Research and Improvement of Multi-granularity Method for Pedestrian Tracing System

Ruixuan Ni

1Beijing Engineering Research Center for IOT Software and Systems, Beijing University of Technology, Beijing, 100124, China
nrx@email.bjut.edu.com

Abstract. In the field of video surveillance, with the continuous development of the Internet and artificial intelligence, the related technology for single device has been very mature, but how to accurately realize pedestrian tracing under cross-devices is still an active research field. In this paper, a multi-granularity pedestrian traceability method is proposed, which is based on the attribute extraction of pedestrian image and the person re-identification algorithm based on attention mechanism retrieves the same pedestrian image cross a lot of devices. The results show that the video surveillance system has good performance, which reduces the occurrence of tracking loss, and the track is visualized at the platform.

1. Introduction

In the field of computer vision, the traditional video surveillance system only provides the display and playback function of video, which requires a lot of manpower to ensure public safety. However, intelligent video surveillance system can automatically extract the key information of surveillance by identifying, tracking and semantic analysis. How to improve the efficiency of surveillance is the key problem to be solved in modern intelligent video surveillance system. One of the important tasks in surveillance video is to discover people and explain their behavior. We need to know if the target pedestrian appears in the multi-channel camera monitoring system where to form an efficient visual network [1]. The identification of pedestrian in video sequences is a pedestrian detection problem. Considering the temporal and spatial correlation, it is a pedestrian tracking problem to identify and track in surveillance video. The problem of retrieving a pedestrian target in different surveillance is called person re-identification problem. Pedestrian detection technology is the basis of pedestrian tracking and pedestrian recognition. The effective pedestrian tracing method proposed in this paper is based on the related theory of re-identification for further research and analysis.

In recent years, deep learning methods have achieved great success in the field with their excellent feature modelling ability. The video frame in the surveillance video has the problems of low resolution, obvious illumination change and insufficient occlusion. How to reduce the influence of these shortcomings and factors is the key to solve the problem. At present, the research on algorithms for complex scenarios is still in the primary stage, and there is still room for further research. Although the existing methods have proposed many solutions for feature extraction, the researches on strategy are not deep enough. However, in practical application, this is the key problem to limit robustness. One of the main detection algorithms is two-stage methods, such as SPP-Net and Faster-RCNN[2], by extracting candidate boxes and then classifying and regression, the accuracy of this kind of detection is higher. Moreover, there are one-stage methods, such as YOLO[3] and SSD, which are classified and regressed directly by dense sampling. Although the accuracy of this kind of detection is not as good as
that of the previous class, the speed of detection is faster. The tracking algorithm is mainly based on two ideas: on the one hand, it does not rely on any prior knowledge, directly detects moving pedestrians in static background or dynamic background from the sequence of images, on the other hand, it depends on the prior knowledge of moving pedestrians. One kind of person re-identification algorithm based on deep learning is representation learning, learning the feature of the same pedestrian even if the field of vision is different, the other is metric learning, which makes the feature of the same pedestrian similar. The feature gap of different pedestrians is larger. Based on the environmental problems of video by multi-channel cameras and the vast field of vision of surveillance video, tracking loss occurs frequently. This article visualizes the trajectory on the platform based on the existing algorithm.

2. Multi-granularity method for pedestrian tracing system

2.1. Architecture

The whole architecture of this paper is divided into four parts: (1) perception layer: the sensing environment data of this system is mainly based on high-definition video camera and processing terminal, or infrared cameras can be added later which can detect of night data. (2) network layer: through intelligent gateway and 4G/5G or NB-IOT wireless network, this system uploads 256*128 pixel data to server after pedestrian detection at processing terminals. (3) storage layer: images are stored on server as PNG file in hierarchical storage catalogs. Important data such as image address, collecting time, collect deviceID and so on in database indexed of device location. Basic database storage the user data and device state data for platform. Storage layer associates feature data with picture address by pedestrianID. (4) platform layer: including the function of setting device access, attribute analysis and pedestrian trace display. It is a platform which have good expansion.

![Figure 1. System architecture of video surveillance](image)

2.2. Data association

In this paper, a method for pedestrian tracing is proposed, which integrates algorithm for pedestrian detection on the processing terminals. They report the data to the platform at a certain number of frames at each interval, uploads the pedestrian picture to the image database which indexed the device location index. By using the person re-identification algorithm based on spatial and channel attention mechanism, users choose the image to extract the feature, and the feature database is constructed by extracting the feature value in the other cameras. To calculate the cosine distance matrix between features, user defines input value-n, extract the range of Rank-n pictureID in feature matching. And then, the platform uses pictureID to find their pedestrian attribute through API to accurate the range of
pedestrians. Eliminate pedestrian images with different attributes in Rank-n pictures before display. The remaining results are considered as the same pedestrian, which can identify suspicious persons efficiently, realize safety warning and improve site safety and ensure public safety.

![System data flow chart](image)

Figure 2. System data flow chart

3. Implementation of pedestrian tracing system

3.1. Attention

Attention mechanism was first used in natural language processing, and later used in computer vision. In neural network, attention mechanism is an additional neural network branch, which assigns different weights to different parts of the input. It is really important to form the different weights called excitation, and then put the weight on the input called scale. Zhang[4]'s idea is to exploit the pairwise relation related to the $i^{th}$ feature to represent this feature node’s global structural information. Specifically, we use $r_{ij}$ to represent the affinity between the $i^{th}$ feature and the $j^{th}$ feature. Then, we use the feature itself and the pairwise relations, $y_i = [x_i, r_i]$, as the feature used to infer its attention using a shared transformation function. More specifically, we use 1*1 Conv2d, BatchNorm2d and ReLU activation to get $(r_{i1}, r_{i2}, ...r_{iN})$, permute dimension to get $(r_{1i}, r_{2i}, ...r_{Ni})$, and cat these two with $x_i$, and squeeze to form weight.

3.2. Person re-identification based on attention mechanism

The model training process adopts the deep residual network which is resnet50 [5] and loads the pre-training weight. By directly transmitting the input $x$ to the output as the initial result, the residual error block is formed, which will insure the accuracy while deepening the neural network. A perceptual of space and channel attention is added between each residual block. In particular, the original feature is spliced with the space, channel feature to form a weight value, and the weight value is multiplied to the original feature. Then enter the full connection layer for the classification task, including Cross entropy loss and triple hard loss, using by adam or sgd optimizer training 600 epoches on CUHK03 dataset [6] to get a trained weight file. The test accuracy on the CUHK03 dataset is Rank1 78.9%, Rank5 91.1%, Rank10 93.9%.
4. Experiment
Server hardware is using Nvidia GTX 2080Ti GPU, Intel CPU Xeon E5-2682v4, 40G SSD system disk, peak bandwidth as 5 M. This paper uses the back-end technology which constructed the platform as Spring, SpringMVC, Mybatis, Netty framework, the platform uses Java Language and the MVC mode to carry on the stratification, Mybatis can separate SQL from code which improves the maintainability, Netty is easier to deal with large data streams and support I/O timeout and idle state detection, Spring can support IOC and AOP, and SpringMVC can handle many requests from the front-end. The front-end technologies are HTML, CSS, JavaScript, and uses Baidu API which can recognize person attributes, whether to wear a hat, whether to wear a mask, whether to wear glasses, whether there is a backpack. The front-end interaction adopts a lightweight data interaction format-json format for requesting and responding data, the device access section uses a TCP protocol and a custom application layer protocol. the communication message includes header tags, data segment length, data segment, CRC check and packet tail, and defines its type and length. Person re-identification uses the trained model to feature extraction and gets the interface to save in database, and then when you choose one picture and search the database for matching feature and get the labels for the rank result, and among picture labels we use attributes to take out the corresponding results from the Mysql database and display them in the interface.

5. Conclusion
In this paper, a multi-granularity pedestrian tracing method is proposed to improve the accuracy of pedestrian tracking by using a deep learning model based on attention mechanism with good performance and pedestrian attribute analysis which is used to further reduce the result set. In addition, in the subsequent research, the false detection in person re-identification algorithm will be further reduced, and we have encapsulated the interface for the upper platform which has certain expansibility.

References
[1] Huang K, Chen X, Kang Y, Tan T. (2015) A Review of Intelligent Video Surveillance Technology. Journal of Computer Science, 38(06):1093-1118.
[2] Ren S, He K, Girshick R, Sun J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 39(6):91-99.
[3] Redmon J, Divvala S, Girshick R, et al.(2016) You only look once: Unified, real-time object detection. In: Computer vision and pattern recognition. Las Vegas. pp. 779-788.
[4] Zhang Z, Lan C, Zeng W, et al. (2020) Relation-Aware Global Attention for Person Re-Identification. In: Computer vision and pattern recognition. Seattle. pp. 3183-3192.
[5] He K, Zhang X, Ren S, et al. (2016). Deep Residual Learning for Image Recognition. In: Computer vision and pattern recognition. Las Vegas. pp. 770-778.

[6] W. Li, R. Zhao, T. Xiao, and X. Wang. (2014) Deepreid: Deep filter pairing neural network for person re-identification. In: Computer vision and pattern recognition. Columbus. pp. 152-159