Research Progress and Trend of Person Re-identification

Guankun Wang

Shenzhen University, China

Keywords: Person re-identification; image features; deep learning; semantic information.

Abstract: Person re-identification refers to the matching technology of the same pedestrian image under different non-overlapping cameras, which has important application value in strengthening social management, preventing the occurrence of criminal acts and realizing event reconstruction. Person re-identification mainly relies on human visual representation and artificial design features and is greatly influenced by illumination, image resolution, pedestrian posture and shooting angle. Therefore, person re-identification faces great challenges. In this paper, the existing pedestrian representation feature learning technology and measurement technology are reviewed and analyzed, and the existing problems and possible solutions are pointed out. Person re-identification has great significance for researchers in this field to grasp the status quo and put forward new research ideas.

1. Introduction

It has become a consensus to using high-tech means to strengthen social management and prevent crime. In order to achieve the purposes, local governments have installed a large number of cameras at key points in public places, traffic intersections, living quarters, parking lots, etc. to strengthen the observation of pedestrian behavior and identity recognition. Camera generates huge amounts of data every day, and it is important to analyze these data. However, it is particularly difficult to acquire biological features such as face and gait in complex scenes, so person re-identification technology has emerged. Different from the traditional face recognition technology, person re-identification (ReID) establishes the corresponding relationship between the same pedestrian images from different cameras. At present, person re-identification mainly relies on visual information of human appearance, but the video image is affected by illumination, lucidity change, pedestrian posture and shooting angle of view and other factors. Even for the same pedestrian, the body appearance images taken by different cameras are quite different. The same pedestrian images taken by the same camera at different times are also various. Therefore, person re-identification is facing enormous challenges and has become a hot research topic in the field of video recognition. It has broad application prospects in social management, emergency reconstruction and so on. Therefore, person re-identification research emerged as the times required, attracting many researchers to invest in this task [1-2].

Person re-identification is a combination of pedestrian detection and person re-identification. The purpose of traditional pedestrian detection is to judge whether there are pedestrians in the input pictures or videos. It is mainly used in the fields of intelligent driving, assistant driving and intelligent monitoring. Person re-identification is to identify the designated person from the input pictures or videos. It is mainly used in image retrieval and so on. Person re-identification is the identification of the same person from videos of different cameras. It is mainly used in criminal investigation and missing persons search. In a certain period, a pedestrian may pass through several adjacent cameras. How to identify the same person under different cameras and draw the path of the target is the focus of person re-identification research, as shown in Fig.1.

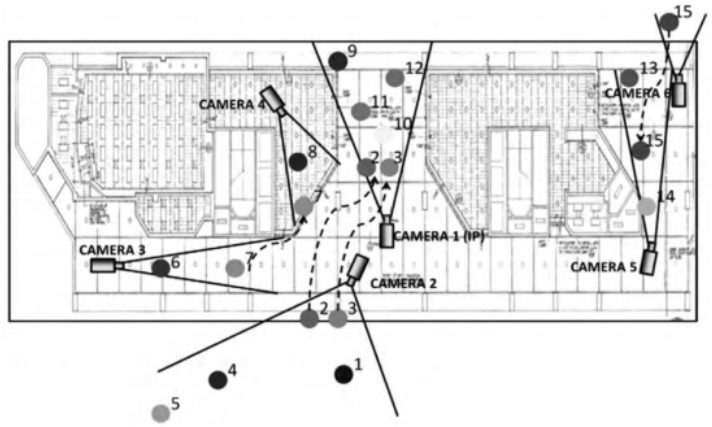

Fig 1. A sketch of person re-identification

Therefore, person re-identification is also a "re-acquisition" process. In addition to conventional video surveillance, person re-identification has also been applied to robotics and multimedia [3]. Person re-identification research began in 2003 and began to be popular in 2008. In recent years, the results have been more abundant [4]. Shangang Gong, Marco Cristani, Vittorio Murino, Fabio Roli and others have put forward different methods in this field and made much progress. In recent years, with the rapid development of machine learning, the introduction of this new technology has also brought person re-identification to a new stage.

2. Research Difficulties and Methods

From computer vision, the most challenging problem in re-recognition is how to match two images of the same person correctly under centralized appearance changes. For example: (1) the occlusion of the target results in the loss of some features. (2) Different visual and illumination conditions lead to different features of the same target. (3) Clothing color approximation and feature approximation of different targets lead to a decrease in discrimination. Based on these three problems, the main solutions are: (1) extracting more suitable features to represent the human body. (2) Choosing the appropriate distance measure function. (3) Parameter training or spatial mapping by training method makes the smaller intra-class distance and the larger inter-class distance.

At present, the major directions of person re-identification are mainly divided into two categories: feature representation and metric learning. The former is devoted to finding an obvious and robust feature for matching. The latter is devoted to finding an ideal distance scale for person re-identification [5].

Due to the different specifications of the cameras, the different shooting environments, and the slight differences in daytime, night, even sunny and rainy days, the quality of the videos may vary greatly. The difficulty of recognizing low-resolution images is much greater than that of high-resolution images.

When training classifiers, images in the Gallery Images are used, which are targeted and have good resolution. In practical applications, the resolution and brightness of the Probe Images extracted from the video captured by the camera will be much worse. As mentioned in document [6], the traditional method of scaling and normalizing low-resolution images cannot increase the effective information in photos. Therefore, the first task is to process the acquired image. Researchers tend to use multi-scale classification methods to solve the problem. At the same time, the author proposes a multi-scale learning framework. The key part of this framework is a criterion for evaluating heterogeneous mean differences in cross-scale image domain queues. Xiao Yuan Jing et al. [7] used a semi-coupled hierarchical discriminant learning method to ensure the image quality in the process of conversion from low to high pixels.

In addition, image segmentation is one of the preprocessing methods for person re-identification image. M. Farenzena et al. [8] used image segmentation to extract human foreground and used the

symmetry of pedestrian area to divide human foreground into different regions. For each region, weighted color histogram features, maximum stable color regions feature and high repeatability structural regions (Recurrent Highly Structured Patches) features are extracted to describe them. Document [9] uses the Pictorial Structure algorithm to locate the area where each "part" of a person is located in the image. For each "component" region, color histogram features like those in reference [8] and maximum stable color region features are extracted to describe them.

3. Method based on Local Features

Feature extraction is a problem that all image recognition must face. The extraction of appropriate and robust features has a great impact on the detection results and execution efficiency. Many features are used for person re-identification, such as color, texture, edge, shape, global feature, local feature and block matching feature. In order to overcome the difficulties of person re-identification, most researchers have chosen to use multiple features to synthesize in order to cope with the complex application environment.

3.1 Local Feature Extraction

Local feature-based learning mainly solves the problems of difficult learning of global features and low efficiency of feature extraction. The commonly used method is local representation. Local representation is usually computed by dividing human boundaries into cells, such as splitting an image into horizontal stripes or grids, and extracting deep features from cells. These solutions are based on the assumption that human posture is similar to the spatial distribution of the human body in the bounding box. For example, in practice, boundary boxes are detected, not hand-marked, so that humans may be in different positions, or their postures may be different. In other words, spatial partitioning is inconsistent with all parts of the human body. Person re-identification based on local features is to extract the local features of the input image, that is, to learn different features for different parts, and then connect them in series. For pedestrian matching, the representation of each part is calculated, and then the similarity between the corresponding parts is aggregated. Commonly used ideas for extracting local features include image segmentation, skeleton key point positioning and attitude correction, etc. Chunxiao Liu et al. [10] have compared several common features, including appearance, texture and color, and concluded that some specific features have better performance in person re-identification. After reading the research papers in recent years, researchers are more inclined to use local features to study person re-identification. The main reason is that the overall posture of people is quite different under different cameras. However, unless intentionally done, human limbs, trunk, clothing and accessories will not change significantly in a short period of time and space. Doug Gray et al. [11] designed a method to roughly divide the pedestrian image into three parts: head, upper body and lower body, and then describe the person in series with the color histogram of each part. SiWu et al. [12] improved the histogram of Oriented Gradients (HOG) feature. HOG is a histogram describing the distribution of gradient intensity and gradient direction in the local area of the image. The distribution can well represent the appearance and shape of the target in the local area. Therefore, HOG features can be applied to person re-identification [13]. The author notes that for a pedestrian, symmetrical features appear in the same camera, such as two arms, two shoulders and two legs. By using this method, some asymmetric targets can be excluded first, which greatly reduces the amount of calculation.

3.2 Feature Extraction based on New Technology

However, this does not mean that the study of non-local features will be stopped. When new technologies (especially hardware) have made considerable breakthroughs, researchers timely apply the latest technology to person re-identification. Reference [14] designed a method of recognition on RGB-D sensor. RGBD sensor generally refers to the sensor that can obtain both environmental color value (RGB) and depth value (Depth). It can be said that it is a sensor system which integrates the functions of TOF camera, laser sensor and ordinary camera [15]. The RGB-D sensor can extract

biological features better, and make the traditional person re-identification based on appearance features better. As mentioned earlier, appearance features are one of the most frequently used features, but they cannot be extracted when the light is not good at night, rainy day and so on. Soonmin Hwang and his team [16] began to focus on solving this problem, and they came up with a unique solution: using infrared sensors to assist in thermal spectral line recognition. They also set up a data set for their method. Because the human body radiates infrared rays to the outside all the time, this method can achieve better results when the light is not good.

4. Metric Learning based Method

Metric learning is a widely used method in image retrieval. Unlike representational learning, metric learning aims to learn the similarity of two pictures through the internet [17]. In the problem of person re-identification, the similarity between different pictures of the same pedestrian is greater than that of different pictures of different pedestrians. Specifically, a mapping f(x): $\mathbf{R}^F \rightarrow \mathbf{R}^D$ is defined to map the image from the original domain to the feature domain, and then a distance measure function D (x, y): $\mathbf{R}^D \times \mathbf{R}^D \rightarrow R$ is defined to calculate the distance between the two feature vectors. Finally, by minimizing the measurement loss of the network, an optimal mapping f (x) is found to minimize the distance between two pictures of the same pedestrian (positive sample pair) and two pictures of different pedestrians (negative sample pair). And this mapping f (x) is the deep trained convolution network.

Deep Learning is the most suitable method for person re-identification in machine learning. By building a model structure like the human brain, deep learning can extract features from the bottom to the top step by step, to establish a good mapping relationship from the bottom signal to the high-level semantics [18].

4.1 Segmenting Different Features for Learning

The key of metric learning is to get a good similarity function. In open and closed environments, it is not feasible of the same re-recognition method because of the differences of the background, perspective and light. These factors must be taken into account when learning. For example, feature extraction, if the human body is "segmented" and part of the similarity function is calculated, the result will be more excellent than that of the whole human body, which is also a preferred method for researchers. Jorge Garc.1a et al. [19] proposed a LOMO method, which analyzed the Horizontal Occurrence of Local Features and maximized the probability of occurrence of events in order to obtain stable representativeness and overcome the change of perspective. At the same time, a scale-invariant texture feature operation and a homomorphic filtering transformation are established to deal with the change of light. Dapeng Chen et al. [20] used a similarity function to maximize the probability of matching the same pedestrian and proposed a clear binomial feature kernel mapping to describe all similar information. Yonglong Tian et al. [21] first used the traditional method to segment the human body in the photograph, and then established a pool to incorporate the different scales of each segment into the pool. Convolutional Neural Network (CNN) and photo training machines of different scales were used in training. In recognition, when there was blocking, the machine would choose the best scale for recognition. Douglas Gray and HaiTao [22] defined the color and texture features of all horizontal strips as feature pools, and used Adaboost algorithm to learn the optimal feature combination to measure the similarity of a pair of pedestrian images. Boosting, also known as reinforcement learning or promotion method, was an important ensemble learning technology, which could enhance the weak learner whose prediction accuracy was only slightly higher than that of random guess to a strong learner with high prediction accuracy. This provided an effective new idea and method for the design of a learning algorithm when it was very difficult to construct a strong learner directly. AdaBoost method was the representative of the most successful approach [23-24], but in the case of only a small number of training samples, this algorithm was often over-fitting. In order to modify this shortcoming, Yuning Du et al. [25] used the method in document [22] for reference and used a Random Ensemble of Color Features to learn the

similarity measure function through Random Forest algorithm. By doing so, better experimental results could be obtained than in document [21]. Bryan Prosser et al. [26] divided the human image into six horizontal bands on average. For each horizontal strip, the corresponding color and texture features were extracted. All features were connected in series, and the similarity measure function was learned by using a support vector machine (SVM).

4.2 Learning by Hierarchical Approach

Hierarchical learning is also a learning method, which can reduce background interference and directly obtain the most needed results. For surveillance video with a complex background, this method can undoubtedly improve the accuracy of recognition.

Ejaz Ahmed et al. [27] proposed a method for simultaneous learning of features and corresponding similarity measures. This method captures the local relationship through the middle layer of each input image, and then calculates the similarity. A near dissimilarity layer is used to compare the features of convolution image. For each patch of the input image, a follow-up layer is used to summarize the differences of the adjacent layer of each patch. Wei-Shi Zheng et al. [28] proposed a feature classifier concept with a patched fuzzy model to provide hierarchical output information through an overall part-based model. Xingyu Zeng et al. [29] proposed a new deep learning model, which could be classified by several stages of training a common backpropagation algorithm. Through a special training strategy, this algorithm can train the network layer by layer by mining hard Samples to achieve the purpose of simulating cascaded classifier. Whether from theoretical analysis or experimental proof, this method can avoid over-fitting. Ping Luo et al. [30] proposed a Switchable Deep Network for Pedestrian Detection (SDN), which combined feature learning, saliency mapping and hybrid feature representation in different parts of the human body through the hierarchy. Unlike other methods, this method combines each part allocated through a selection layer before it is selected. Meanwhile, convolution layer is used as a feature extraction of low and middle levels, and then SDN is used for fusion. Compared with the original method, this method uses the characteristics of automatic learning and is also a good innovation.

5. Future Research Directions

Person re-identification usually has high recognition accuracy in strong surveillance scenarios, but in high-difficulty data sets, the performance often decreases dramatically, especially to the complex real environment. At present, most of the re-recognition work is based on two hypotheses: given candidate box and high-precision manual annotation, which cannot be verified in practice. The end-to-end system of deep learning also puts forward higher requirements for person re-identification detection and pre-stage tracking.

In view of the impact of pedestrian detection and tracking on the accuracy of re-identification, it can be defined and optimized the loss function of location, and integrate it into the final recognition score to reduce the detection error. In the aspect of tracking, face recognition, color and non-background information are conducive to improving accuracy. In the process of tracking, pedestrians will change greatly. Using sequence diagrams can reduce the dependence on large-scale monitoring information. Moreover, it is impossible to calibrate the data collected by each camera in a practical application. It is also important to study whether the method has enough generalization ability to utilize uncalibrated camera data. In addition, adding attribute learning, video-based person re-identification, using language to retrieve pedestrians, a training set and GAN to generate data are all directions to be improved and extended.

From the latest progress listed in this paper, the focus of person re-identification research is still focused on feature extraction and training. With the improvement of computer performance and the popularity of high-definition camera, local features will still be the focus of research, and multi-feature fusion for recognition is the general trend. At the same time, some of the latest technology will bring some new feature extraction methods, and even some new features. As for training, with the rapid development of machine learning, hierarchical training has become a popular

trend. However, due to the difficulty of solving the problems mentioned above, further improvement is still needed in person re-identification. In the process of research, the accuracy of recognition, the speed of recognition and the consumption of resources are also need more consideration. In addition, pedestrian re-recognition is the re-recognition of human as a target object. If the features are slightly modified, the program can be used to recognize other target objects, which has good expansibility.

References

[1] Liu Cheng. Key Technologies of Person re-identification [D]. Beijing: Beijing University of Posts and Telecommunications, 2013.

[2] Zhu Boyi. Pedestrian detection and recognition based on depth and visual information fusion [D]. Shanghai: Donghua University, 2013.

[3] Bedagkar-Gala, Apurva, Shah Shishir K. A survey of approaches and trends in person reidentification [J]. Image and Vision Computing 32.4,2014:270-286.

[4] Gong S, Cristani M, Yan S, et al. Person re-identification[M]. London: Springer,2014.

[5] Jing X, Zhu X, Wu F. Super-resolution Person re-identification with semi-coupled low-rank discriminant dictionary learning [C]. Computer Vision and Pattern Recognition,2015.

[6] Li X, Zheng W, Wang X. Multi-Scale Learning for Low Resolution Person Re-Identification[C]. International Conference on Computer Vision,2015.

[7] Jing X, Zhu X, Wu F. Super-resolution Person re-identification with semi-coupled low-rank discriminant dictionary learning [C]. Computer Vision and Pattem Recognition,2015.

[8] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features [C] // Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010:2360-2367.

[9] Cheng DS, Cristani M, Stoppa M, et al. Custom Pictorial Structures for Re-identification [C] // BMVC. 2011,1(2):6.

[10] Liu C, Gong S, Changeloy C. Person Re-identification: what features are importance [J]. European Conference on Computer Vision,2012.

[11] Gray D, Brennan S, Tao H. Evaluating appearance models for recognition, reacquisition, and tracking[C]//Proc. IEEE Intemational Workshop on Performance Evaluation for Tracking and Surveillance (PETS).2007, 3(5).

[12] WuS, Laganiere R, Payeur P. Improving pedestrian detection with selective gradient self-similarity feature[J]. Pattem Recognition,2015,48(8):2364-2376.

[13] Yao Xueqin, Li Xiaohua, Zhou Rapids. Pedestrian detection method based on edge symmetry and HOG [J]. Computer Engineering, 2012, 38 (5): 179-182.

[14] Barbosa I B. Cristani M. Del Bue A. et al. Re -identification with rgb -d sensors [C]//Computer Vision -ECCV 2012. Workshops and Demonstrations. Springer Berlin Heidelberg.

[15] Zhu Xiaoxiao, Cao Qixin, Yang Yang, et al. Real-time creation of 3D indoor environment map based on RGB-D sensor [J] Computer Engineering and Design, 2014, 35(1):203-207.

[16] Hwang S, Park J, Kim N, et al. Multispectral pedestrian detection: Benchmark dataset and baseline[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1037-1045.

[17] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: past, present and future [DB /OL]. [2018-10-22]. https: //arxiv.org /pdf /1610.02984.pdf.

[18] Yu Kai, Jia Lei, Chen Yuqiang, et al. Yesterday, today and tomorrow of in-depth study [J]. Computer Research and Development, 2013, 50 (9): 1799-1804.

[19] Liao S, Hu Y, Zhu X. Person re-identification by Local Maximal Occurrence representation and metric learning[C]. Computer Vision and Pattern Recognition,2015.

[20] Chen D, Yuan Z, Hua G. Similarity learning on an explicit polynomial kernel feature map for person re-identification[C]. Computer Vision and Pattern Recognition,2015.

[21] Tian Y, Luo P, Wang X. Deep Learning Strong Parts for Pedestrian Detection[J]. International Conference on Computer Vision,2015.

[22] Gray D, Tao H. Viewpoint invariant person re-identification with an ensemble of localized features[M]//Computer Vision-ECCV2008.Springer Berlin Heidelberg,2008:262-275.

[23] Cao Ying, Miao Qiguang, Liu Jiachen, et al. Research progress and Prospect of AdaBoost algorithm [J]. Journal of Automation, 2013, 39 (6): 745-758.

[24] The top ten algorithms in data mining[M]. CRC Press,2009.

[25] Du Y, AiH, Lao S. Evaluation of color spaces for person re-identification[C]//Pattern Recognition (ICPR),2012 21st International Conference on. IEEE,2012:1371-1374.

[26] Prosser B, Zheng WS, Gong S, et al. Person Re-Identification by Support Vector Ranking [C]// BMVC. 2010, 2(5):6.

[27] Ahmed E, Jjones M, Kmarks T. An improved deep learning architecture for person re-identification [C]. CVPR, 2015:3908-3916.

[28] Zheng W, Li X, Xiang T. Partial Person Re-Identification [C]. International Conference on Computer Vision,2015.

[29] Zeng X, Ouyang W, Wang X. Multi-stage contextual deep learning for pedestrian detection[C]// Proceedings of the IEEE International Conference on Computer Vision.2013:121-128.

[30] Luo P, Tian Y, Wang X, et al. Switchable deep network for pedestrian detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2014:899-906.