Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 12, Number 1, pages 31–32, January 2023

Implementation of Pedestrian Tracking in Low-Resolution Video using Multi-Camera

Gukjin Son
ICT Research Institute
DGIST
Daegu, Korea
sudopop@dgist.ac.kr

Junkwang Kim
ICT Research Institute
DGIST
Daegu, Korea
kjk1208@dgist.ac.kr

Youngduk Kim*
ICT Research Institute
DGIST
Daegu, Korea
ydkim@dgist.ac.kr

***Abstract*— The utility of intelligent CCTV has been verified in many social domains. One of the main purposes of intelligent CCTV is pedestrian tracking. In this paper, we concentrate on tracking pedestrians using multiple cameras. Especially, we present a method that not only recognizes pedestrians from high-resolution camera views but also continuously tracks the recognized pedestrians from low-resolution camera views. Deep learning-based object detection model(YOLO-v4) and multi-object tracking algorithm(DeepSORT) are used in the suggested method to track pedestrians. Pedestrians are matched from the views of high-resolution and low-resolution cameras using a perspective transformation.**

Keywords—Object Detection, Multiple object tracking, Re-identification, Deep learning, Computer vision

I. Introduction

Nowadays, intelligent CCTV (Closed Circuit Television) is widely used in various social fields such as pedestrian recognition, crime scene prediction, and abnormal behavior detection. Especially, thanks to the multiple information, multi-camera system serves to improve the precision and coverage of intelligent CCTV systems. However, processing the vast amount of video captured by multiple cameras requires more advanced resources than a single-camera approach.

Pedestrian tracking is a technology that detects a specific person or multiple people at the same time and seamlessly confirms their movement. Recently, many studies on pedestrian tracking have been conducted and various algorithms have been proposed. For pedestrian tracking, it is common to simply use a single-camera system to track, but if the tracking range is wide and diverse, it is better to use multiple cameras.

One of the main issues for pedestrian tracking from multiple cameras is an extension of the problem of pedestrian tracking from a single-camera. In this paper, we focus on solving this scalability problem. In an environment where heterogeneous cameras coexist, we propose a method that not only detects pedestrians from high-resolution camera views but also continuously tracks the detected pedestrians from low-resolution camera views. The proposed method improves pedestrian tracking performance and ensures the scalability of cameras with various infrastructures. Our method tracks pedestrians using YOLO (You Only Look Once)-v4 [1] and DeepSORT algorithm [2]. In addition, Re-ID (re-identification) technology is used to match pedestrians from the viewpoint of high-resolution and low-resolution cameras.

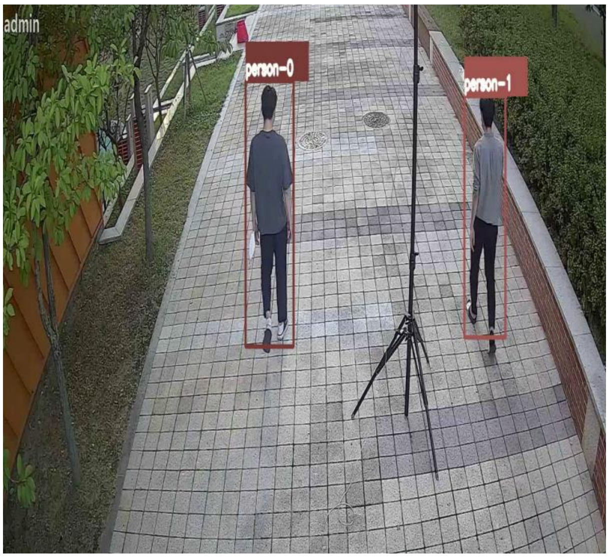

Figure 1. Pedestrian tracking results from the low-resolution image using single-camera.

II. METHODS

The object detection algorithm is divided into a single-stage detector that performs classification and detection at the same time and a two-stage detector that performs classification and detection separately. One stage detector has a fast execution speed like YOLO [3] and RetinaNet [4], but has slightly lower accuracy. The two-stage detector has a relatively slow processing speed like FPN (Feature Pyramid Networks) [5] and Cascade R-CNN [6] but has high accuracy. We used YOLO-v4 as an object detector to extract pedestrian features and draw bounding boxes. YOLO-v4 model is divided into three parts: CSPDarknet53, neck, and head. CSPDarknet53 which greatly reduces the computational amount of CNN structure is a backbone to ensure real-time performance. The neck structure is composed of spatial pyramid pooling and personal area networking, and the head structure is composed based on the anchor of YOLO-v3 [7].

Object tracking is a process of estimating bounding boxes and the identities of objects in videos. In the proposed method, DeepSort is adopted as a tracker to track moving objects throughout the image sequence. DeepSORT predicts the object position of the next frame using the kalman filter based on the information of the previous frame and obtains the mahalanobis distance and cosine distance through visual similarity. This task is associated with Re-ID.

Pedestrian Re-ID is associating images of the same pedestrian taken from different cameras. The core of Re-ID is to compare the similarity between the features of pedestrians taken from different cameras. However, we study a method of tracking pedestrians by considering the geometrical relationship of the background by placing overlapping areas between multi-cameras. The geometric relationship of the background is transformed through perspective transformation. This type of transformation keeps straight lines straight after transformation but does not preserve parallelism, length, and angles.

III. Results and Discussion

The proposed methods were compared through two experiments: (1) pedestrian tracking results of different-resolution images using a single-camera and (2) pedestrian tracking results of low-resolution images using multiple cameras were compared. We use a pre-trained network using the COCO (Microsoft Common Objects in Context) dataset [8] to train pedestrians. Figure 1 shows the result of detecting and tracking a pedestrian with a high-resolution single-camera. After detecting a pedestrian, a unique ID (Identification) is assigned, and ID switching in which the initially assigned ID is changed to another ID does not occur. With a high-resolution single-camera, it is not difficult to detect and continuously track pedestrians.

Figure 2 shows the result of detecting and tracking a pedestrian with a low-resolution single-camera. Although a unique ID is assigned after detecting a pedestrian, an ID switching problem occurs in which the initially assigned ID is continuously changed to another ID.

Figure 2. Pedestrian tracking results from the low-resolution image using single-camera.

Figure 3 shows the result of matching the pedestrian detected by the high-resolution camera with the pedestrian of the low-resolution camera using perspective transformation. After detecting a pedestrian, a unique ID is assigned, and ID switching does not occur continuously. However, due to the image blurring of the low-resolution camera, false positives occur in which the system incorrectly detects the background as a pedestrian.

Figure 3. Pedestrian tracking results from the low-resolution image using multi-cameras.

IV. Conclusion

We propose a method to improve pedestrian tracking performance in an environment where heterogeneous cameras coexist by using the geometric relationship among multiple cameras. Although the proposed method has the disadvantage by the fact that the target areas of the high-resolution camera and the low-resolution camera should overlap, it has the advantage of being able to detect pedestrian features without individually collecting them. In the future, we plan to study a pedestrian tracking method using both the pedestrian information collected from cameras and heterogeneous sensors.

Acknowledgment

This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (22-IT-02, 50%) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00895, 50%)

References

[1] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).

[2] Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric." 2017 IEEE international conference on image processing (ICIP). IEEE, 2017.

[3] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[4] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[5] Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[6] Cai, Zhaowei, and Nuno Vasconcelos. "Cascade r-cnn: Delving into high quality object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[7] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[8] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.