Object Detection Method for Drone Videos Using Optical Flow

Takahiro Shindo Kein Yamada Taiju Watanabe Hiroshi Watanabe
Graduate School of Fundamental Science and Engineering, Waseda University

Abstract

The application of object detection in aerial video taken by drones is crucial. Since objects appear small in drone-captured video, object detection models with many parameters are required for accurate object detection. Therefore, video compression and information reduction methods are considered for transmitting videos to the object detection model on a server. Our paper suggests an approach of utilizing optical flow to estimate object positions to reduce the number of frames transmitted to the server. Experimental results reveal that frame reduction by the proposed method does not result in degradation of image recognition accuracy.

Keywords: video transmission, object detection, edge camera, optical flow, yolov5

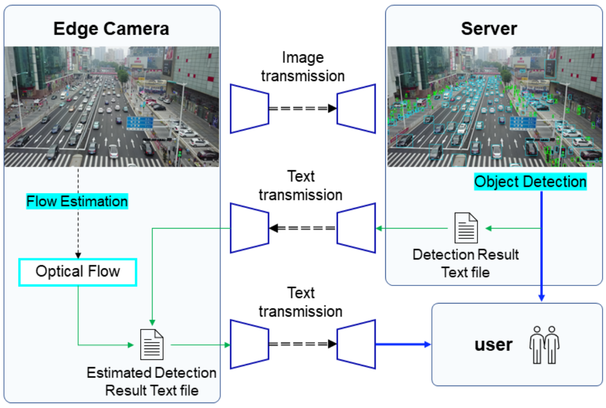

Figure 1: Proposed Video Transmission system for Object Detection.

1. Introduction

In the era of Beyond 5G, more efficient video transmission methods are needed in terms of image quality and image recognition accuracy. In particular, research on methods for transmitting video captured by edge cameras is active. Although HEVC[1] and VVC[2] have been standardized as video coding methods for human vision, video coding standards for image recognition have not yet been created. Therefore, MPEG has positioned video coding for image recognition as Video Coding for Machines (VCM) and started standardization activities. In this standardization activity, video compression methods for image recognition such as object detection and segmentation are studied.

Generally, the computational power of edge cameras is not so high. Due to the weak computational power, these cameras are insufficient for neural network-based object detection. Therefore, it is necessary to transmit videos from edge cameras and perform image recognition processing on a server. Image coding and information reduction techniques for image recognition are essential in this transmission. In this paper, we propose a method to reduce the number of transmitted frames by using optical flow to infer object positions. By reducing the number of transmitted frames, the amount of information in the transmitted video can be reduced. In experiments, we investigate the effect of our method on the accuracy of object detection. We show that the proposed method can reduce the number of transmitted images with almost no loss of image recognition accuracy.

2. Related Works

Optical flow is the calculated motion of pixels between adjacent frames as a vector. OpenCV has a function that calculates the optical flow for all pixels in an image. This function is based on the paper by Gunnar[3], implemented as function cv2.calcOpticalFlowFraneback(). Unlike optical flow estimation using the Lucas-Kande method[4], it can detect dense flows in images. In addition, many flow estimation methods using neural networks have been proposed in recent years, but the problem with all of them is that they are computationally expensive.

3. Proposed Method

In this paper, we propose a method to reduce the number of transmission frames by using optical flow to infer object location. An overview of the transmission method is shown in Fig.1. First, the first frame is transmitted from the edge device to the server, where object detection is performed on the image using an image recognition model. The results are then returned from the server to the edge device. Then, at the edge device, the optical flow is calculated from the video information, and the position of the object is inferred based on the detection results received from the server. In other words, the position coordinates of the object are shifted by the amount of optical flow. This process eliminates the need to transmit frame images to the server and perform the object detection process since the exact position of the object can be estimated.

Figure 2: Left image: An example of an image from the VisDrone dataset. Right image: optical flow calculated by flow estimation.

4. Experiment

4.1. Evaluation Method

To confirm the effectiveness of the proposed method, we investigate the accuracy of object location estimation using optical flow. In the validation experiment, seven sequences of the VisDrone dataset[5] for validation are used. VisDrone dataset consists of aerial videos taken by drones, and an example of a frame image is shown in Fig.2. These videos are published as multiple images, with a total frame count of 2846. All seven videos were captured by an edge camera mounted on the drone and annotated for object detection.

For each video, the video will be read consecutively by 12 frames. We first input only one of the 12 frames of the video into the object detection model. Then, we apply our proposed method for the remaining frames in which optical flow infers the object location. The following process is repeated until the the end of the video. An example of optical flow is also shown in Fig.2. The object detection model is YOLOv5[6], trained using training images from the VisDrone dataset. As a comparison method, we also measure the detection accuracy when all frames of the video are input to YOLOv5.

4.2. Results

The experimental results are shown in Fig.3 and Table 1. The figure shows the object detection accuracy of the proposed and compared methods for each of the seven sequences, and the table shows the average of their accuracy. Detection accuracy is expressed in mAP, and 0.5 and 0.9 are used for the IoU thresholds when calculating mAP. In this figure and table, prop_mAP50 and prop_mAP90 represent the object detection accuracy using the proposed method when the IoU threshold is set to 0.5 and 0.9, respectively. conv_mAP50 and conv_mAP90 represent that of the existing methods to be compared.

These results show that even if the object position is estimated using optical flow instead of YOLOv5, the degradation in detection accuracy is minimal. In addition, when we use the proposed method, a larger IoU threshold value results in a decrease in object detection accuracy. However, the fall in accuracyis only a 5.0% decrease in object detection accuracy compared to the case where all frames are input to YOLOv5.

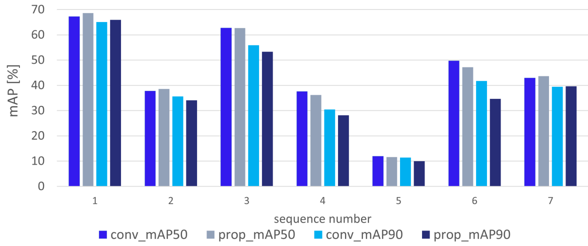

Figure 3: Comparison of object detection accuracy for each of the 7 sequences when using existing and proposed methods.

Table 1: Comparison of average object detection accuracy for all sequences when using existing and proposed methods.

conv_mAP50	prop_mAP50	conv_mAP90	prop_mAP90
44.3%	44.1%	40.0%	38.0%

5. Conclusion

We proposed a method to reduce the number of transmitted image frames by using optical flow to infer object location. We showed that the proposed method hardly reduces the accuracy of object detection in video, even if the number of frame images input to the object detection model is reduced. For future research, it is necessary to actually encode the images and calculate the bit rate required for transmission, and then investigate the effect of the proposed method on the reduction of transmission volume.

Acknowledgment

The results of this research were obtained from the commissioned research (JPJ012368C05101) by National Institute of Information and Communications Technology (NICT), Japan.

References

[1] High Efficiency Video Coding, Standard ISO/IEC 23008-2, ISO/IEC JTC 1, Apr. 2013.

[2] Versatile Video Coding, Standard ISO/IEC 23090-3, ISO/IEC JTC 1, Jul. 2020.

[3] F. Gunnar. "Two-frame motion estimation based on polynomial expansion." Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29-July 2, 2003 Proceedings 13. Springer Berlin Heidelberg, 2003.

[4] Baker, Simon, and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework." International journal of computer vision 56, 2004, pp. 221-255.

[5] D.Du *et al.*, "VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 0-0.

[6] G. Jocher *et al.*, "ultralytics/yolov5: v7.0-yolov5 sota realtime instance segmentation," Zenodo, Nov., 2022.