-Derived Thermal and RGB Images

Seunghyeon Lee ^{1,2} , Youngkeun Song ^{1,2} and Sung-Ho Kil ^{3,*}

¹ Department of Landscape Architecture, Graduate School of Environmental Studies, Seoul National University, Seoul 08826, Korea; shlee5598@snu.ac.kr (S.L.); songyoung@snu.ac.kr (Y.S.)

² Integrated Major in Smart City Global Convergence, Seoul National University, Seoul 08826, Korea

³ Department of Ecological Landscape Architecture Design, College of Forest and Environmental Sciences, Kangwon National University, Chuncheon 24341, Korea

* Correspondence: sunghokil@kangwon.ac.kr

Abstract: Wildlife monitoring is carried out for diverse reasons, and monitoring methods have gradually advanced through technological development. Direct field investigations have been replaced by remote monitoring methods, and unmanned aerial vehicles (UAVs) have recently become the most important tool for wildlife monitoring. Many previous studies on detecting wild animals have used RGB images acquired from UAVs, with most of the analyses depending on machine learning–deep learning (ML–DL) methods. These methods provide relatively accurate results, and when thermal sensors are used as a supplement, even more accurate detection results can be obtained through complementation with RGB images. However, because most previous analyses were based on ML–DL methods, a lot of time was required to generate training data and train detection models. This drawback makes ML–DL methods unsuitable for real-time detection in the field. To compensate for the disadvantages of the previous methods, this paper proposes a real-time animal detection method that generates a total of six applicable input images depending on the context and uses them for detection. The proposed method is based on the Sobel edge algorithm, which is simple but can detect edges quickly based on change values. The method can detect animals in a single image without training data. The fastest detection time per image was 0.033 s, and all frames of a thermal video could be analyzed. Furthermore, because of the synchronization of the properties of the thermal and RGB images, the performance of the method was above average in comparison with previous studies. With target images acquired at heights below 100 m, the maximum detection precision and detection recall of the most accurate input image were 0.804 and 0.699, respectively. However, the low resolution of the thermal sensor and its shooting height limitation were hindrances to wildlife detection. The aim of future research will be to develop a detection method that can improve these shortcomings.

Keywords: thermal sensing; unmanned aerial vehicle; object-based animal detection; instant and automated detection; mixed image analysis; wildlife monitoring; multiple height shooting



Citation: Lee, S.; Song, Y.; Kil, S.-H. Feasibility Analyses of Real-Time Detection of Wildlife Using UAV-Derived Thermal and RGB Images. *Remote Sens.* **2021**, *13*, 2169. <https://doi.org/10.3390/rs13112169>

Academic Editor: Lefei Zhang

Received: 20 April 2021

Accepted: 28 May 2021

Published: 1 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For wildlife detection and monitoring, traditional methods such as direct observation [1] and capture–recapture have been carried out for diverse purposes [2]. However, these methods require a large amount of time, considerable expense, and field-skilled experts [3,4] to obtain reliable results. Furthermore, performing a traditional field survey can result in dangerous situations, such as an encounter with wild animals. Remote monitoring methods, such as those based on camera trapping [5], GPS collars [6], and environmental DNA sampling [7], have been used more frequently, mostly replacing traditional survey methods, as the technologies have developed. Camera-trapping methods can track the life cycle of animals at the nest level. Camera networks can be created by installing multiple cameras, and high-quality data can be acquired across the region of interest [8]. However,

these methods still have limitations, such as the inability to cover an entire region [9] or detect individual targets [10].

As a means of overcoming such limitations, unmanned aerial vehicles (UAVs) are becoming popular for conducting wildlife censuses [11]. The main benefits of UAVs are that they can detect animals remotely, covering a wider region with fine spatiotemporal resolution [11,12]. In addition, UAVs can be used to investigate hard-to-access or dangerous areas [13]. However, UAVs clearly have some limitations. The study site, and the UAV flying height and speed, can limit the ability to detect small animals [14] and targets in dense forest [15] and to track fast moving animals [16]. The weather can also limit UAV operations [17], and flight time is constrained by the battery [18]. Although detailed detection data using UAVs are somewhat lacking, some studies have used UAVs to detect terrestrial mammals [19–23], marine [24] mammals, birds [25], and reptiles [26].

The most common type of data acquired by UAVs is RGB images. Using these images, manual counting—of elephant seals [27] and Antarctic shag [28], for example—provides the most accurate results. Automated detection studies mainly used machine-learning and deep-learning (ML–DL) methods for wildlife detection [19]. Cattle [20–22], wild animals in the savannah [23], and various mammals [11] have been targets. The studies targeting cattle and other mammals used convolutional neural network (CNN) deep-learning models, and the study that detected wild animals in the savannah used an exemplar support vector machine (ESVM) machine-learning model. ML–DL methods provide relatively accurate results, but at least 1000 images are required to develop a proper detection model for specific species [11,21]. Moreover, producing training data and training the model require a lot of time. To detect mammal species, one study [11] spent 4 days training the machine-learning model. Therefore, such ML–DL methods using big data cannot be used in the field for real-time acquisition. Furthermore, the detection models developed are fitted to training images, so they cannot detect different species or targets on different types of land cover.

Instead of RGB images, by changing the existing camera to a thermal camera or mounting an additional camera, thermal images can be acquired by a UAV. The development of the thermal sensor technology and reduction in sensor prices have attracted the interest of wildlife researchers [29]. Using a thermal camera, homeothermic animals can be detected based on the temperature difference between their bodies and the surrounding environment. This new technology has already been used to detect animals such as hippopotami [30], seals [24], deer [31], and cattle [22]. Furthermore, research on detecting marine mammals [24] and avian species [25] has been conducted. Additionally, the thermal sensor feature of detecting infrared radiation makes it possible to locate animals at night [11] and camouflaged targets [29]. Although the technology and data are new, the same ML–DL methods are typically used for animal detection [32]. However, new methods, such as isoline creation [30] and using two images shot at different times to identify changes [31], have been suggested. The former method improves the detection rate by considering the degree of growth and overlapping conditions, but it requires preprocessing for geo-referencing and image merging and clipping. The latter method, unlike other methods used for thermal sensing research, uses images shot at a height greater than 1 km, and this feature has the strength of being able to cover a very broad region. However, preprocessing is still needed. This data preprocessing limits the use of thermal cameras for real-time wildlife detection in the field [33].

Another limitation of previous studies that used thermal cameras is that most only used thermal images for detection; however, some detection research targeting avian species [25] and white-tail deer [34] used thermal and RGB images simultaneously, which resulted in higher detectability than using thermal images alone. However, in these studies, an RGB camera and thermal camera were mounted together on the UAV. This method increases research costs, and using the two types of datasets together mandates an additional preprocessing step and additional time to match their data properties.

This paper proposes a new method for detecting animals. There were three main objectives, to address the limitations of previous research:

(1) Reduce the animal detection time

The main limitation of previous animal detection methods is that they cannot not be applied in the field in real time. ML–DL-based methods need an enormous number of training images, and it takes a long time to train the detection model. Methods using thermal images require preprocessing to detect animals. To address these limitations, the proposed method can detect animals based on single images, and image preprocessing is simplified.

(2) Enable detection in more environments

ML–DL-based methods are only suitable for certain species and land cover types or environments. To improve detection versatility, the proposed method considers target size and surface temperature when detecting animals. Theoretically, the method can be adapted to all homeothermic animals if the body size and surface temperature are known. Here, we focused on detecting mid-sized animals (alpaca).

(3) Use thermal and RGB images acquired from the same thermal camera

When a detection method needs both thermal and RGB images, separate thermal and RGB cameras are used. However, any thermal camera can save thermal and RGB images simultaneously, and the centroid is the same because the shooting time is the same. Therefore, by modifying the distortion caused by focal length, shooting area, and spatial resolution, thermal and RGB images can be used simultaneously for research without the requirement of two cameras [35].

The main goal of this study was to develop an automated method for detecting animals using a thermal image dataset, to apply it under in situ conditions in real time, and to achieve similar detection ability to previous methods. The fastest detection time was 0.033 s, the maximum detection precision was 0.804, and the detection recall rate was 0.699.

2. Study Site and Data

2.1. Study Site

An animal farm (37.827° N, 127.882° E) in the middle of a natural forest in Hongcheon, Republic of Korea, was used as the study site for data collection (Figure 1). To determine the animal species and their locations, the UAV operated over the entire farm. Through this process, the distribution of land cover was also confirmed. The major species on the farm is *Vicugna pacos* (alpaca), so these animals were mainly used to develop the detection and analysis method. The farm also has a few *Cervus nippon* (sika deer), *Struthio camelus* (ostrich), and *Camelus bactrianus* (camel). The barns for each species are located on grassland or bare land, and they are mainly moving on those land covers. The area of the farm is approximately 12.02 ha, and the main cover type is forest (50%), followed by grassland (35%). The remaining contributors to land cover comprise artificial structures such as roads and buildings, and bare land. The minimum and maximum elevations on the farm are 450.56 and 512.00 m, respectively.

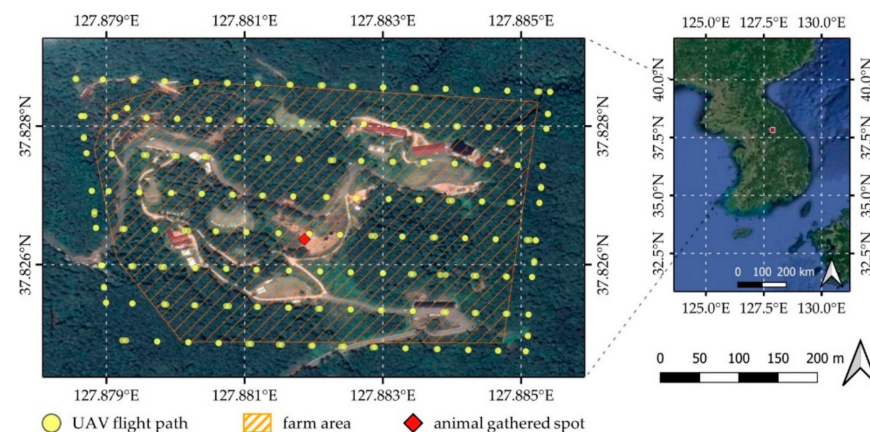


Figure 1. Overview of the study site.

2.2. Data Acquisition

UAV flights were conducted using a MATRICE 210 UAV (DJI, Shenzhen, China), and the thermal camera was a FLIR ZENMUSE XT2 (DJI). The thermal camera has both an RGB sensor and a thermal sensor, and images are captured by both sensors at different resolutions. Each RGB image contains 4000×3000 pixels, and each thermal image contains 640×512 pixels. The spatial resolution of each RGB image at 25 m above the ground is 0.59 cm/pixel, whereas the resolution of each thermal image is 2.24 cm/pixel. Due to the increased focal length of the thermal sensor, each thermal image covers a narrower region [36].

The data were acquired on 25 November 2020. In Korea, November is considered to fall within the winter season, and snow typically falls from the middle of November. Although snow cover provides advantages, in the sense that a lower land-surface temperature is beneficial in automated animal detection and photographs can show not only the animals but also their tracks, thereby improving detection rates [37], a lack of adequate snow cover can inhibit animal detection, requiring the images to be filmed again [38]. Therefore, the shooting date was selected to occur when the air and land surface temperatures were low and there was no snow cover. This decision maximized the temperature difference between the targets and land cover types and facilitated more accurate detection of animals. Furthermore, by shooting images around noon, the shadow size of individual targets was minimized, which reduced the possibility of error from shadows.

After a programmed drone flight over the entire study site, the drone was controlled manually to capture the locations of the main target animals (alpaca). After finding a spot, 26 images were acquired from heights of 25–275 m above the ground at 10-m intervals to aid the development of a method to be used under various circumstances. The body lengths of the main target animals range from 80 to 100 cm when fully grown, and they have various fur colors, including black, gray, white, dark brown, and light brown. Based on the UAV results, the targets were sorted into four categories according to their visible condition. The category “isolated” indicated that the target stood alone, not touching any other target or obstacle. “Bordering” meant that two targets were touching each other, and “overlapping” meant that the targets’ body parts were crossing each other’s. “Partial” indicated that the target was partly visible at the edge of the image (Figure 2).

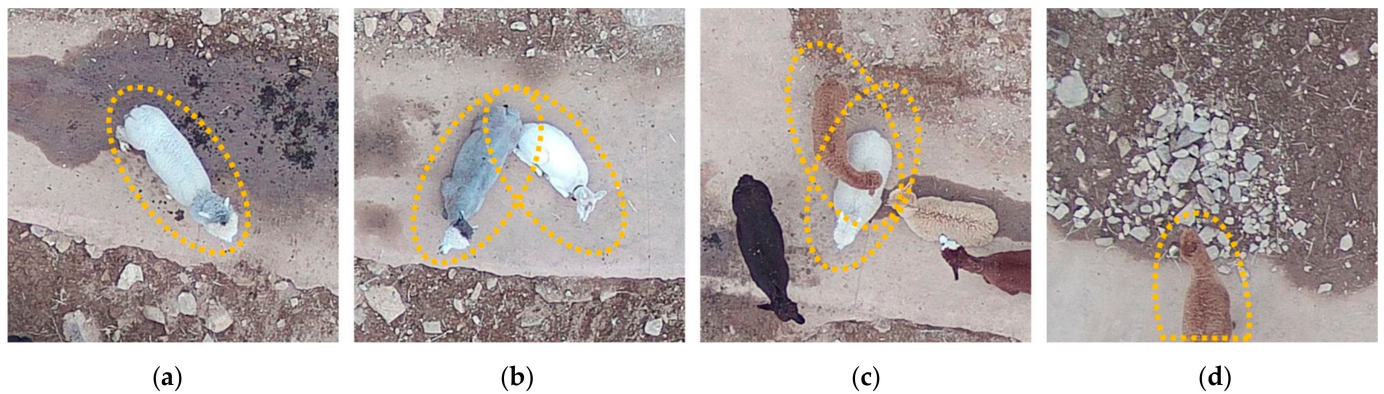


Figure 2. Target shape categories: (a) isolated, (b) bordering, (c) overlapping, and (d) partial.

2.3. Data Preprocessing

As the outputs of the XT2 sensors have different pixel sizes, spatial resolutions, and coverage areas (Figure 3), they need to be modified to have the same properties. Furthermore, to acquire accurate results, temperature correction of the thermal images and masking of non-target regions are required.

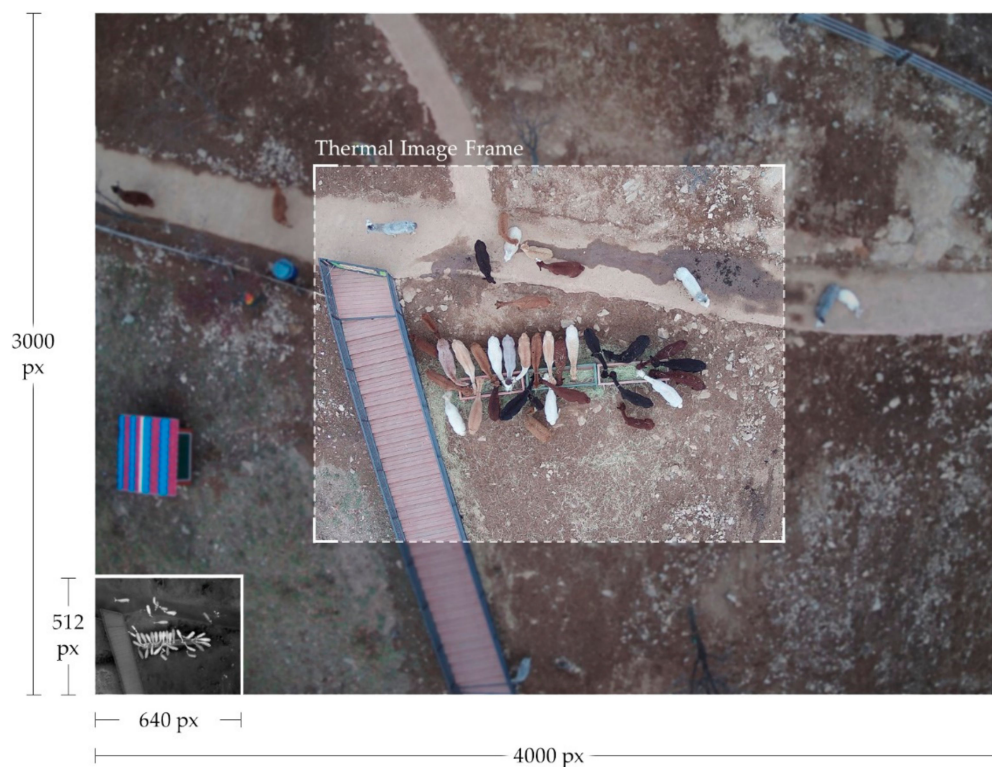


Figure 3. Representative outputs from the thermal sensor. The small image at the bottom-left is the thermal image (640×512 px), and the larger image is the RGB image (4000×3000 px). The coverage area of the thermal image is marked by the white rectangle in the RGB image.

2.3.1. RGB Lens Distortion Correction and Clipping

Due to the difference in focal length, the distortion in the images also differs [39]. The RGB sensor of XT2 has a focal length of 8 mm, but the thermal sensor has a focal length of 19 mm. When the focal length is shorter, the image is subject to barrel distortion compared with an image with longer focal length [40]. Therefore, to use the thermal and RGB images together, we had to correct the distortion in the RGB images. Python and the OpenCV2

library [41] were used for this purpose. After correction, the corrected RGB images were clipped and rescaled to have the same coverage as the thermal images (Figure 4).

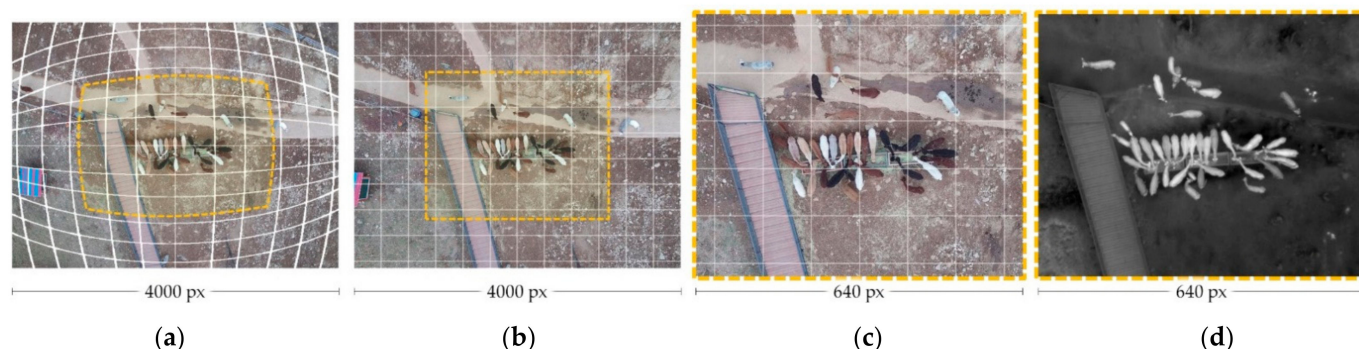


Figure 4. Example of focal length distortion and correction. (a) Original RGB image (distorted), (b) corrected RGB image, (c) clipped RGB image, and (d) original thermal image.

2.3.2. Thermal Image Correction by Fur Color

Although the body temperature of the target animals is the same across individuals, the surface temperature can differ because of the fur color. The surface temperatures of animals with brighter fur were lower [42] because of higher reflectance [43]. Surface temperature differences can cause errors in the detection process and must be corrected for.

The pixel value of each RGB channel is needed to identify bright targets. Based on our measurements, we found that the surface temperature of white animals was approximately 25% lower than that of animals with darker fur. Therefore, the pixels of thermal images located at the same locations as white pixels from RGB images were adjusted to have higher values (Figure 5).

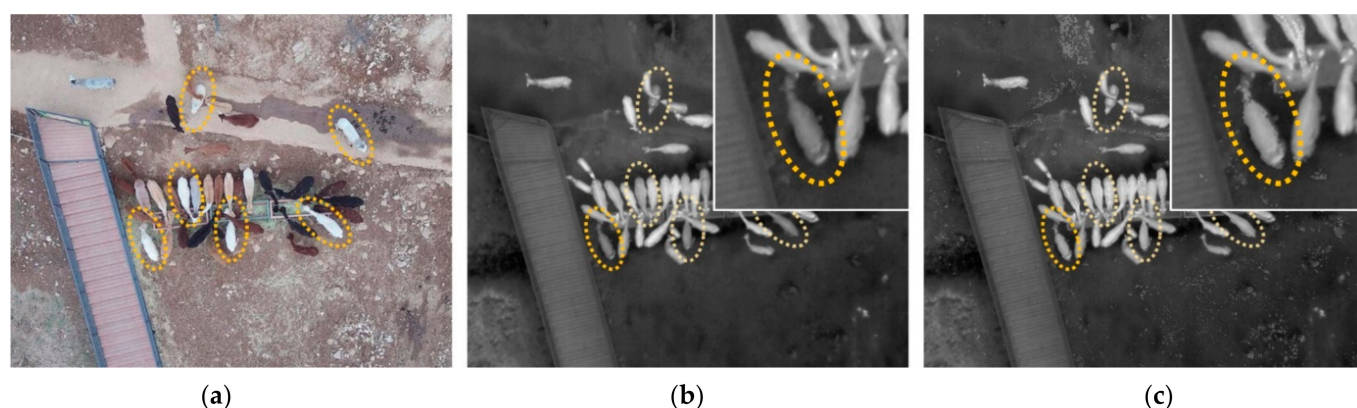


Figure 5. Example of thermal image correction (fur color). (a) Original RGB image, (b) original thermal image, and (c) corrected thermal image.

2.3.3. Unnatural Object Removal

The principle of animal detection using thermal images is to locate spots where the temperature is different, because homeothermic animals always have the same body temperature and this consistency creates a temperature gap between animals and their surrounding environment. However, artificial structures, e.g., buildings and roads, have a much higher surface temperature compared with animals or natural surfaces. Therefore, when these types of artificial land cover are included in a thermal image, numerous errors in animal detection occur [31]. To eliminate this error, artificial structures should be masked.

However, it is difficult to tell which parts of the image should be removed, since one of main purposes of this study was to develop a method that can be used for instant

detection under in situ conditions, and pursuing this objective limited the time available to analyze images and locate artificial structures. Therefore, as an alternative to artificial cover detection, the unnatural color masking method was used. Fortunately, more than half of the artificial structures at the study site have unnatural colors, such as vivid red, vivid blue, and vivid orange (Figure 6). As when correcting the temperature for fur color, for this step, temperature values were removed according to pixel color. Many possible errors can be prevented by removing these high-temperature artificial structures.

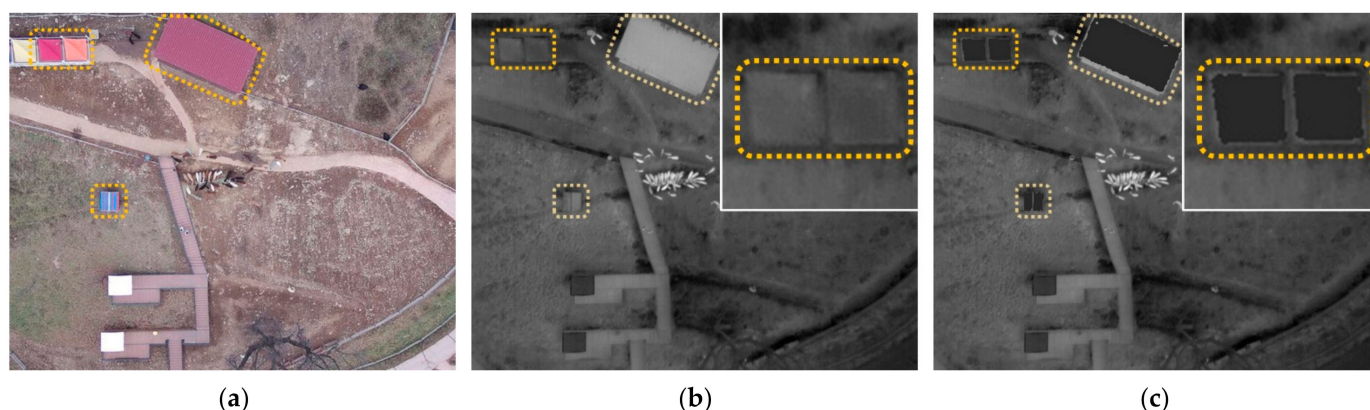


Figure 6. Example of thermal image correction (unnatural). (a) Original RGB image, (b) original thermal image, and (c) corrected thermal image.

3. Methods

Our method requires both thermal and RGB images but especially thermal images, as these contain more useful information.

The open-source programming language Python was used in the Google Colab [44] environment to develop the proposed method. Google Colab, a cloud service based on Jupyter Notebooks, executes Python code using both CPU and GPU resources, thus enabling quantitative analysis on a scale that exceeds the limitations of personal computers. The main functions of the proposed method are Sobel edge creation [45] and contour drawing. OpenCV2, an optimized computer vision library, was used for image processing.

The automated detection results obtained using the proposed method were categorized based on shooting height and target shape, i.e., isolated, bordering, overlapping, or partial.

3.1. Sobel Edge Detection and Contour Drawing

Sobel edge creation refers to a method that finds edges simply. This gradient operator works vertically and horizontally [46]. When the difference in pixel value is larger, the Sobel edge has a higher value. By combining the vertical and horizontal Sobel edges, a biaxial Sobel edge can be made. This biaxial Sobel edge was used to draw the binary contours. After applying a threshold to the biaxial Sobel, the segmented image was used for contouring. At the same time as contours were drawn, the centroid point of each contour was marked on the images (Figure 7). The accuracy of the contours was high, but some were wrongly drawn around non-target objects, such as stones, wet soil, and artificial structures; therefore, to eliminate these false-positive results and obtain accurate results, the contours had to be sorted.

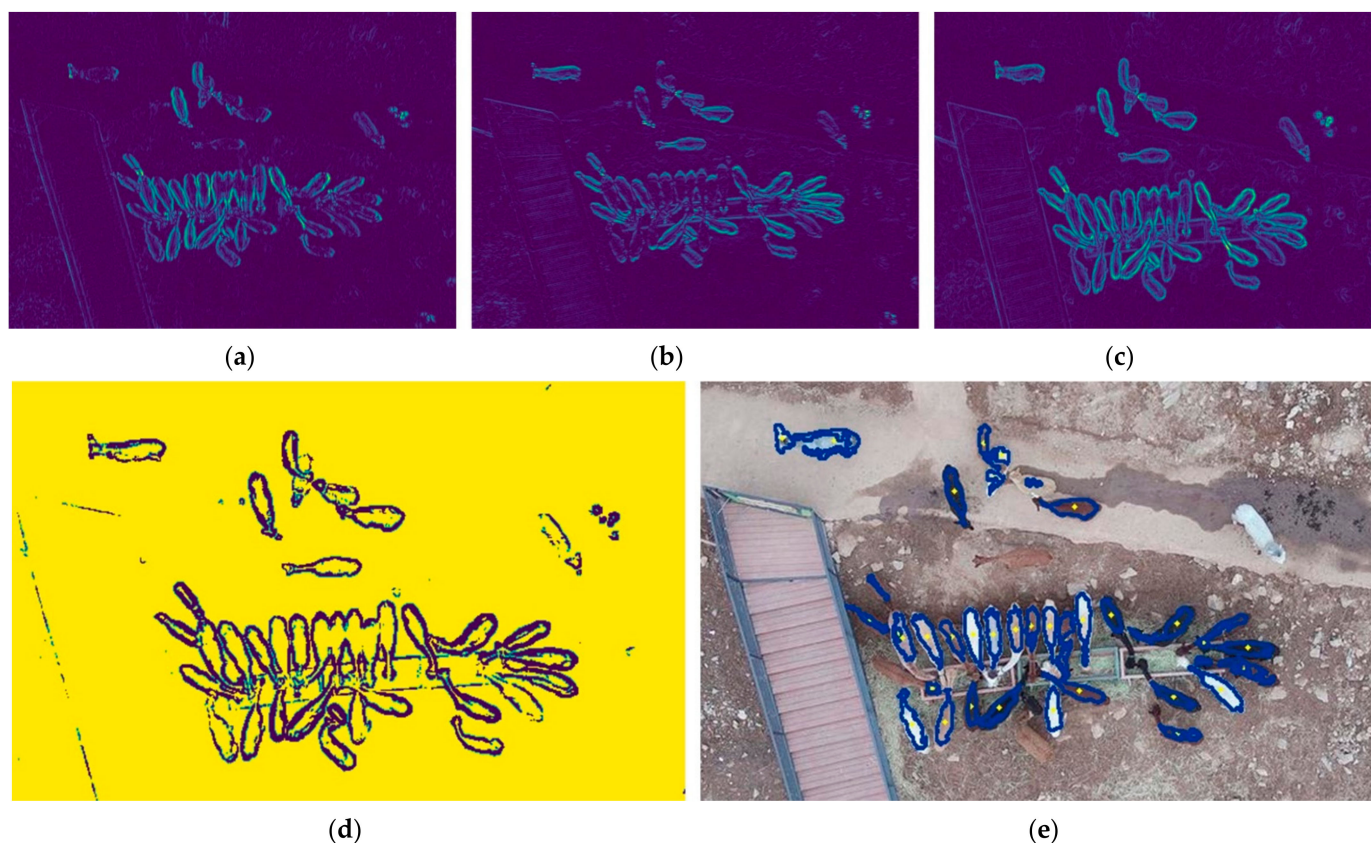


Figure 7. Example of Sobel edge detection and contour drawing. (a) Sobel x-axis outcome, (b) Sobel y-axis outcome, (c) Sobel biaxial outcome, (d) thresholded image, and (e) contoured image.

3.2. Object Detection and Sorting

To eliminate wrongly drawn contours, size–temperature filtering was used. The mean body length of the target animal was approximately 0.9 m, and the top-view area was approximately 4500 cm². However, the body of the animal is fully covered with thick, curly fur, so its body heat is not shown clearly in the thermal image, making the animal look smaller than normal. Hence, the area filter was set to detect contours smaller than 3500 cm² and larger than 100 cm². The minimum criterion was set much smaller than the common size of the target to find segmented body parts such as overlapping or partial targets. Additionally, the size of drawn contours can be small because of the animal's body shape. Therefore, to obtain a high probability of animal detection, the area filter was set with a large range.

For contours sorted by the area filter, the centroid temperature filter was used again. The maximum and minimum body temperatures of the targets 25 m above the animal's body were nearly 20 °C and 10 °C, respectively. Therefore, the filtering option was set to find contours warmer than 9 °C to ensure that every target was filtered. In addition, the temperature also changes with changes in shooting height. To minimize this error, we corrected the temperature by height. The shooting height and maximum temperature of the targets are linearly related (Figure 8). Temperature filtering was adapted using Equation (1).

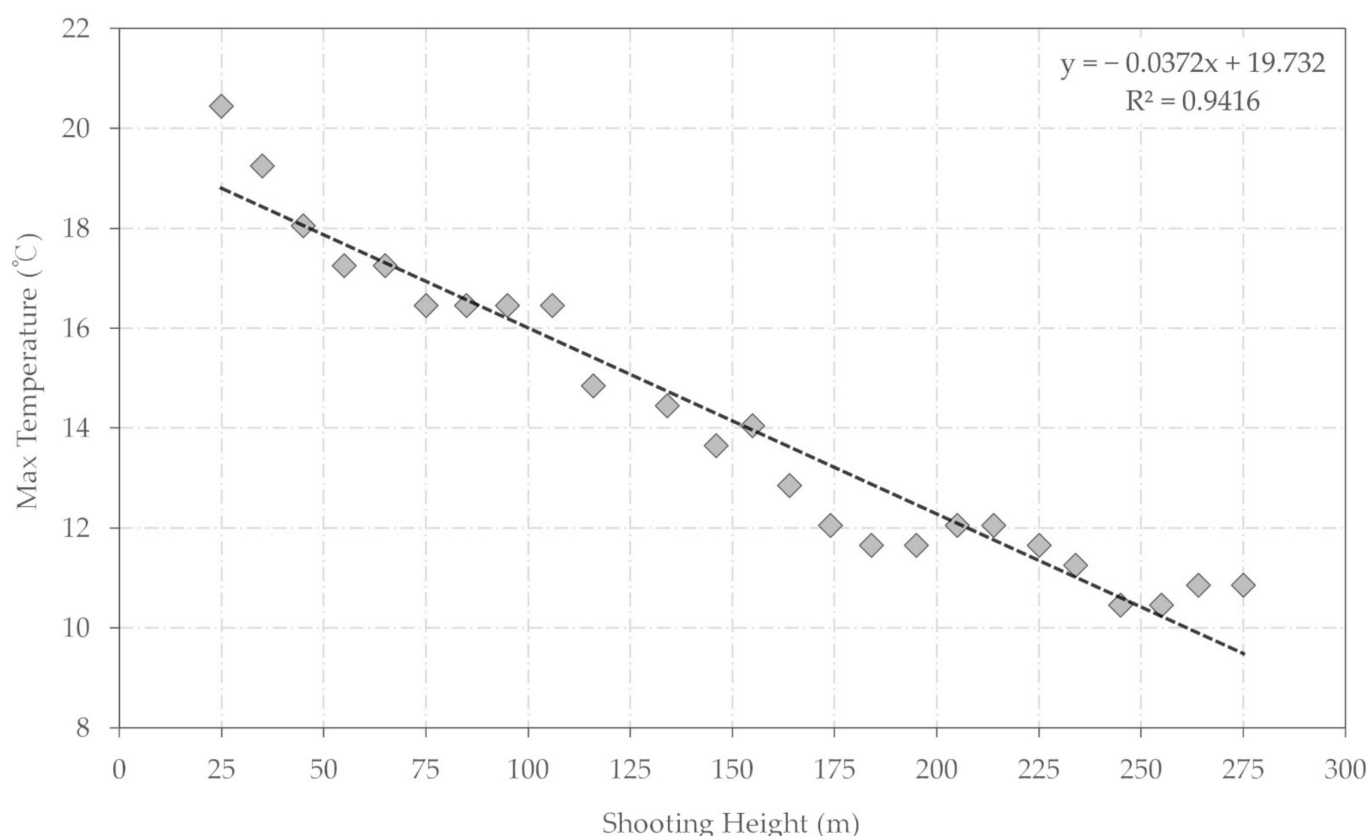


Figure 8. Regression analysis of maximum temperature and shooting height.

The main target animal in this study was the alpaca. Therefore, size-temperature filtering was designed and adapted to this species. However, this object detection and sorting method can be adapted to target other species by changing the filter criteria.

$$t_{corrected} = -0.0372 \cdot \text{height} + 19.732 \quad (1)$$

3.3. Input Images Generation

As mentioned previously, six kinds of input images were used for the automated detection method (Figure 9). These input images were generated to enhance the detection ability, shorten the detection time, and determine which type of input image produces the most accurate detection performance. The six kinds of input images were corrected RGB images, original thermal images, thermal images corrected for fur color, thermal images with masked unnatural colors, corrected RGB images \times original thermal images, and corrected RGB images \times all correction-applied thermal images.

The thermal and RGB images were processed using contour and centroid generation, size-temperature filtering, a target counting process after Sobel edge creation, and image binarization. These images were combined after image binarization, and each combined image could be used to generate contours corresponding to those of the two kinds of images. Therefore, combined images allowed for a more accurate detection ability.

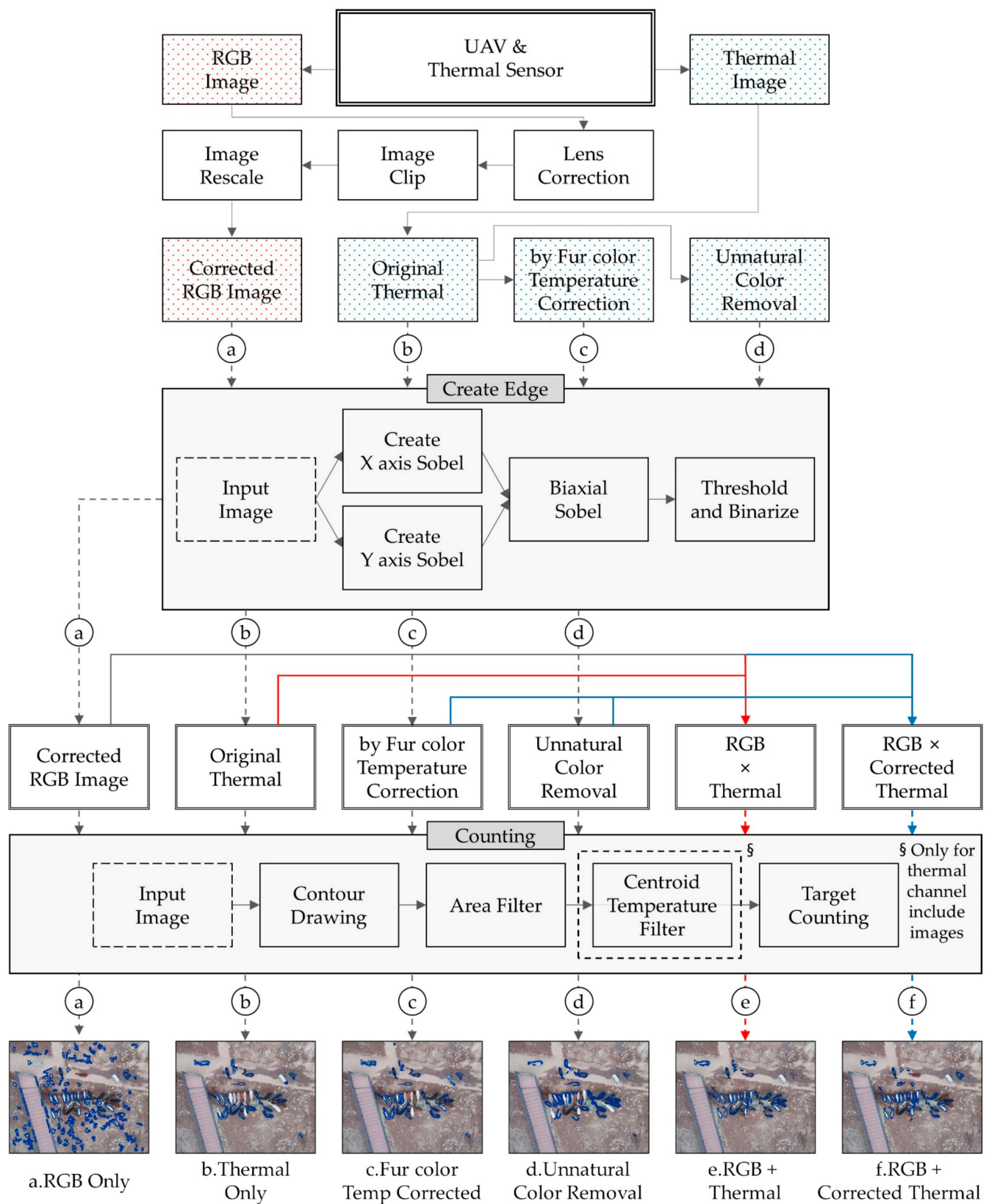


Figure 9. Flowchart of the detection method for all image types.

4. Results

The automated detection results obtained using the proposed method were categorized based on shooting height and target shape, i.e., isolated, bordering, overlapping, or partial. The detection recall, precision, and time were also analyzed.

The detection results were assessed based on the detection precision and detection recall rate (Figure 10). The detection precision was calculated as the number of real animals among the automatic detections divided by the total number of detections. The detection recall rate was the number of real animals among the automatic detections divided by the number of animals in the image. These two values have the same range, from 0 to 1, and higher values indicate higher detection ability.







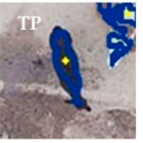
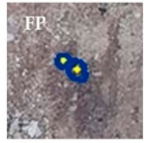

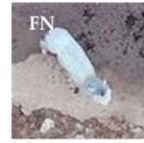
		Manually counted		Detection Precision	Detection Recall
		True	False		
Automatically counted	True	 True-positive	 False-positive	$\frac{TP}{TP + FP}$ 	$\frac{TP}{TP + FN}$ 
	False	 False-negative	 True-negative	 + 	 + 

Figure 10. Calculation of detection precision and recall rate.

To compare the detection precision and detection recall rates of the six kinds of images, the number of targets in each image was counted manually (Figure 11), and targets were labeled according to their shape category (i.e., isolated, bordering, overlapping, or partial). The number of targets in each of the 26 individual original images was about 40. However, for every type of input image, the numbers of targets detected tended to decrease with increased shooting height. At shooting heights greater than 100 m, fewer than 10 targets could be detected in each type of image, and at heights greater than 125 m, fewer than five targets could be detected.

Based on the results of manual and automatic counting, the detection precision and recall rate were evaluated. As the detection precision decreased dramatically above a height of 100 m, we focused on detection results at shooting heights lower than 100 m (Table 1). The total number of targets was 316, consisting of 56 isolated targets, 243 bordering targets, 17 overlapping targets, and three partial targets. Of the 316 targets, 5 were ostriches.

When only RGB images were used for detection, the detection recall rate was 0.367, and the detection precision was 0.013. RGB images cannot be subjected to temperature filtering. Therefore, false-positive detection results such as soil, rocks, roofs, and roads could not be eliminated. This uncertainty in sorting led to the poor detection recall result.

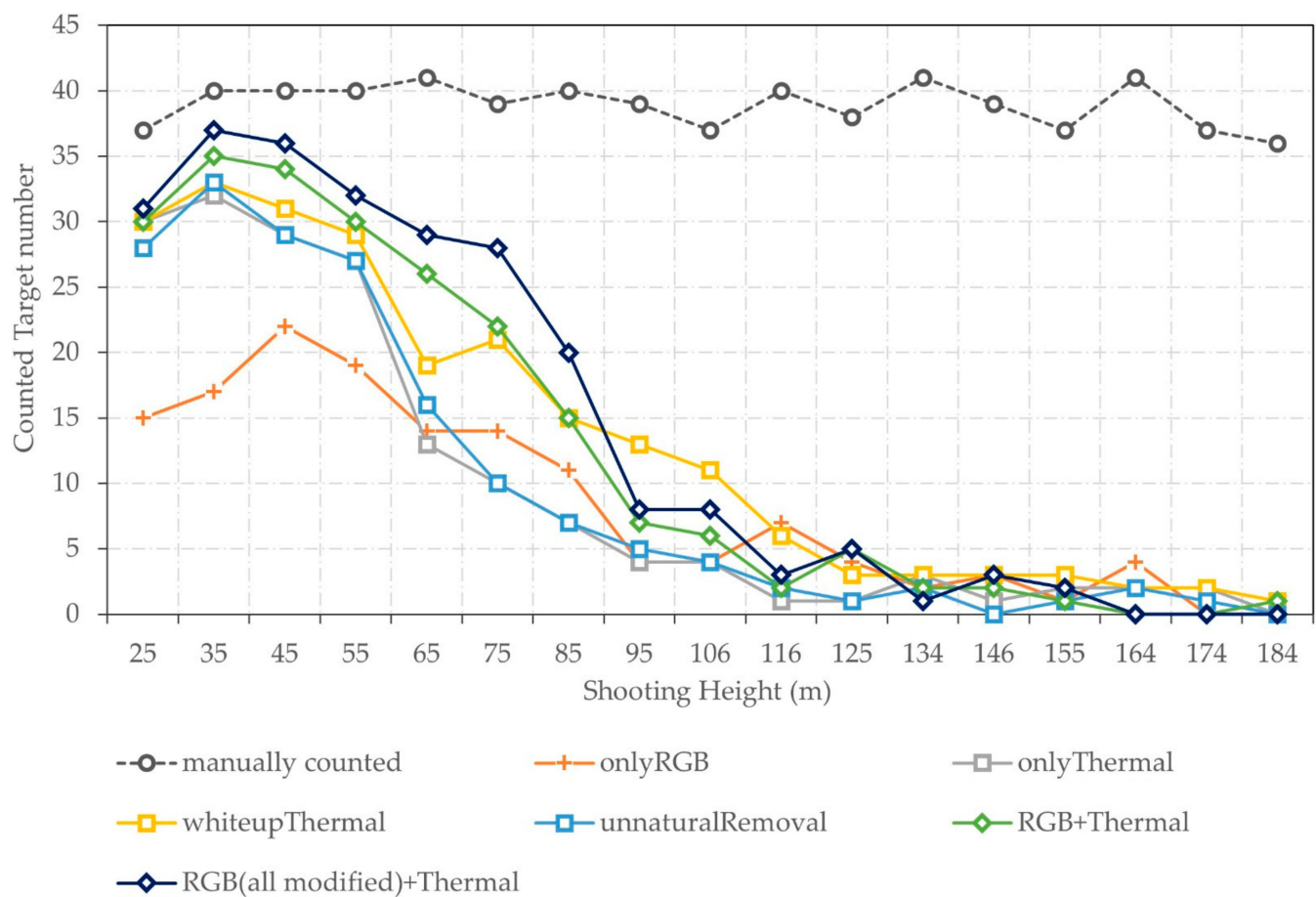


Figure 11. Detection results by height and type of input image.

Table 1. Total numbers of targets detected and the detection precision and recall rates (for images with a shooting height of less than 100 m).

	Isolated	Bordering	Over-Lapping	Partial	Detected	Error	Total Count	Detection Precision	Detection Recall
Manual count	56	243	17	3	316				
Corrected RGB only	17	95	0	4	116	9099	9215	0.013	0.367
Thermal only	32	120	0	0	152	40	192	0.792	0.481
Corrected for fur thermal	42	149	0	0	191	108	299	0.639	0.604
Unnatural color removal thermal	31	122	2	0	155	38	193	0.803	0.491
Corrected RGB + Thermal	28	168	2	1	199	795	994	0.200	0.630
Corrected RGB + Corrected thermal	34	183	3	1	221	54	275	0.804	0.699

When the input image contained thermal information, the detection precision and recall rate were higher. In particular, compared with the RGB-only detection results, the precision increased by at least 50-fold. The original thermal images and the two types of corrected thermal images also produced similar precision results of approximately 0.8. However, detection recall increased by approximately 20% when images corrected for fur color and temperature were used. Moreover, there were two types of combined thermal and RGB images. The first type was created by multiplying a corrected RGB image with the original thermal image, and the second type was obtained by multiplying a corrected RGB image with the all-corrections-applied thermal image. The detection recall rate using

these images exceeding 0.6, and the detection precisions were 0.200 and 0.804, respectively. When corrected thermal images were used, the detection precision was approximately four-fold higher.

Use of the six types of input images resulted in different detection times. To calculate the detection time of an individual image by image type, the total detection times of the 26 images shot for each height range were summed, and then the sum was divided by 26. After repeating this process 50 times, the average detection time was calculated (Table 2). Of the four processing methods used in Google Colab, parallel processing had the fastest detection time for all image input types. The image type associated with the fastest detection time was the thermal image with unnatural color removal, which was associated with a detection time of 0.033 s. The image type associated with the slowest detection time was the corrected RGB image \times corrected thermal image combined image. This image type was associated with a detection time that was three times slower than the fastest time. In general, when the input image had RGB channels, more detection time was required. This occurred because the detection method had to consider more channels and because the large numbers of errors associated with RGB images prolonged the true–false decision-making time of the method. Converting the detection time to frames per second (FPS), the input images including RGB channels were acquired at 9 FPS. The other input images were acquired at 25–30 FPS.

Table 2. Time needed for detection by type of image under four kinds of processing environment.

CPU and RAM		Intel(R) Xeon(R) CPU @ 2.20 GHz and 12.69 GB						
Running Environment		Single CPU		GPU Accelerated (Tesla T4_16 GB)		GPU Accelerated (Tesla P100_16 GB)		CPU Parallel Processing (2 Cores)
Detection Time and Applicable FPS		Time (s)	FPS	Time (s)	FPS	Time (s)	FPS	Time (s) FPS
Corrected RGB only		0.192	5	0.159	6	0.143	7	0.109 9
Thermal only		0.047	21	0.038	26	0.036	28	0.036 28
Corrected for fur color and temperature		0.063	16	0.051	20	0.047	21	0.040 25
Unnatural color removal		0.048	21	0.038	26	0.036	28	0.033 30
RGB + Thermal		0.194	5	0.151	7	0.146	7	0.109 9
RGB + Corrected thermal		0.197	5	0.158	6	0.145	7	0.111 9

5. Discussion

5.1. Detection Precision and Recall

For wildlife detection, detection precision and recall are fundamentally important. The 26 images shot at each height range were used to generate six kinds of input images. The same detection method was used for each of these input images, and it detected between one- and two-thirds of the targets. When the input image had a thermal channel, the maximum detection precision increased approximately two-fold. Additionally, a detection recall rate of 0.699 was obtained when using the corrected RGB image and corrected thermal image together.

Previous studies have shown a diverse range of detection precision (Table 3). A method applied for hippopotamus detection performed best [30]. Studies of cattle [47], monkeys [48], and white-tailed deer [34] detected between 60% and 70% of their targets. Fur seal [49] and human [32] studies detected approximately 40% of their targets. Considering the differences in site environment, target size and shape, and thermal image shooting conditions, the detection method proposed here has above-average performance.

Table 3. Comparison of the proposed method with previous studies.

		Proposed Method	Chrétien, L.P., et al., 2016 [34]	Hambrecht, L., et al., 2019 [32]	Lhoest, S., et al., 2015 [30]	Longmore, S.N., et al., 2017 [47]	Seymour, A.C., et al., 2017 [24]	Gooday, O.J., et al., 2018 [49]	Oishi, Y., et al., 2018 [31]	Spaan, D., et al., 2019 [48]
Used Dataset		UAV Derived RGB and Thermal Images				UAV Derived Thermal Images				
Site	location	animal farm, Hongcheon, Republic of Korea	Falardeau Wildlife Observation and Agricultural Interpretive Centre, Canada	Issa study site, Tanzania	Garamba National Park, Democratic Republic of Congo	Arrowe Brook Farm Wirral, UK	Hay Island & Saddle Island, Canada	Kaikoura, New Zealand	Nara Park, Japan	Los Arboles Tulum, Mexico
	area (m ²)	120,200	2215	-	-	6500	160,000	-	5,510,000	40,000
	numbers	1	1	24	4	1	2	3	1	3
Data Acquisition	Date	25 11 2020	06 11 2011	03 2017	09 2014, 05 2015	14 07 2015	29 01 2015~02 02	19 02 2015~27	11 09 2015	10 06 2018~23
	Time	11:00~13:00	07:00~13:00	-	-	-	07:30, 19:00	07:00, 12:00, 16:00	19:22~20:22	17:30~19:00
	Altitude (m)	25~275	60	70, 100	39, 49, 73, 91	80~120	-	50	1000, 1300	70
Target	name	alpaca	white-tailed deer	human	hippopotamus	cattle	grey seal	New Zealand fur seal	sika deer	spider monkey
	body length (m)	0.8~1.0	1~1.9	0.3~0.5	3~5	2.4	1.0~2.5	1.0~2.5	1~1.9	0.7
Results (best or average)	Accuracy	0.804	0.650	0.410	0.860	0.700	0.750	0.430	0.753	0.650
	Detection time (s)	0.033	-	-	-	-	-	-	-	-

Among the previous studies, only the white-tailed deer study [34] provided detection precision and recall results. As was found here, the previous study found very large numbers of false-positive detection results when RGB images were used as input images. This previous study used unsupervised pixel-based and object-based methods. When the unsupervised pixel-based classification method was used with RGB images, the detection precision was 0.046; when the object-based method was used, it was 1.0. However, the detection recall of the two methods had the same value of 0.484. Thus, according to this result, the object-based method did not detect more targets compared with the unsupervised pixel-based method. However, the method proposed here increased both the detection recall rate and detection precision by using different kinds of input images.

5.2. Instant Detection

Detection time is also a major factor in wildlife detection. To detect animals in real-time, detection time is a more important factor than detection precision or recall. The government of the United States limits the capture rate of thermal video equipment for export to 9 FPS, and most products have this capture rate [50] including the thermal camera used here. To apply our method to 9 FPS videos, the detection time should be less than 0.12 s. The full frame rate is 30 FPS. To be able to detect animals in real-time, the detection time should be less than 0.034 s.

The methods of previous studies based on machine learning and deep learning are difficult to use in real-time, and the authors have discussed these limitations [51]. The studies listed in Table 3 did not provide detection times, and their methods require a preprocessing step so cannot be used in real-time. A study of koalas [52] provided detection-time results. When shooting from altitudes of 20, 30, and 60 m, the detection times were 1.3, 1.6, and 2.1 s, respectively, but these times are insufficiently fast for real-time use.

With parallel processing, the fastest detection time for a single input image using the method presented here was 0.033 s, and the slowest was 0.111 s. Converted to FPS, these times correspond to 30 and 9 FPS, respectively. When the input image had only a thermal channel, the FPS range was 25–30; when the input image had RGB channels, the rate was 9 FPS. Thus, all input image types can be used to analyze exported thermal videos. Single thermal channel images can detect almost every frame during real-time shooting.

Furthermore, the sensor always shoots thermal and RGB images simultaneously, so both types of input image can be used according to preference. The best way to use the method developed here is to check for the presence of wildlife in a thermal image with unnatural colors removed. For the frames with a confirmed presence of wildlife, the corrected RGB image combined with the corrected thermal image can be used to clearly determine numbers and locations.

5.3. Using the Proposed Method to Supplement Previous Methods

The proposed method can detect animals regardless of color, shape, or size and does not need to generate a training dataset. This advantage reduces the total time needed for detection, and, at the same time, the method can be used to generate the training dataset itself. As a result of the automated detection process, our method marks the outline and centroid of each target. Then, instant target sorting can be used to form sets of images of detected animals, and this stacked result can be employed for ML–DL training, even while simultaneously conducting UAV surveys in the field.

This quick and in-field detection method can be used to supplement the relatively precise and advanced existing methods. Not only is our method useful for creating a training dataset but also, when a trained model is used for detection, the region of interest in the RGB image can be minimized. This areal reduction can lead to time saving.

5.4. Utility of Thermal Sensors

The use of thermal sensors provides several benefits for wildlife detection, especially time saving in the detection process [53], enhanced detection performance [32,34], and

wider application across many species. However, thermal sensors still have limitations and drawbacks. An important limitation is that thermal sensors cannot sense through obstacles such as tree canopies, hideouts, and bushes. RGB image-based detection methods also have this limitation. However, thermal images can be used to detect the body temperature of a camouflaged target and may have an advantage over RGB images.

Another more critical drawback is the sparse resolution of the thermal camera. Compared with an RGB image, the image generated from a thermal sensor has approximately 40-fold fewer pixels and one-quarter of the coverage area. Furthermore, the spatial resolution at the same shooting height is approximately four times lower. This drawback limits the shooting height for obtaining images for use in detecting wildlife. At a height of 100 m, the pixel resolution is approximately 9 cm, and at a height of 200 m, the resolution is approximately 18 cm (Figure 12). If the shooting height is higher than 100 m, a target of the size in this paper will be represented by only a few pixels, and if multiple targets are in contact with each other or overlapped, their edges become more difficult to distinguish, and blurred targets are not detected properly. In this study, a height of 100 m seemed to be the maximum height for significant detection of wildlife. If the target size or shape was different or a high-quality thermal sensor could be used, the maximum height would be higher.

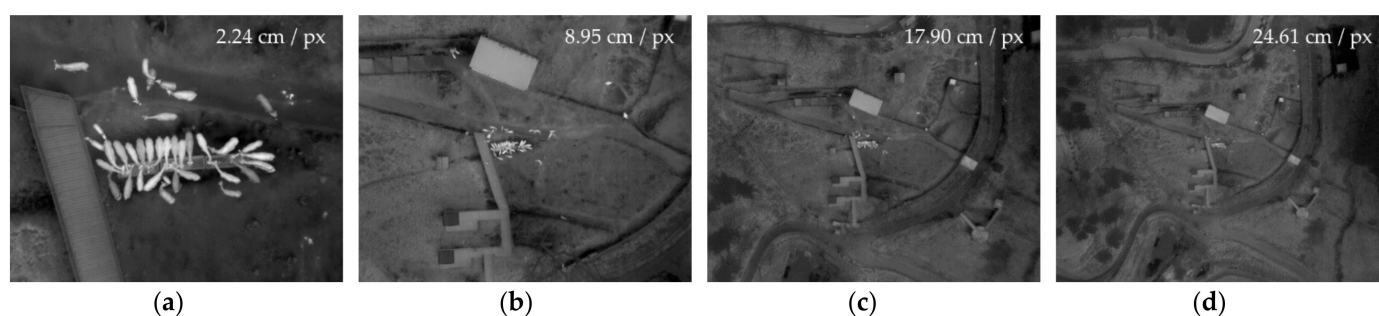


Figure 12. Thermal images taken at shooting heights of (a) 25 m, (b) 100 m, (c) 200 m, and (d) 275 m, and their pixel resolution.

5.5. Method Overview

Our method overcomes three limitations of previous studies: it can detect target animals in real-time with minimal data preprocessing, use two types of images for advanced detection ability, and be applied in diverse situations.

Size-temperature filtering enables our method to be applied to different species and land cover types. However, a lack of data means that further validation of its applicability to different land cover types is necessary. In addition, shooting altitude remains a limitation, similar to previous methods. Another drawback of the proposed method is that it cannot merge acquired images to minimize preprocessing and reduce detection time. However, although merging allows for more rapid detection of targets, partial targets might not be detected accurately. This could be overcome by using slower UAV flight speeds and higher frame rates.

6. Conclusions

This paper developed a new method for detecting animals using thermal and RGB images. The maximum detection precision was 0.804, and the recall rate was 0.699. The major improvement in detection time enables real-time usage.

This method has two limitations. The environments and conditions, such as the detection target, shooting time, and land cover, were not diverse when the raw data were acquired, so the method was developed using only limited data. Nevertheless, the method might be applicable to many species and circumstances. The sparse resolution of the thermal sensor is another limitation that limits the shooting height. Nonetheless, if high-

resolution thermal images are used, the method may be able to detect smaller targets and it may be possible to fly the UAV at a higher altitude.

The focus of future work will be to diversify the target species and shooting conditions to clarify how versatile a thermal sensor-mounted UAV system would be in conducting wildlife surveys. Using these data, an advanced method that can detect targets at greater heights will be proposed.

Author Contributions: Conceptualization, S.L.; Data curation, S.L.; Formal analysis, S.L.; Funding acquisition, Y.S. and S.-H.K.; Investigation, S.L. and S.-H.K.; Methodology, S.L.; Project administration, S.-H.K.; Resources, S.-H.K.; Software, S.L.; Supervision, Y.S. and S.-H.K.; Validation, S.L.; Visualization, S.L.; Writing—original draft, S.L.; Writing—review & editing, Y.S. and S.-H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was conducted with the support of the Korea Environment Industry & Technology Institute (KEITI) through its Urban Ecological Health Promotion Technology Development Project and funded by the Korea Ministry of Environment (MOE) (2019002760001).

Acknowledgments: This work is financially supported by Korea Ministry of Land, Infrastructure and Transport (MOLIT) as (Innovative Talent Education Program for Smart City).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Witmer, G.W. Wildlife population monitoring: Some practical considerations. *Wildl. Res.* **2005**, *32*, 259–263. [\[CrossRef\]](#)
2. Caughley, G. *Analysis of Vertebrate Populations*; Wiley: London, UK, 1977.
3. Kellenberger, B.; Volpi, M.; Tuia, D. Fast animal detection in UAV images using convolutional neural networks. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Fort Worth, TX, USA, 23–28 July 2017; pp. 866–869.
4. Pollock, K.H.; Nichols, J.D.; Simons, T.R.; Farnsworth, G.L.; Bailey, L.L.; Sauer, J.R. Large scale wildlife monitoring studies: Statistical methods for design and analysis. *Environmetrics* **2002**, *13*, 105–119. [\[CrossRef\]](#)
5. O’Connell, A.F.; Nichols, J.D.; Karanth, K.U. *Camera Traps in Animal Ecology: Methods and Analyses*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
6. Bowman, J.L.; Kochanny, C.O.; Demarais, S.; Leopold, B.D. Evaluation of a GPS collar for white-tailed deer. *Wildl. Soc. Bull.* **2000**, *28*, 141–145.
7. Bohmann, K.; Evans, A.; Gilbert, M.T.P.; Carvalho, G.R.; Creer, S.; Knapp, M.; Yu, D.W.; De Bruyn, M. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **2014**, *29*, 358–367. [\[CrossRef\]](#)
8. Hinke, J.T.; Barbosa, A.; Emmerson, L.M.; Hart, T.; Juárez, M.A.; Korczak-Abshire, M.; Milinevsky, G.; Santos, M.; Trathan, P.N.; Watters, G.M.; et al. Estimating nest-level phenology and reproductive success of colonial seabirds using time-lapse cameras. *Methods Ecol. Evol.* **2018**, *9*, 1853–1863. [\[CrossRef\]](#)
9. Burton, A.C.; Neilson, E.; Moreira, D.; Ladle, A.; Steenweg, R.; Fisher, J.T.; Bayne, T.; Boutin, S. Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *J. Appl. Ecol.* **2015**, *52*, 675–685. [\[CrossRef\]](#)
10. Ford, A.T.; Clevenger, A.P.; Bennett, A. Comparison of methods of monitoring wildlife crossing-structures on highways. *J. Wildl. Manag.* **2009**, *73*, 1213–1222. [\[CrossRef\]](#)
11. Kellenberger, B.; Marcos, D.; Tuia, D. Detecting mammals in UAV images: Best practices to address a substantially im-balanced dataset with deep learning. *Remote Sens. Environ.* **2018**, *216*, 139–153. [\[CrossRef\]](#)
12. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sens.* **2015**, *7*, 4026–4047. [\[CrossRef\]](#)
13. Bayram, H.; Stefan, N.; Engin, K.S.; Isler, V. Tracking wildlife with multiple UAVs: System design, safety and field experiments. In Proceedings of the 2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS), Los Angeles, CA, USA, 4–5 December 2017; pp. 97–103.
14. Caughley, G. Bias in aerial survey. *J. Wildl. Manag.* **1974**, *38*, 921–933. [\[CrossRef\]](#)
15. Bartmann, R.M.; Carpenter, L.H.; Garrott, R.A.; Bowden, D.C. Accuracy of helicopter counts of mule deer in pinyon-juniper woodland. *Wildl. Soc. Bull.* **1986**, *14*, 356–363.
16. Mutalib, A.H.A.; Ruppert, N.; Akmar, S.; Kamarusuzman, F.F.J.; Rosely, N.F.N. Feasibility of Thermal Imaging Using Unmanned Aerial Vehicles to Detect Bornean Orangutans. *J. Sustain. Sci. Manag.* **2019**, *14*, 182–194.
17. Thibbotuwawa, A.; Bocewicz, G.; Radzki, G.; Nielsen, P.; Banaszak, Z. UAV Mission planning resistant to weather uncertainty. *Sensors* **2020**, *20*, 515. [\[CrossRef\]](#)
18. Cesare, K.; Skeeel, R.; Yoo, S.H.; Zhang, Y.; Hollinger, G. Multi-UAV exploration with limited communication and battery. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 2230–2235.

19. Kellenberger, B.; Marcos, D.; Lobry, S.; Tuia, D. Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9524–9533. [\[CrossRef\]](#)
20. Barbedo, J.G.A.; Koenigkan, L.V.; Santos, P.M.; Ribeiro, A.R.B. Counting cattle in uav images—dealing with clustered animals and animal/background contrast changes. *Sensors* **2020**, *20*, 2126. [\[CrossRef\]](#)
21. Barbedo, J.G.A.; Koenigkan, L.V.; Santos, T.T.; Santos, P.M. A study on the detection of cattle in UAV images using deep learning. *Sensors* **2019**, *19*, 5436. [\[CrossRef\]](#)
22. Rivas, A.; Chamoso, P.; González-Briones, A.; Corchado, J.M. Detection of cattle using drones and convolutional neural networks. *Sensors* **2018**, *18*, 2048. [\[CrossRef\]](#)
23. Rey, N.; Volpi, M.; Joost, S.; Tuia, D. Detecting animals in African Savanna with UAVs and the crowds. *Remote Sens. Environ.* **2017**, *200*, 341–351. [\[CrossRef\]](#)
24. Seymour, A.C.; Dale, J.; Hammill, M.; Halpin, P.N.; Johnston, D.W. Automated detection and enumeration of marine wildlife using unmanned aircraft systems UAS and thermal imagery. *Sci. Rep.* **2017**, *7*, 45127. [\[CrossRef\]](#)
25. Lee, W.Y.; Park, M.; Hyun, C.-U. Detection of two Arctic birds in Greenland and an endangered bird in Korea using RGB and thermal cameras with an unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0222088. [\[CrossRef\]](#)
26. Bevan, E.; Wibbels, T.; Najera, B.M.; Martinez, M.A.; Martinez, L.A.; Martinez, F.I.; Cuevas, J.M.; Anderson, T.; Bonka, A.; Hernandez, M.H.; et al. Unmanned aerial vehicles (UAVs) for monitoring sea turtles in near-shore waters. *Mar. Turt. Newsl.* **2015**, *145*, 19–22.
27. Fudala, K.; Bialik, R.J. Breeding Colony Dynamics of Southern Elephant Seals at Patelnia Point, King George Island, Antarctica. *Remote Sens.* **2020**, *12*, 2964. [\[CrossRef\]](#)
28. Pfeifer, C.; Rümmler, M.C.; Mustafa, O. Assessing colonies of Antarctic shags by unmanned aerial vehicle (UAV) at South Shetland Islands, Antarctica. *Antarct. Sci.* **2021**, *33*, 133–149. [\[CrossRef\]](#)
29. Kays, R.; Sheppard, J.; Mclean, K.; Welch, C.; Paunescu, C.; Wang, V.; Kravit, G.; Crofoot, M. Hot monkey, cold reality: Surveying rainforest canopy mammals using drone-mounted thermal infrared sensors. *Int. J. Remote Sens.* **2019**, *40*, 407–419. [\[CrossRef\]](#)
30. Lhoest, S.; Linchant, J.; Quevauvillers, S.; Vermeulen, C.; Lejeune, P. How many hippos HOMHIP: Algorithm for automatic counts of animals with infra-red thermal imagery from UAV. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-3/W3*, 355–362. [\[CrossRef\]](#)
31. Oishi, Y.; Oguma, H.; Tamura, A.; Nakamura, R.; Matsunaga, T. Animal detection using thermal images and its required observation conditions. *Remote Sens.* **2018**, *10*, 1050. [\[CrossRef\]](#)
32. Hambrecht, L.; Brown, R.P.; Piel, A.K.; Wich, S.A. Detecting ‘poachers’ with drones: Factors influencing the probability of detection with TIR and RGB imaging in miombo woodlands, Tanzania. *Biol. Conserv.* **2019**, *233*, 109–117. [\[CrossRef\]](#)
33. Chabot, D. Systematic Evaluation of a Stock Unmanned Aerial Vehicle (UAV) System for Small-Scale Wildlife Survey Applications. Doctoral Dissertation, McGill University, Montreal, QC, Canada, 2009.
34. Chrétien, L.-P.; Théau, J.; Ménard, P. Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system. *Wildl. Soc. Bull.* **2016**, *40*, 181–191. [\[CrossRef\]](#)
35. López, A.; Jurado, J.M.; Ogayar, C.J.; Feito, F.R. A framework for registering UAV-based imagery for crop-tracking in Precision Agriculture. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102274. [\[CrossRef\]](#)
36. Heikkilä, J.; Silven, O. Calibration procedure for short focal length off-the-shelf CCD cameras. In Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 25–29 August 1996; Volume 1, pp. 166–170.
37. Oishi, Y.; Matsunaga, T. Support system for surveying moving wild animals in the snow using aerial remote-sensing images. *Int. J. Remote Sens.* **2014**, *35*, 1374–1394. [\[CrossRef\]](#)
38. Kellie, K.A.; Colson, K.E.; Reynolds, J.H. *Challenges to Monitoring Moose in Alaska*; Alaska Department of Fish and Game, Division of Wildlife Conservation: Juneau, AK, USA, 2019.
39. Třebický, V.; Fialová, J.; Kleisner, K.; Havlíček, J. Focal length affects depicted shape and perception of facial images. *PLoS ONE* **2016**, *11*, e0149313. [\[CrossRef\]](#)
40. Neale, W.T.; Hessel, D.; Terpstra, T. *Photogrammetric Measurement Error Associated with Lens Distortion*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2011.
41. Hongzhi, W.; Meijing, L.; Liwei, Z. The distortion correction of large view wide-angle lens for image mosaic based on OpenCV. In Proceedings of the 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), Jilin, China, 19–22 August 2011; pp. 1074–1077.
42. Synnefa, A.; Santamouris, M.; Apostolakis, K. On the development, optical properties and thermal performance of cool colored coatings for the urban environment. *Sol. Energy* **2007**, *81*, 488–497. [\[CrossRef\]](#)
43. Griffiths, S.R.; Rowland, J.A.; Briscoe, N.J.; Lentini, P.E.; Handasyde, K.A.; Lumsden, L.F.; Robert, K.A. Surface re-reflectance drives nest box temperature profiles and thermal suitability for target wildlife. *PLoS ONE* **2017**, *12*, e0176951. [\[CrossRef\]](#)
44. Apolo-Apolo, O.E.; Pérez-Ruiz, M.; Martínez-Guanter, J.; Valente, J. A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a deep learning technique. *Front. Plant Sci.* **2020**, *11*, 1086. [\[CrossRef\]](#)
45. Sobel, I. *An Isotropic 3 × 3 Gradient Operator, Machine Vision for Three-Dimensional Scenes*; Academic Press: New York, NY, USA, 1990.
46. Russ, J.C.; Matey, J.R.; Mallinckrodt, A.J.; McKay, S. The image processing handbook. *Comput. Phys.* **1994**, *8*, 177–178. [\[CrossRef\]](#)

-
47. Longmore, S.N.; Collins, R.P.; Pfeifer, S.; Fox, S.E.; Mulero-Pázmány, M.; Bezombes, F.; Goodwin, A.; De Juan Ovelar, M.; Knapen, J.H.; Wich, S.A. Adapting astronomical source detection software to help detect animals in thermal images obtained by unmanned aerial systems. *Int. J. Remote Sens.* **2017**, *38*, 2623–2638. [[CrossRef](#)]
 48. Spaan, D.; Burke, C.; McAree, O.; Aureli, F.; Rangel-Rivera, C.E.; Hutschenreiter, A.; Longmore, S.N.; McWhirter, P.R.; Wich, S.A. Thermal infrared imaging from drones offers a major advance for spider monkey surveys. *Drones* **2019**, *3*, 34. [[CrossRef](#)]
 49. Gooday, O.J.; Key, N.; Goldstien, S.; Zawar-Reza, P. An assessment of thermal-image acquisition with an unmanned aerial vehicle (UAV) for direct counts of coastal marine mammals ashore. *J. Unmanned Veh. Syst.* **2018**, *6*, 100–108. [[CrossRef](#)]
 50. Luo, R.; Sener, O.; Savarese, S. Scene semantic reconstruction from egocentric rgb-d-thermal videos. In Proceedings of the 2017 International Conference on 3D Vision 3DV, Qingdao, China, 10–12 October 2017; pp. 593–602.
 51. Van, G.; Camiel, R.V.; Pascal, M.; Kitso, E.; Lian, P.K.; Serge, W. Nature Conservation Drones for Automatic Localization and Counting of Animals. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 255–270.
 52. Gonzalez, L.F.; Montes, G.A.; Puig, E.; Johnson, S.; Mengersen, K.; Gaston, K.J. Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors* **2016**, *16*, 97. [[CrossRef](#)]
 53. Witczuk, J.; Pagacz, S.; Zmarz, A.; Cypel, M. Exploring the feasibility of unmanned aerial vehicles and thermal imaging for ungulate surveys in forests-preliminary results. *Int. J. Remote Sens.* **2018**, *39*, 5504–5521. [[CrossRef](#)]