

Utilizing Semantic Information in Person Re-Identification

孫, 山前

<https://hdl.handle.net/2324/7363802>

出版情報 : Kyushu University, 2024, 博士 (工学) , 課程博士
バージョン :
権利関係 :



氏 名 : 孫 山前

論 文 名 : Utilizing Semantic Information in Person Re-Identification

(意味論に基づく人物再同定方法に関する研究)

区 分 : 甲

論 文 内 容 の 要 旨

In recent decades, closed-circuit television (CCTV) surveillance has emerged as a mainstream crime prevention measure used around the world. Take the United States as an example: Even ten years ago, 49% of local police departments reported using CCTV, with usage increasing to 87% for agencies servicing jurisdictions with populations of 250,000 or more. Moreover, it is said that the number of surveillance cameras installed across Japan has exceeded five million. One application of video surveillance systems is suspect tracking. Manual methods require many police forces to watch a vast amount of surveillance videos, which is inefficient and makes it easy to miss some details. To solve this problem, we proposed an automatic pedestrian detection and match method in this research. In addition, our method introduced direction information and improved the matching accuracy compared with traditional methods.

In the field of computer vision and image processing, pedestrian re-identification refers to matching the same person appearing in different photos captured by unrelated cameras. The process of pedestrian re-identification can be divided into moving object detection, foreground segmentation, and pedestrian re-identification. This study is also based on the above process and improved some of its steps.

Chapter 1 introduces the background of our study.

Chapter 2 introduces moving object detection methods. The target of moving object detection is to divide each video frame into foreground (moving object) and background (static object). In this study, we proposed discrete Fourier transform and Otsu method-based moving object detection methods corresponding to Chapter 2.2 and Chapter 2.3. In the study based on discrete Fourier transform, we regard the pixels at the same position in consecutive video frames as a time-domain signal. Then, we transform the time-domain signal into the frequency domain and remove the part with frequency 0 (i.e., background). After the inverse transform, we get an image with only the foreground. In addition, we compared the discrete Fourier transform and fast Fourier transform. The experimental results show that the fast Fourier transform can significantly reduce the time cost when the number of images increases. The discrete Fourier transform-based moving object detection method has some problems. First, it requires manual threshold setting. In the experiment, we found that the threshold setting directly affects the integrity of the detected foreground. The second is that there will be shadows around the detected foreground. To solve the above problems, we proposed the Otsu method-based moving object detection method. The Otsu method is used to perform automatic image thresholding. Specifically, we calculate the background image by a median filter. Then,

we subtract the background image from the original image to obtain the foreground image. In binarization, we compared the Otsu method with the entropic, minimum error, and moment-preserving methods, and the experimental result shows that the Otsu method performed best for general images. However, this method performed poorly on images containing only background because it forcibly separated the background into foreground and background. Therefore, we proposed an improved method based on the average nearest neighbor method. The average nearest neighbor index describes the white pixels (foreground) as clustered or dispersed. The image containing only the background can be distinguished as white pixels scattered after binarization, while they are gathered in the image containing both foreground and background.

In Chapter 3, we proposed a metric to determine the k -value in the k -means method to segment the foreground into pedestrians. We named our metric “clusters-distance balance method.” While we classify the foreground with the k -means method, when the number of clusters increases, the average of the sum of the distances from the points to the centroid in each cluster decreases. We regard the sum of these distances as our metric. It will significantly reduce before reaching the appropriate number of clusters. However, after reaching the appropriate number of clusters, the average of the sum of the distances from the points to the centroid decreases slowly; instead, the total distance increases. We select the number of clusters corresponding to the value with the smallest total distance as the optimal number of clusters. Compared with other methods, such as the elbow method, our method considers the case where the number of classes is only 1. In addition, we also proposed a multi-angle pedestrian image matching method based on perspective transformation.

In Chapter 4, we proposed a simple template-based pose detection method in Chapter 4. In this study, we manually selected some images as the template images. Then, we calculated the sum of the distances (while the traditional Hausdorff distance-based method uses the minimum distance) between the edge pixels of the query image and the nearest edge pixel in the template image. The posture corresponding to the template image with the minimum distance is used as the posture of the query image. Our method is more robust than the traditional Hausdorff distance-based method.

Chapter 5 talks about utilizing semantic information in person re-identification. In this section, we mainly study the connection between color spaces, direction information, and matching accuracy. In our former research, we transformed images into Androutsos color space and calculated the color histograms and matching. The experimental results show that the transformed images are far different visually from the original images. And the matching accuracy is low. So, we replaced the Androutsos color space with HTML and web-safe color spaces, and the matching accuracy increased. Moreover, we regarded the direction of two adjacent images as the moving direction and separated the images into eight groups with different moving directions. We match the images only in the same groups. The experimental results show that the matching accuracy significantly improved with all of the three color spaces.

Chapter 6 summarizes this paper.