Published: 06 January 2026

Citation in BibTeX format

ICDCN 2026: 27th International Conference on Distributed Computing and Networking
January 6 - 9, 2026
Nara, Japan

EXTENDED-ABSTRACT

Wildlife Detection using Motion History Information Captured by Camera Trap in the Dark

SHOGO YOSHIDA, Nara Institute of Science and Technology, Ikoma, Nara, Japan

YUKI MATSUDA, Okayama University, Okayama, Okayama, Japan

KOTA TSUBOUCHI

HIROHIKO SUWA, Nara Institute of Science and Technology, Ikoma, Nara, Japan

KEIICHI YASUMOTO, Nara Institute of Science and Technology, Ikoma, Nara, Japan

ICDCN Companion '26: Companion Proceedings of the 27th International Conference on Distributed Computing and Networking (January 2026)
https://doi.org/10.1145/3737611.3776961
ISBN: 9798400719691

Wildlife Detection using Motion History Information Captured by Camera Trap in the Dark

Shogo Yoshida
Nara Institute of Science and Technology
Ikoma, Nara, Japan
yoshida.shogo.yo4@is.naist.jp

Yuki Matsuda
Okayama University
Okayama, Okayama, Japan
Nara Institute of Science and Technology
Ikoma, Nara, Japan
yukimat@okayama-u.ac.jp

Kota Tsubouchi
LY Corporation
Tokyo, Japan
ktsubouc@lycorp.co.jp

Hirohiko Suwa
Nara Institute of Science and Technology
Ikoma, Nara, Japan
h-suwa@is.naist.jp

Keiichi Yasumoto
Nara Institute of Science and Technology
IKoma, Nara, Japan
yasumoto@is.naist.jp

Abstract

This paper proposes a method to analyze infrared camera trap images captured at night for wildlife monitoring. Infrared images are typically noisy, making daytime image analysis ineffective. Our approach extracts motion information from continuous frames to enhance classification accuracy. Experiments using real nocturnal data demonstrate that the proposed method outperforms existing models that analyze single images, achieving higher efficiency and accuracy in wildlife detection under low-light and noisy conditions.

CCS Concepts

• **Computing methodologies** → *Computer vision.*

Keywords

Image Classification, CNN, Wildlife Monitoring, Infrared Cameras

ACM Reference Format:
Shogo Yoshida, Yuki Matsuda, Kota Tsubouchi, Hirohiko Suwa, and Keiichi Yasumoto. 2026. Wildlife Detection using Motion History Information Captured by Camera Trap in the Dark. In *Companion volume of the 27th International Conference on Distributed Computing and Networking (ICDCN Companion 2026), January 06–09, 2026, Nara, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3737611.3776961

1 Introduction

Crop damage caused by wild animals is a global issue, emphasizing the need for effective wildlife monitoring. Camera traps are widely used to observe nocturnal wildlife [4]. However, manually inspecting the large number of infrared images collected is highly labor-intensive. Many studies [8] have employed machine learning to classify animals automatically, but most focus on daytime RGB images. In contrast, infrared images are affected by noise, low contrast, and partial visibility of animals, often making recognition challenging even for human observers [9].

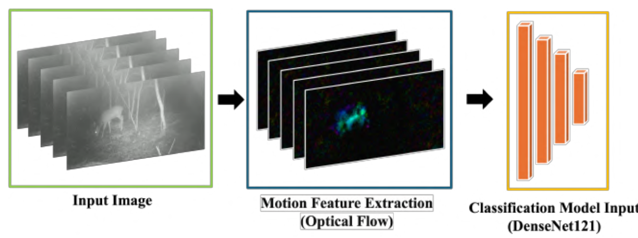


Figure 1: Overview of proposed method

To address this limitation, we propose a method that utilizes Optical Flow [5] to extract motion information from sequential nighttime frames. Unlike conventional CNN-based static image analysis, our approach leverages temporal dynamics to improve detection robustness under noisy conditions.

2 Methodology

This study provides a two-class classification of wildlife and background from infrared images captured at night by camera traps. Data captured at night by infrared cameras are rendered in grayscale with reduced luminance and noise and other differences in image appearance compared with RGB camera data captured during the daytime. Therefore, wildlife data captured by infrared cameras do not clearly show the entire body, rather only the eyes or a part of the body, making it difficult for the human eye to judge whether wildlife has been captured or not. Our method employs dynamic motion information to classify targets rather than relying on static appearance. As illustrated in Figure 1, motion information is extracted using Optical Flow from video sequences captured by infrared cameras at night. This information is then used as input features for the classification model, enabling highly accurate detection. Optical Flow, specifically the Gunnar Farneback method [3], represents pixel motion between consecutive image frames as vectors. Although computationally expensive due to per-pixel motion estimation, it captures detailed movement vectors and rich motion

ACM ISBN 979-8-4007-1969-1/26/01
https://doi.org/10.1145/3737611.3776961

Table 1: Overall Performance

Data	RGB	Proposed method
Camera Trap	74.4 %	**79.5 %**
Prof. Camera	**98.0 %**	96.7 %

information. Note that algorithms for extracting motion information are not limited to Optical Flow; the Motion History Image (MHI) [1] technique is another well-known example. The extracted Optical Flow data are used as input features to a classification model that identifies wildlife in nighttime infrared images. We employ DenseNet121 [6] as the classification architecture.

3 Experiment

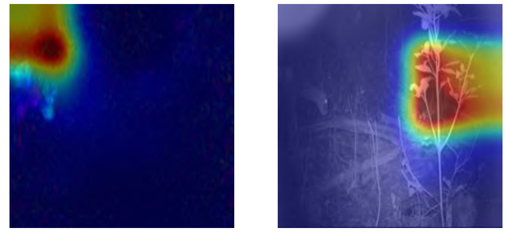
The performance of the proposed method was compared with a model that was trained on infrared images in the same ways as if they were regular RGB images (described as RGB below). The data used in the experiments were images actually captured by camera traps set up in mountainous areas, and only infrared images, which are prone to contain noise and have low resolution, were used. In total, 14316 images were used. 6941 images were used for training, 2975 for validation, and 4400 for evaluation. We chose DenseNet121 as a learning model. The batch size during training was set to 16, and adaptive moment estimation (Adam) [7] was used as the optimization method. The learning rate was 1×10^{-6}, cross entropy loss was used as the loss function, and transfer learning was performed based on models trained using ImageNet [2]. The results are summarized in Table 1. The accuracy of the baseline RGB model was 74.41%, while the proposed method achieved 79.53%. This improvement confirms that incorporating motion information enhances performance compared with static RGB-based analysis.

4 Discussion

Evaluation using high-resolution images: To enhance the effectiveness of the proposed method, the same experiment was conducted on another high-resolution camera dataset. The high-resolution infrared images were taken by a professional-use low-noise camera, and the full-body images of wildlife were clearly visible and easily recognized by the human eye. The total number of high-resolution images was 11,476. 5,390 images were used for training, 2,310 for validation, and 3,776 for evaluation.

On the high-resolution camera dataset, the accuracy of the RGB model was 98.0% whereas that of the proposed method was 96.7% (See the "Prof. Camera" result in Table 1). The proposed method was comparably lower in accuracy possibly because the existing method works sufficiently well on high-resolution data.

Limitations: A model using motion information was successful in classification when the motion of the target was large, but tended to be less accurate when the motion of the target was slight, such as when swaying of branches and leaves was small or when the wild animal was stationary. This is believed to be due that the classification focuses only on the motion information of the target, which reduces the amount of available information when the target is not moving. It is possible to compensate for the decline in classification accuracy by combining appearance information as a supplement, rather than relying solely on motion information.



(a) Proposed Method **(b) RGB**

Figure 2: Grad-CAM Result

Qualitative evaluation using Grad-CAM: A method based on Grad-CAM [10] was used to compare the outputs of the RGB model and the proposed method (Figure 2). Grad-CAM visualizes the regions of interest in an input image as a heatmap of the regions that contributed the most to the model's prediction. The comparison indicated that the model using RGB images focused on the background vegetation instead of the target animal, incorrectly classifying it into the background class, whereas the proposed method accurately focused on the animal and correctly classified it. This is because, by using motion information from OpticalFlow as input for training, background information is eliminated and specific parts and behavioral features of the animal are effectively captured. These results confirm that the proposed method can effectively learn and classify by using motion data of the target.

5 Conclusion

We developed a method for determining the presence of wildlife in noisy images taken by an infrared camera at night. The method is based on the idea that, by using optical flow to capture animal movement information, wildlife can be accurately detected even in noisy data. We compared our method with a model that judges static RGB image-based images and validated it in experiments on images from actual camera traps installed in a mountainous area.

References

[1] Aaron F. Bobick and James W. Davis. 2001. The recognition of human movement using temporal templates. *IEEE Transactions on pattern analysis and machine intelligence* 23, 3 (2001), 257–267.
[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
[3] Gunnar Farnebäck. 2003. Two-frame motion estimation based on polynomial expansion. In *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*. Springer, 363–370.
[4] Neil A Gilbert, John DJ Clare, Jennifer L Stenglein, and Benjamin Zuckerberg. 2021. Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology* 35, 1 (2021), 88–100.
[5] Berthold KP Horn and Brian G Schunck. 1981. Determining optical flow. *Artificial intelligence* 17, 1-3 (1981), 185–203.
[6] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
[7] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[8] Scott Leorna and Todd Brinkman. 2022. Human vs. machine: Detecting wildlife in camera trap images. *Ecological Informatics* 72 (2022), 101876.
[9] Frank Schindler and Volker Steinhage. 2021. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. *Ecological Informatics* 61 (2021), 101215.
[10] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.