令和4年度 知能システム学専攻修士論文要旨

工藤 研究室	氏　　名	CHEN Yichen
修士論文題目	Person Retrieval System Using Attributes of Pedestrians in Surveillance Video Images	

In the field of video surveillance, pedestrian retrieval systems are a focus of technical study. For example, in the Boston Marathon bombing, it took the police three days to find the suspect from hundreds of hours of surveillance video. It is undoubtedly an inefficient method to use manpower to find the suspect from surveillance video. If there was a pedestrian retrieval system at that time, it would save a lot of investigation time and avoid the waste of human and financial resources. The more time police spend in the search, the more likely the suspect will escape. In today's security-conscious society, a system for finding target pedestrians from surveillance video makes perfect sense. There are also two situations for pedestrian retrieval. One situation is that there is image information about the target pedestrian. Then, to perform pedestrian retrieval, it is necessary to compare the similarity between the pedestrian appearing in the video and the target. However, having image information about the target pedestrian is unusual. Generally speaking, the witnesses are more likely to have the pedestrian's description. In this case, only the attribute information of the target pedestrian exists, such as the target pedestrian's age, gender, the color of the clothes, the color of the pants, etc. This thesis will focus on the latter situation, using pedestrian attribute information to search for the target pedestrian.

In order to achieve this goal, I need to detect all the pedestrians in the surveillance video and then track them to avoid repeated detection. For each tracked person, I predict the attributes of the pedestrian. The more the attribute prediction result matches the attributes of the target pedestrian, the more likely it is the target pedestrian. Therefore, I divide this process into four parts. The first is the person detection in the surveillance video. In this thesis, YOLOv5s is used for person detection. Then I track the detected pedestrians using the ByteTrack algorithm. Pedestrians generally appear in the surveillance video for a period of time, not just one frame, so it is necessary to track the pedestrians to eliminate the repetition. After that, I propose the Attributes Correlation Network (ACN) to identify the attributes of pedestrians, and ACN is the main contribution of this thesis. At last, the pedestrian retrieval system's output is the pedestrian images that best match the attribute information of the target pedestrian. In order to get the output, the system retrieves the target pedestrian image according to the similarity between the recognition attributes of pedestrians and the target attributes.

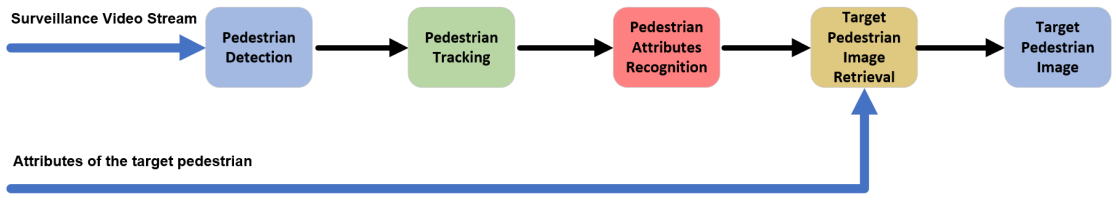

Fig1: Overall flowchart of the proposed system.

令和4年度 知能システム学専攻修士論文要旨

工藤 研究室	氏　　名	CHEN Yichen
修士論文題目	Person Retrieval System Using Attributes of Pedestrians in Surveillance Video Images	

To predict the attributes of a pedestrian, it is necessary to localize the areas from an image that are relevant to attributes since region annotations for this task are unavailable. Previous methods have tried to introduce heuristic body-part localization processes to improve local feature representations. However, they didn't use attributes to define local feature regions. In addition, correlations between attributes have not been utilized. In real-world video surveillance scenarios, visual pedestrian attributes such as gender, clothing color, and so on, are crucial for pedestrian attribute recognition. According to the correlation of attributes, it is considered useful to localize the regions. I propose an attribute estimation network for pedestrians based on spatial attention maps and attribute correlations. Spatial attention can assist the network in focusing on where each attribute should be concerned, and attribute correlation can help to reduce some incorrect predictions. In the experiments using pedestrian attribute datasets, I obtained the results that the proposed method estimates binary attributes with roughly 90% accuracy and categorical attributes with around 75% accuracy. For the mA and F1 metrics, the attribute correlation attention module can enhance performance by 1.36 and 1.57 percentage point, respectively. The spatial attention module can also improve performance by 0.57 and 0.67 percentage point.

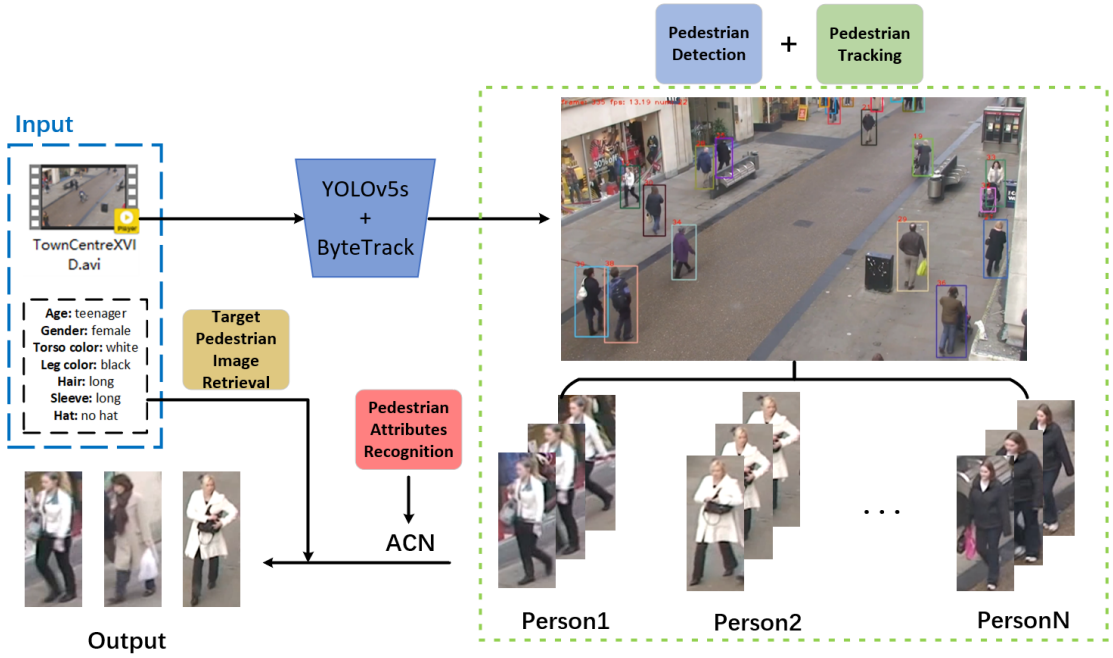

Fig2: Overall architecture of the proposed system.