Journal of New Media
DOI:10.32604/jnm.2020.09823
Review

Tech Science Press

A Review of Person Re-Identification

Tong Jiang[*]

Nanjing University of Information Science and Technology, Nanjing, 210000, China
[*]Corresponding Author: Tong Jiang. Email: 20161243582@nuist.edu.cn
Received: 05 August 2020; Accepted: 19 August 2020

Abstract: Recently, person Re-identification (person Re-id) has attracted more and more attention, which has become a research focus of computer vision community. Person Re-id is used to ascertain whether the target pedestrians captured by cameras in different positions at different moments are the same person or not. However, due to the influence of various complex factors, person Re-id still has a lot of holes to be filled. In this paper, we first review the research process of person Re-id, and then, two kinds of mainstream methods for person Re-id are introduced respectively, according to the different types of training data they used. After that, we introduce some specific methods for different kinds of person Re-id, including handcrafted feature descriptor and metrics learning based methods as well as neural network based methods. Then, some commonly used datasets and their performance evaluation criteria are introduced. Finally, we compare these methods in order to display their advantages and disadvantages. Last but not list, depending on the current research status and development tendency, we make a prospect for person Re-id research.

Keywords: Video surveillance system; person re-identification

1 Introduction

Along with the high speed development of modern society, massive people have migrated to cities, which not only promotes urban economy, but also boosts the pressure of safeguard in public areas. In order to timely prevent or deal with these public safety problems, lots of video surveillance equipment have been installed in every corner of public areas. However, in the face of so mass and frequent crowd flow, a video surveillance system can produce huge amounts of daily monitoring data. It is difficult to analyze quickly and accurately by manual handling or traditional computer aided analysis. Therefore, an efficient person Re-id system is demanded to deal with these plenty video data generated by multi-camera video surveillance systems, so as to reduce the labor burden in video surveillance work. Fortunately, along with the research of computer since and technology going deep, person Re-id emerges as the time require.

Person Re-id, which is regarded as a core issue of video surveillance systems, in recent years has received extensive concern. From the number of relevant papers published in top-level international conferences (see Fig. 1), it can be found that person Re-id is becoming a hotspot in related fields.

The purpose of person Re-id is to retrieve a specific pedestrian in images or videos taken by multiple cameras distributed in different locations, which can also be simply regarded as the identification and tracking of specific pedestrians in cross-camera scenes [1]. Due to the appearances of pedestrians have the characteristics of rigid and flexible, the pedestrians' appearance is susceptible to the influence of clothing, posture, visual angle, lighting, shielding, background as well as other complex factors, which makes person Re-id facing great technical challenges.

The traditional frameworks adopted by most person Re-id methods cover two core parts. The first is feature extraction i.e., extracting robust and discriminated feature vectors from images or videos to

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

represent the characteristics of pedestrians' appearance. The second part is similarity measurement. In order to compare the differences between different pedestrians, similarity comparison between feature vectors is needed to judge pedestrians' similarity. It is obvious that person Re-id and image retrieval are roughly similar. In general, there are three kinds of categorizing methods for person Re-id algorithms. First, person Re-id methods can be categorized into image-based methods and video-based methods depending on the different data types of training samples they used. The latter can achieve better performance, since the video contains more temporal information. Apart from that, person Re-id methods can also be categorized into supervised, semi-supervised as well as unsupervised person Re-id according to their needs for label information in training process. Additionally, we can also divide person Re-id methods into handcrafted feature descriptor based methods and deep learning feature based methods, according to the different feature extraction methods they used. Recently, since the deepening development of deep learning, a variety of neural network based person Re-id methods have been presented, some of them have achieved better performance than traditional methods [2], and gradually become the focus of research.

In this paper, a review of person Re-id is proposed, which is written as follows: First, in Section 2, the key historical process of person Re-id will be introduced. After that, we will introduce some image-based methods and some commonly used datasets and evaluation criteria in section 3. In Section 4, some video-based methods will be introduced. Section 5 includes the prospect of the future development tendency of person Re-id. Last but not least, a conclusion will be made in Section 6.

2 The Research Process of Person Re-id

The research of person Re-id dates from the beginning of this century. In 2003, F. Porikli et al. [3] built a non-parametric model of camera pairs by right of correlation coefficient matrix, which can explore the difference of color distribution of targets under different camera views, and realized cross-view target matching. Three years later, N. Gheissari et al. [4] first proposed the concept of "person Re-identification", and realized person Re-id by using color and salient edge histograms. After that, in 2010, for the first time, A. Krizhevsky et al. [5] published an article on person Re-id in CVPR. Since then, person Re-id have been occupying an important proportion (see Fig. 1) in either top international conferences or authoritative journals of computer vision community.

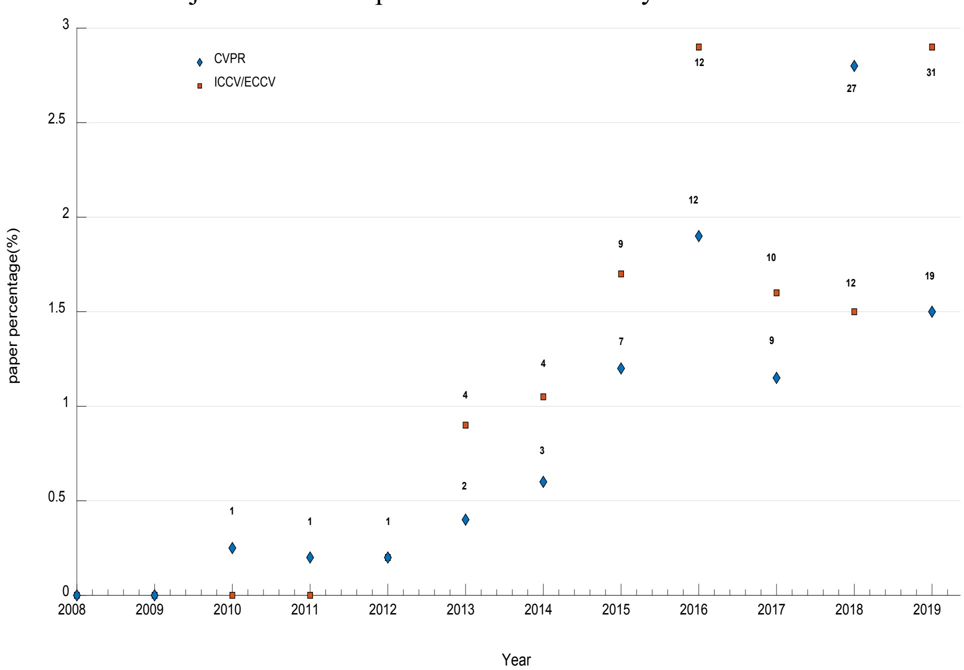

Figure 1: Numbers and percentages of person Re-id related papers on top conferences over the past years

After more than a decade of research, person Re-id has achieved a lot of significant results, and the research on person Re-id is becoming more and more impeccable. Especially in the last few years, many person Re-id specific datasets have been published, which have greatly promoted the research on person Re-id. Moreover, with the development of deep learning technology, some deep networks have also been proposed for person Re-id and obtained excellent results. Tab. 1 lists some far-reaching significant landmarks of person Re-id research.

Table 1: Landmarks of person Re-id research

Time	Author	Content
1997	T. Huang et al. [6]	Multi-cam tracking
2005	W. Zajdel et al. [7]	"Person Re-identification" first appears
2006	N. Gheissari et al. [4]	Re-id became an independent CV problem
2010	M. Farenzena et al. [1]	Video-based multi-shot-based Re-id
2014	W. Li et al. [8]	Deep learning for Re-id
2014	Y. Xu et al. [9]	Combine "Detection" and "Re-id"
2017	X. Zhang et al. [10]	Person Re-id outperform human experts

3 Image-Based Person Re-id

In 2006, N. Gheissari et al. [4] first studied person Re-id using person images. Since then, image-based methods have become a major route for person Re-id. Suppose that there is a gallery (dataset) G, represented as $G=\{g_i\}_{i=1}^{N}$, in which there are N person images belonging to N different identities. Now, for a specific probe image q, whose identity can be computed as:

$$l^* = \underset{l\in 1,2,\cdots,N}{\arg\max}\ sim(q, g_l) \tag{1}$$

where l^* is the identity of probe q, and $sim(\cdot,\cdot)$ is a specific similarity measure function. This is a common run mode of most image-based person Re-id methods.

3.1 Handcrafted Feature Based Person Re-id

For handcrafted feature based person Re-id, the attentions of researchers are mainly paid to two directions: one is the representation learning for samples, the other is metrics learning. The former focuses on how to extract more effective features, while the latter focuses more on how to better measure the similarities between samples according to the extracted features. The development process of these two directions will be introduced below.

3.1.1 Handcrafted Features for Person Re-id

There are two types of features: one is low-level feature and the other is high-level feature. The most frequently-used low-level feature is color, followed by texture. M. Farenzena et al. [1] first separated the foreground from the background, and then divided the foreground into several different blocks according to colors, after that they extracted and integrated the features of different blocks to form a whole feature vector, they regard it as the feature of the sample. Different to the above work, N. Gheissari et al. [4] calculated the HS color histogram and contour histogram for each segmented region, then, different regions were coded according to different significance levels. Moreover, D. Gray et al. [11] used 8 color channels (RGB, HS, YCbCr) and 21 texture filters to divide pedestrian's image into multiple horizontal striated areas, then extracting their features followed by concatenation. Tab. 2 is a summary of some handcrafted features.

Recent works still use low-level features for Re-id, but they have been further processed to enhance the discriminant power of low-level features. For example, R. Zhao et al. [12] extracted 32-dimensional color histogram and 128-dimensional SIFT (Scale-Independent Feature Transform) features from images for person Re-id. A. Das et al. [13] divided a person image into three parts, including head, trunk and legs, after that, they extracted HSV color feature histogram for each part. Similar to Z. Li et al. [14] also extracted the local color features, and used the Gaussian Stratification algorithm to get the spatial correlation information between the parts, while S. Pedagadi et al. [15] extracted the color histogram and used PCA method to reduce its dimension. In order to eliminate the influence of the light on the low-level features such as color, A. Datta et al. [16] took light factors into account in the person appearance model and modified the appearance model according to the distribution of light, so as to improve the robustness of the model to light. Moreover, Y. Yang et al. [17] proposed the global SCNCD feature (a color descriptor based on significant color) to mitigate the interference caused by illumination.

Since the low-level features are susceptible to factors such as illumination and occlusion, researchers are going to use middle-level features with stronger robustness. For example, E. P. Xing et al. [18] marked 15 attributes for pedestrians according to their clothing and biological characteristics, and then used these attributes for Re-id. C. X. Liu et al. [19] proposed the concept of pedestrian prototype, using some common attributes to form a pedestrian prototype for Re-id. Recently, a lot of new data and methods for attribute learning have be proposed, such as C. Su et al. [20] extracted semantic features of pedestrian images to improve feature discrimination.

Table 2: A summary of handcrafted features

Author	Year	Features	Time Info	Name
D. Gray et al. [11]	2008	color, texture	none	ELF (RGB, YCbCr, HSV, Gabor filters)
A. Krizhevsky et al. [5]	2012	CNN color, shape	none	CNN
O. Oreifej et al. [21]	2013	color, shape, texture	none	Color, LBP, HOG
R. Zhao et al. [12]	2013	color	none	dColorSIFT (Dense Color, Dense SIFT)
B. Ma et al. [22]	2014	appearance, texture, biological excitation characteristic	none	gBiCov(BIF, Gabor, Covariance Discriptor)
M. G. Gou et al. [23]	2016	color, local, shape, track	exist	LBP, HOG3D, DynFV
T. Matsukawa et al. [24]	2016	color, local, shape, gradient	none	GOG（Regional Gaussian Distribution, LAB, HSV, nRGB）
N. Mclaughlin et al. [25]	2016	color, track,CNN	exist	CNN, RNN

3.1.2 Metric Learning Methods for Person Re-id

For handcrafted feature based person Re-id, the metrics also have an important impact on results. Therefore, many researchers have turned to study the metric learning based person Re-id. Generally, metric learning can be classified into supervised and unsupervised learning, or global and local metric learning. In person Re-id problem, the key point of metric learning is to learn a global metric. In brief, the principle of metric learning is to pull the feature vectors of the same identity closer, at the same time, push the different classes' feature vectors further. Among them, the most commonly used is the metric function based on Markov distance [18], the Markov distance between and is computed as follows:

$$d\left(x,y\right)=\left(x-y\right)^{T}M\left(x-y\right) \tag{2}$$

Where M is a semidefinite matrix. E. P. Xing et al. [18] pointed out that the solution of the above equation can be convert to a convex optimization problem.

Based on the Markov distance, a lot of metrics have been proposed for person Re-id. For example, the KISSME [26] uses maximal likelihood to judge whether two vectors are similar. K. Q. Weinberger et al. [27] sets a boundary for a specific matching neighbor and punishes the non-target matching neighbor within the boundary and proposed a new nearest neighbor classification method. In order to avoid falling into overfitting, J. V. Davis et al. [28] proposed a metric learning method based on information theory, with which one can find the optimal value between overfitting and accuracy, so as to get the solution.

Due to the high computational overhead of the semi-positive matrix in Eq. (2), M. Hirzer et al. [29] proposed to use a sufficient approximate similar matrix to substitute the semi-positive matrix to reduce the burden of calculation. Z. Zhong et al. [30] introduced a method termed k-reciprocal encoding, which recoded the training images and the test images, then to compare their distances based on the recoded codes. In addition to learning a metric matrix, some researchers focus on learning discriminant subspaces instead. S. C. Liao et al. [31] presented a method to project the original feature to a low dimensional subspace, which is similar to LDA (Linear Discriminant Analysis), and the transforming formula is:

$$J(\omega)=\frac{\omega^T S_b \omega}{\omega^T S_w \omega} \tag{3}$$

where S_b is the inter-class matrix and S_w represents intra-class scatter matrix.

Based on the idea of above works, C. Zhao et al. [32] proposed a method of joint transfer constraint to learn the similarity function by combining multiple common subspaces, each in charge of a sub-region. In the common subspaces, the original samples can be reconstructed based on MvVW under the low-rank or sparse representation constraints, which can enhance the robustness and noise resistance.

Table 3: A summary of metric learning methods

Year	Author	Methods
2009	K. Q. Weinberger et al. [27]	LMNN
2009	M. Guillaumin et al. [33]	LDML
2010	B. Prosser et al. [34]	RankSVM
2012	M. Köstinger et al. [35]	KISSME
2013	W. S. Zheng et al. [36]	RDC & PRDC
2013	S. Pedagadi et al. [15]	LFDA
2015	S. C. Liao et al. [31]	XQDA
2016	X. Q. Gu et al. [37]	WLML
2018	J. Wang et al. [38]	EquiDML
2019	B. Nguyen et al. [39]	k-KISSME

3.2 Deep Learning Based Person Re-id

Since 2012, A. Krizhevsky et al. [5] proposed in ILSVRC, inspired by which, deep learning method has becoming a popular method in computer vision community. In 2014, W. Li et al. [8] at the first time applied deep learning in person Re-id, using CNN for feature extraction and Re-id. Subsequently, continuous studies have been conducted to improve the CNN's performance for person Re-id.

Soon after, lots of works have been carried out to improve the CNN network. L. Zheng et al. [40] took the invariance of pedestrian gait into consideration to extract the gait features. In [41], the feature map generated by the first two layers of the CNN is analyzed for the feature salience, so as to increase the weight of salient features in comparing phase. R. R. Varior et al. [42] proposed to integrate the LSTM (long-term memory) into the network to add the ability of network to extract temporal features. In addition, H. Liu et al. [43] also integrated the attentional mechanism into CNN network to improve the feature extracting ability of the model.

However, there is a common shortcoming of the above methods that they only consider the label information of the pedestrian images and do not fully use the attribute information. Therefore, many attribute based person Re-id methods were proposed. For instance, Y. T. Lin et al. [44] employed ResNet to extract attribute features additionally, then classified each attribute feature followed by comparing them with the corresponding attribute features of the test samples, so as to realize the person Re-id. Similarly, with the help of CNN, T. Matsukawa et al. [24] extracted the attribute features of the training samples in advance, then they used multi-attribute loss layer to classifying attribute features, and realized the person Re-id at last.

All the works mentioned above are input the original samples to the CNNs directly to learn a high-dimensional feature. Whereas, the handcrafted features can also be used as the input of the CNN networks. L. Wu et al. [45] synthesized SIFT features and color features into a high-dimensional fisher vector, then input it into the network to learn a low-dimensional feature map with the advantages of both low-dimensional features and high-dimensional features.

In 2018, many wonderful works have been presented. With the purpose of ensuring that the pixel-level features are consistent, L. X. He et al. [46] introduced a network called FCN, to generate certain-sized spatial feature maps. With FCN, they can match person images under different sizes. Furthermore, they also presented a method called DSR to avoid explicit alignment, so that the FCN can reduce the similarity between coupled images of different persons and increase the similarity between coupled images of the same person. In this way, they can identify an arbitrary patch of a pedestrian image. Since that high-level embeddings from deep CNNs may limit their accuracy on difficult examples or make them needlessly expensive for the easy ones, Y. Wang et al. [47] proposed a model, which combines effective embeddings built on multiple CNN layers, based on this model they can remedy the above drawbacks of prevailing person Re-id methods. Noticing the interference of the drastic variations in illumination, S. Bak et al. [48] introduced a new dataset containing hundreds of illumination conditions. Moreover, in order to achieve better accuracy under unseen illumination conditions, they also proposed a domain adaptation method to take the advantage of the dataset and perform an unsupervised fine-tuning. Z. Zhong et al. [49] introduced a data augmentation approach termed CamStyle to alleviate the risk of over-fitting and smooth the camera style disparities. By analyzing the different poses and viewpoints of different person, Y. Cho et al. [50] proposed a framework for person Re-id to estimate different poses and it perform multi-shot matching depending on the pose information. In [51], X. T. Zhu et al. developed a personal identification method called X-ICE hash, which is used to learn binarization and recognition of identity representation across perspectives in a unified way.

In 2019, deep learning based person Re-id methods have been further developed. Many pose-guided methods have been proposed. With the help of proposed PAC-GAN, M. X. Li et al. [52] firstly tried to enhance the unsupervised cross-view person Re-id by pose augmentation. Based on PCB, Y. F. Sun et al. [53] proposed a pooling operator called RPP, which allows the parts to be more precisely located. J. N. Li et al. [54] proposed a Part-Guided Representation (PGR) composed of Pose Invariant Feature (PIF) and Local Descriptive Feature (LDF), respectively. The PIF can approximate a pose invariant representation inferred by pose estimation and pose normalization, while LDF focuses on discriminative body parts by approximating a representation learned with body region segmentation. In this way, extra pose extraction is only introduced during the training stage to supervise the learning of PGR. In addition to pose-guided methods, there are other kinds of person Re-id methods. X. L. Qian et al. [55] proposed a deep Re-id network termed MuDeep which includes two types of layers. The first one is a multi-scale CNN layer, the other is an attention learning layer called leader-based attention learning layer. With MuDeep, they can clearly learn the importance of different spatial locations for feature extraction. In [56], Y. T. Shen et al. proposed a novel operation, called KPM (Kronecker Product Matching), to match and warp feature maps of different persons to handle viewpoint and pose variations between compared person images. Since that comparing warped feature maps results in more accurate P2G affinities, so in order to fully utilize all available P2G and G2G affinities for accurately ranking gallery person images, they also proposed a group-shuffling random walk operation. Both KPM and GRW (Group-shuffling Random Walk) operations are end-to-end trainable.

3.3 Datasets and Evaluation for Person Re-id

In order to facilitate the comparison and analysis of person Re-id methods, some commonly used datasets and performance evaluation criteria will be introduced followed by the introduction of specific person Re-id works.

3.3.1 Datasets

Many datasets for image-based person Re-id have been published. Before the emergence of deep learning technology, most person Re-id methods adopted handcrafted features [57] and verified the performance of the methods on small-sized datasets. With the development of deep learning, many large-sized datasets and deep learning based person Re-id methods have been presented. The commonly used datasets and their main indicators are shown in Tab. 4.

Table 4: Common datasets for image-based person Re-id

Dataset	Time	ID	scale	camera	label	evaluation
VIPeR	2007	632	1264	2	Hand	CMC
iLIDS	2009	119	476	2	Hand	CMC
GRID	2009	250	1275	8	Hand	CMC
CAVIAR	2011	72	610	2	Hand	CMC
PRID2011	2011	200	1134	2	Hand	CMC
WARD	2012	70	4786	3	Hand	CMC
CUHK01	2012	971	3884	2	Hand	CMC
CUHK02	2013	1816	7264	5 pairs	Hand	CMC
CUHK03	2014	1467	13164	2	Hand/DPM	CMC
RAiD	2014	43	1264	4	Hand	CMC
PRID-450S	2014	450	900	2	Hand	CMC
Market-1501	2015	1501	32668	6	Hand/DPM	CMC
PRW	2016	920	34304	6	Hand	CMC
Airpot	2017	9651	39902	6	Hand	CMC

VIPeR [58], CUHK01 [59] and Market-1501 [60] are all the datasets used for image-based person Re-id. The VIPeR is the most widely used one, which consisting of 632 people, each of them takes two photos with varying degrees of changes in lighting, perspective and posture, which makes VIPeR very challenging. CUHK01 adopts the method of manual box selection to select the pedestrian images, including 3884 images of 971 pedestrians, all of which are normalized to the pixel of 160 × 60, which have better image quality compared to VIPeR. Market-1501 contains 32668 images, which are obtained by shooting 1501 pedestrians captured by six cameras on the campus of Tsinghua University. The person frame is detected by the Deformable Part Model (DPM) automatically, so many images only contain parts of the pedestrian's body.

Different from image-based person Re-id datasets, PRID-2011 [61], iLIDS-VID [62] and MARS [63] are the datasets for video-based person Re-id methods. Two fixed non-overlap cameras are used to collect videos for PRID-2011, 385 pedestrians collected from camera A, while camera B contains 749 pedestrians. There are 200 pedestrians appeared in both cameras. Compared with PRID-2011, the iLIDS-VID is more regular and of challenging. The iLIDS-VID was gathered from the surveillance video of airport arrival hall, containing 600 videos of 300 pedestrians under two cameras. In the videos, there are light and perspective changes, complex background and occlusion, so it is difficult to identify correctly. MARS is an expanded version of the Market-1501, the number of images expanded from 32668 to 1191003. As shown in Tab. 4 and Tab. 5, with the continuous progress of person Re-id research, the scale of datasets is getting larger and larger, and the number of cameras is increasing. So far, Market-1501 is an image dataset for person Re-id that can provide enough samples for deep network training. There is another point worth noting, the bounding boxes of small datasets are manually demarcated, while in large datasets, they are generally obtained automatically by means of pedestrian detection methods such as DPM. Tab. 4 lists some commonly used datasets for image-based methods, and Tab. 5 lists some commonly used datasets for video-based methods.

Table 5: Common datasets for video-based person Re-id

Dataset	Time	ID	frame	bbox	camera	label	evaluation
ETHZ	2007	148	148	8580	1	Hand	CMC
3DPES	2011	200	1000	200k	8	Hand	CMC
PRID2011	2011	200	400	40k	2	Hand	CMC
iLIDS-VID	2014	300	600	44k	2	Hand	CMC
MARS	2016	1261	20715	1M	6	DMP/GMMCP	CMC/mAP

3.3.2 Evaluation Criteria

At present, the CMC (the cumulated matching characteristic) [64] is the most commonly used evaluation index to measure the advantages and disadvantages of a person Re-id algorithm. When a query sample is compared to the candidates, the samples in the gallery set are sorted ascendingly according to the distance between the query sample and each of them. Assuming that there are N samples in gallery set i.e. N times of query and sorting operation are performed, and the sorting result of target sample in each query is represented by $r=\left(r_1, r_2, \cdots, r_N\right)$, then CMC can be computed as:

$$CMC\left(R\right)=\frac{1}{N}\sum_{i=1}^{N}\begin{cases}1, & r_i \le R\\ 0, & r_i > R\end{cases} \tag{4}$$

It can be seen from Eq. (4) that the CMC reflects the probability of finding the correct result in the first k matching results. The horizontal axis is k, the vertical axis is the recognition rate. Generally, the recognition results of several methods will be drawn in a coordinate system, so that the difference of performances will be conspicuous.

Fig. 2 shows five CMC curves of different methods on the VIPeR dataset. When k = 1, it represents the first-round recognition rate, i.e., it represents the classification accuracy in the traditional sense: the proportion of accurately matched results. The larger value the CMC reached, the better it performed. As the number of matching candidates increases, the accuracy will inevitably improve. Therefore, CMC curve shows an increasing trend with the increase of abscissa. Since that it is difficult to make a clear distinction if there is little difference in performance between different person Re-id methods, researchers often adopt another more accurate measurement method: Rank-n table, which gives the cumulative matching accuracy of key matching points in numerical form.

The commonly used indicators are Rank-1, Rank-5, Rank-10 and Rank-20. For example, Rank-5 represents the probability of matching correctly in the first 5 images, and the higher the probability reached, the better the algorithm performed. In addition to CMC curve, L. Zheng et al. [65] proposed another evaluation index–mAP (the mean average precision), and they further pointed out that combining mAP and CMC, one can better compare the advantages and disadvantages of the method. In recent years, more and more works use CMC combine with mAP to evaluate their performances [50-57].

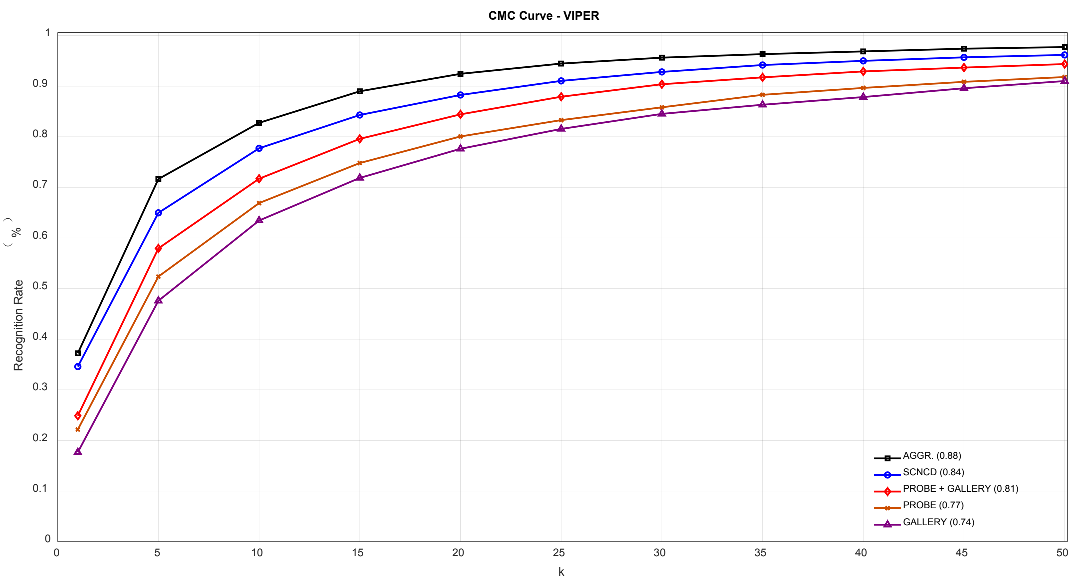

Figure 2: Five CMC curves of different methods on the VIPeR dataset

3.4 Comparison of Image-Based Person Re-id Methods

In recent years, the accuracy of person Re-id has gradually improved. Since Market-1501 was proposed in 2015, the rank-1 accuracy has been improved from 44.02% [65] to 70.04% [43]. Additionally, deep learning based person Re-id method has been the holder of the highest accuracy rate. From the Tab. 6 we can see that the deep learning based methods are much better than traditional feature descriptor based methods according to their performances on rank-k recognition accuracies, especially on CUHK03 and Market-1501. However, there is still much room for person Re-id to improve, especially with the larger datasets. For instance, it can be found from the result of Market-1501 that the highest rank-1 accuracy has reached 70.04% [43], but the mAP is only 39.55%. In addition, although existing methods have achieved a 60-70% accuracy rate on most datasets, they are still far from practical application since there are many problems remain to be solved.

Table 6: The state-of-the-art methods on image-based person Re-id datasets

Dataset	Alg.	Hand/Deep	Rank-1(%)	Rank-5(%)	Rank-10(%)	Rank-20(%)	Year
VIPeR	SCSP	Handcrafted	53.5	82.6	91.5	96.9	2016
	FFN	Deep	51.5	81	91.4	96.9	2016
	HIPHOP	Deep	54.2	82.4	91.5	96.9	2017
CUHK01	Zhang	Handcrafted	65	85	89.9	94.4	2016
	FFN	Deep	55.5	78.4	83.7	92.6	2016
	HIPHOP	Deep	78.8	92.6	95.3	97.8	2017
Market-1501	Zheng	Deep	85.8	94.4	96.4	97.5	2016
	SOMA Net	Deep	81.3	92.6	95.3	97.1	2017
	WARCA	Handcrafted	45.1	68.1	76	84	2016

In the field of deep learning based person Re-id, there are two commonly used matching methods. One is to identify all the training samples, then find the most similar one to the target sample, i.e., identification mode. The other is verification mode: given a pair of samples, the purpose of this mode is to verify whether they belong to the same person. General speaking, the identification mode is better than the verification mode.

4 Video-Based Person Re-id

With the development of deep learning based person Re-id, the size of datasets has become larger and larger, providing a firm foundation for the study of video-based person Re-id, so that video-based methods have become an emerging research focus.

The principle of video-based methods are also based on the Eq. 1, except that, two sets of bounding boxes $\{q_i\}_{i=1}^{m}$ and $\{g_j\}_{i=j}^{n}$ are used to replace the prob q and the gallery set G, where m and n represent the number of bounding boxes in the video. Different from image-based methods, video-based methods can extract more features, including gait, trajectory and other dynamic features.

4.1 Handcrafted Feature Based Video Person Re-id

4.1.1 Handcrafted Features for Video-Based Person Re-id

In 2010, people began to study video-based person Re-id [66,67,69], which commonly used color features and foreground segmentation to detect pedestrians. For example, O. Hamdoun et al. [68] used SURF (speeded up robust features) to detect and describe the points of interest. With the help of the manifold structure of video, D. N. T. Cong et al. [69] attempted to modify the color features. Considering the different resolutions of different videos, X. Y. Jing et al. [70] adjusted the resolutions of different videos to facilitate the features extracting process. Dynamic features of videos are also gradually being used. T. Q. Wang et al. [71] proposed to use dynamic features such as HOG3D and gait energy diagram to describe pedestrians. Differently, K. Liu et al. [72] proposed to add video sequence information to the final features. Moreover, depending on the periodicity of pedestrian gait as well as time sequence, C. Gao et al. [73] tried to describe the characteristics of pedestrian gait.

4.1.2 Metric Learning for Video-Based Person Re-id

Based on the feature descriptors, some researchers have investigated the metric of video-based person Re-id. W. S. Zheng et al. [74] proposed to re-rank the matching images to obtain higher accuracy. In order to reduce the interference caused by the cross-view variations, W. Li et al. [75] tried to find the best matching through local matching. X. K. Zhu et al. [76] proposed to learn the metric matrixes of images and videos simultaneously to reduce the computation. J. J. You et al. [77] presented a metric learning method termed top-push, which effectively improved the matching accuracy.

4.2 Deep Learning Based Video Person Re-id

Since video is composed of multiple continuous images, two strategies are adopted when input a video into deep neural network. One is multi-matching strategy [66], but it requires plenty of computation and is unsuitable for large-scale datasets. The second is pooling strategy [78]. The matching video clips are pooled into one unit as input, which has been proved to be effective in large-scale data. Pooling strategy can be maximized pooling or averaged pooling. In addition, many works were proposed to study how to extract dynamic features in video based on deep networks. In [78], the study showed that the appearance features did not perform well on MARS dataset, because the samples were shielded by guardrails, so, it's difficult to extract the visual features. So that it can explain why there are some works mainly focused on the use of dynamic features. N. Mclaughlin et al. [25] employed CNNs to extract the appearance features followed by the RNN to obtained the dynamic features. Based on the above works, in [79], the RNN-GRU (the gated recurrent unit) with a special structure was used to accelerate the convergence speed. With the help of the handcrafted low-level features, including color and LBP features, Y. C. Yan et al. [80] sent these features into a series of LSTMS to obtain the dynamic features of a sample. In addition, Z. X. Wu et al. [81] also proposed a CNN-RNN hybrid network to extract dynamic pedestrian features. M. Ye et al. [82] proposed the Dynamic Graph Matching (DGM) framework to refine the tag structure and improve the tag estimation process by iteratively learning better similarity measures from the middle estimated tags. In addition, they designed an active reweighting approach to improve the accuracy of intermediate labels.

Table 7: The state-of-the-art methods on video-based person Re-id datasets

Dataset	Alg.	Hand/Deep	Rank-1(%)	Rank-5(%)	Rank-10(%)	Rank-20(%)	Year
PRID-2011	Zhang	deep	83.3	93.3	-	96.7	2017
	N. Mclaughlin et al. [25]	deep	70	90	95	97	2016
	TAPR	handcrafted	68.6	94.6	97.4	98.9	2016
iLIDS-VID	Zhang	deep	60.2	85.1	-	94.2	2017
	N. Mclaughlin et al. [25]	deep	58	84	91	96	2016
	TAPR	handcrafted	55	87.5	93.8	97.2	2016
MARS	Zhang	deep	55.5	70.2	-	80.2	2017
	CNN + XQDA	deep	65.3	80.2	-	89	2016
	LOMO + XQDA	handcrafted	30.7	46.6	-	60.9	2016

4.3 Comparison Between Different Video-Based Person Re-id Methods

First, the results on the ETHZ [83] dataset have reached saturation and have been almost out of use for the last two years. In 2015, G. Lisanti et al. [84] achieved almost 100% results. They tested each video 5 times, took the average, the rank-1 accuracy had reached 99.8%, 99.7% and 99.9% respectively. It probably because that there is only a small number of pedestrians in a video sample, which is relatively simple. It can be found that due to the lack of deep learning method to use this dataset for the time being, there is no corresponding annotation.

Secondly, the commonly used datasets in recent studies are PRID-2011 and iLIDS-VID. The rank-1 accuracy of the two datasets has improved significantly in the last 4 years, reaching 50%. Furthermore, deep learning still performs well among these methods. On iLIDS-VID dataset and PRID-2011dataset, the best methods are all based on deep learning. Last but not least, based on the large-scale datasets, it is possible to train a deep learning model with strong discriminant power, which can theoretically achieve high recognition accuracy. Of course, at the present, the rank-1 accuracy and mAP still have been nowhere near their limits, so video-based person Re-id has great developing potential.

5 Future Prospects

By analyzing the development history and research status of person Re-id, we can see that after more than ten years of development, person Re-id has been vigorous growth in research methods, datasets, evaluation criteria and other aspects. Especially since the introduction of deep learning method, the test accuracy has been greatly improved, and the relevant research methods have also been greatly improved. However, there is still much space for person Re-id to improve, which is increasingly becoming the research focus of this field.

5.1 End-to-End Person Re-id

In 2016, T. Xiao et al. [85] and L. Zheng et al. [86] first proposed a real end-to-end person Re-id system. The system inputted videos and prob samples together, then integrated pedestrian detection and Re-id into a framework without manually drawing the bounding box, and directly obtains recognition results. This work not only reduces the requirement for person Re-id datasets and the dependence on manual labeling, but also makes a joint evaluation on the performance of pedestrian detection and Re-id, which greatly improves the accuracy of Re-id.

Compared with the person Re-id implemented step by step, the end-to-end system has obvious advantages. For this reason, a number of specific end-to-end person Re-id datasets were created, such as LSPS (Large scale person search) [85], PRW [86], CUHK Campus [59], and EPFL [87].

However, the end-to-end person Re-id is still in its infancy and will be improved gradually in the future.

5.2 Expanding the Datasets

In recent years, the scale of the datasets has grown rapid, from a few hundred images in the original VIPeP and iLIDS to more than half a million images in PRW and LSPS, but is still far from enough. Deep learning has a great demand for training data. Studies have shown that with the expansion of dataset size, the performance of deep learning system will improve accordingly [88]. The main limit of present person Re-id methods are the size and quality of datasets. It is conceivable that the person Re-id will have great development after more large-scale and high-quality datasets are proposed.

5.3 Real-Time Person Re-id

Real-time person Re-id plays a vital role in video surveillance system, but the existing person Re-id methods, especially the algorithms with higher accuracy, are all suffer from the large amount of computation and are difficult to achieve real-time processing. In the last few years, with the development of deep learning, researchers have made great progress in accelerating the speed of deep learning models. In 2016, D. Gray et al. [89] proposed the YOLO (You only look once) algorithm, and then L. Zheng et al. [88] proposed the SSD (single shot multi-box detector) algorithm. After tested on VOC dataset, these two algorithms both achieved over 40 FPS (frames per second), i.e., real-time detection, and that the mAP were over 70%. Real-time detection has great significance in practical applications. If the ideas of these two methods can be used for person Re-id, it will be possible to fill the gap of real-time detection in the field of person Re-id.

6 Conclusion

This paper mainly introduced the person Re-id from four aspects. The first part introduced the development history and research status of person Re-id. Then kinds of feature extracting methods in the field of image-based person methods and video-based methods were discussed after the introduction of commonly used datasets. Subsequently, the evaluation criterion of person Re-id has been simple introduction, and the CMC or mAP results of person Re-id methods proposed in different periods was compared. Finally, the future development of person Re-id was forecasted and prospected.

It is obvious that the research on person Re-id has made great progress, including the improvement of accuracy and feature extracting methods as well as even the realization of end-to-end learning. However, there is still a long way to go for further research.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Farenzena, L. Bazzani and A. Perina, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR2010*, San Francisco, California, USA, pp. 2360-2367, 2010.

[2] Y. Huang, Q. Wu and J. S. Xu, "Sbsgan: Suppression of inter-domain background shift for person re-identification," in *Proc. ICCV2019*, pp. 9527-9536, 2019.

[3] F. Porikli, "Inter-camera color calibration by correlation model function," in *Proc. ICIP2003,* vol. 2, pp. 133-136, 2003.

[4] N. Gheissari, T. B. Sebastian and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. CVPR2006*, pp. 1528-1535, 2006.

[5] A. Krizhevsky, I. Sutskever and G. E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks,” *Neural Information Processing Systems*, pp. 1097-1105, 2012.

[6] T. Huang and S. Russell, “Object identification in a bayesian context,” in *Proc. IJCAI1997*, vol. 97, pp. 1276–1282, 1997.

[7] W. Zajdel, Z. Zivkovic and B. Krose, “Keeping track of humans: Have i seen this person before?” in *Proc. IEEE International Conference on Robotics and Automation*, pp. 2081–2086, 2005.

[8] W. Li, R. Zhao and T. Xiao, “Deepreid: Deep filter pairing neural network for person re-identification,” in *Proc. CVPR2014*, pp. 152-159, 2014.

[9] Y. Xu, B. Ma, R. Huang and L. Lin, “Person search in a scene by jointly modeling people commonness and person uniqueness,” *in Proc. ACM international conference on Multimedia*, pp. 937–940, 2014.

[10] X. Zhang, H. Luo and X. Fan, “AlignedReID: surpassing human-level performance in person re-identification,” *Computer Vision and Pattern Recognition*, 2017.

[11] D. Gray and H. Tao, “Viewpoint invariant pedestrian recognition with an ensemble of localized features,” in *Proc. ECCV2008*, Marseille, France, pp. 262-275, 2008.

[12] R. Zhao, W. Ouyang and X. G. Wang, “Unsupervised salience learning for person re-identification,” in *Proc. CVPR2013,* Portland, Oregon, USA, pp. 3586-3593, 2013.

[13] A. Das, A. Chakraborty and A. K. Roy-Chowdhury, “Consistent Re-identification in a camera network,” in *Proc. ECCV2014*, Cham, Zug, Switzerland: Springer, pp. 330-345, 2014.

[14] Z. Li, S. Y. Chang and F. Liang, “Learning locally-adaptive decision functions for person verification,” in *Proc. CVPR2013*, pp. 3610-3617, 2013.

[15] S. Pedagadi, J. Orwell and S. A. Velastinm, “Local fisher discriminant analysis for pedestrian re-identification,” in *Proc. CVPR2013*, pp. 3318-3325, 2013.

[16] A. Datta, L. M. Brown, R. Feris and S. Pankanti, “Appearance modeling for person re-identification using Weighted Brightness Transfer Functions,” in *Proc. ICPR2012*, Tsukuba, Japan, 2012.

[17] Y. Yang, J. M. Yang and J. J. Yan, “Salient color names for person re-identification,” in *Proc. ECCV2014*, pp. 536-551, 2014.

[18] E. P. Xing, M. I. Jordan and S. Russell, “Distance metric learning with application to clustering with side-information,” *Neural Information Processing Systems*, pp. 521-528, 2002.

[19] C. X. Liu, S. G. Gong and C. C. Loy, “Person re-identification: what features are important?,” in *Proc. ICCV2011*, Barcelona, Spain, pp. 391-401, 2012.

[20] C. Su, F. Yang and S. Zhang, “Multi-Task Learning with Low Rank Attribute Embedding for Person Re-Identification,” in *Proc. ICCV2015*, Santiago, Chile, pp. 3739-3747, 2015.

[21] O. Oreifej and Z. C. Liu, “Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences,” in *Proc. ICCV2013*, Portland, OR, pp. 716–723, 2013.

[22] B. Ma, Y. Su and F. Jurie, “Covariance descriptor based on bioinspired feature for person re-identification and face verification,” *Image Vision Computing*, vol. 32, pp. 379–390, 2014.

[23] M. G. Gou, X Zhang and A. Ratesborras, “Person Re-identification in Appearance Impaired Scenarios,” *British Machine Vision Conference*, 2016.

[24] T. Matsukawa and E. Suzuki, “Person re-identification using CNN features learned from combination of attributes,” in *Proc. CVPR2016*, pp. 2428-2433, 2016.

[25] N. Mclaughlin, J. M. D. Rincon and P. Miller, “Recurrent convolutional network for video-based person re-identification,” in *Proc.* CVPR2016, pp. 1325-1334, 2016.

[26] M. Kostinger, M. Hirzer and P. Wohlhart, “Large scale metric learning from equivalence constraints,” in *Proc. CVPR2012*, Providence, Rhode Island, pp. 2288-2295, 2012.

[27] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *Journal of Machine Learning Research*, pp. 207-244, 2009.

[28] J. V. Davis, B. Kulis and P. Jain, “Information-theoretic metric learning,” in *Proc. ICML2007*, pp. 209-216, 2007.

[29] M. Hirzer, P. M. Roth and M. Kostinger, “Relaxed pairwise learned metric for person re-identification,” in *Proc. ECCV2012*, Firenze, Italy, pp. 780-793, 2012.

[30] Z. Zhong, L. Zheng and D. L. Cao, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. CVPR2017*, pp. 1318-1327, 2017.

[31] S. C. Liao, Y. Hu and X. Y. Zhu, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR2015*, pp. 2197-2206, 2015.

[32] C. Zhao, X. Wang and W. Zuo, "Similarity learning with joint transfer constraints for person re-identification," *Pattern Recognition*, 2020.

[33] M. Guillaumin, J. Verbeek and C. Schmid, "Is that you? metric learning approaches for face identification," in *Proc. ICCV2009*, pp. 498-505, 2009.

[34] B. Prosser, W. S. Zheng and S. Gong, "Person reidentification by support vector ranking." in *Proc. BMVC,* 2010.

[35] M. Köstinger, M. Hirzer and P. Wohlhart, "Large scale metric learning from equivalence constraints," in *Proc. CVPR2012*, pp. 2288–2295, 2012.

[36] W. S. Zheng, S. G. Gong and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653-668, 2013.

[37] X. Q. Gu and Y. X. Ge, "Weighted local metric learning for person re-identification," in *Chinese Conference on Biometric Recognition*, pp. 686-694, 2016.

[38] J. Wang, Z. Wang and C. Liang, "Equidistance constrained metric learning for person re-identification," *Pattern Recognition*, pp. 38-51, 2018.

[39] B. Nguyen and B. D. Baets, "Kernel distance metric learning using pairwise constraints for person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 589-600, 2019.

[40] L. Zheng, Y. J. Huang and H. C. Lu, "Pose invariant embedding for deep person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4500-4509, 2019.

[41] Y. Chen, X. Zhu and W. Zheng, "Person re-identification by camera correlation aware feature augmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 392-408, 2018.

[42] R. R. Varior, B. Shuai and J. W. Lu, "A siamese long short-term memory architecture for human re-identification," in *Proc. ECCV2016*, Amsterdam, The Netherlands, pp. 135-153, 2016.

[43] H. Liu, J. S. Feng and M. B. Qi, "End-to-end comparative attention networks for person re-identification," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3492-3506, 2017.

[44] Y. T. Lin, L. Zheng and Z. D. Zheng, "Improving person re-identification by attribute and identity learning," *Pattern Recognition*, pp. 151-161. 2019.

[45] L. Wu, C. H. Shen and A. V. D. Hengel, "Deep recurrent convolutional networks for video-based person re-identification: An end-to-end approach," *Computer Vision and Pattern Recognition*, 2016.

[46] L. X. He, J. Liang and H. Q. Li, "Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach," in *Proc. CVPR2018*, pp. 7073-7082, 2018.

[47] Y. Wang, L. Q. Wang and Y. R. You, "Resource aware person re-identification across multiple resolutions," in *Proc. CVPR2018*, pp. 8042-8051, 2018.

[48] S. Bak, P. Carr and J. F. Lalonde, "Domain adaptation through synthesis for unsupervised person re-identification," in *Proc. ECCV2018*, pp. 189-205, 2018.

[49] Z. Zhong, L. Zheng and Z. D. Zheng, "Camstyle: a novel data augmentation method for person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1176-1190, 2018.

[50] Y. Cho and K. Yoon, "PAMM: Pose-aware multi-shot matching for improving person re-identification," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3739-3752, 2018.

[51] X. T. Zhu, B. T. Wu and D. C. Huang, "Fast open-world person re-identification," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2286-2300, 2018.

[52] M. X. Li, X. T. Zhu and S. G. Gong, "Unsupervised tracklet person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1, 2019.

[53] Y. F. Sun, L. Zheng and Y. L. Li, "Learning Part-based Convolutional Features for Person Re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1, 2019.

[54] J. N. Li, S. L. Zhang, Q. Tian, "Pose-guided representation learning for person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1, 2019.

[55] X. L. Qian, Y. W. Fu and T. Xiang, “Leader-based multi-scale attention deep architecture for person re-identification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 371-385 2020.

[56] Y. T. Shen, T. Xiao and S. Yi, “Person re-identification with deep kronecker-product matching and group-shuffling random walk,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1, 2019.

[57] V. Balntas, K. Lenc and A. Vedaldi, “HPatches: a benchmark and evaluation of handcrafted and learned local descriptors,” in *Proc. CVPR2017*, pp. 5173-5182, 2017.

[58] D. Gray, S. Brennan and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, vol. 3, no. 5, pp. 1-7, 2007.

[59] W. Li, R. Zhao and X. G. Wang, “Human reidentification with transferred metric learning,” in *Proc. ACCV2012*, pp. 31-44, 2012.

[60] W. T. Liao, M. Y. Yang and N. Zhan, “Triplet-based deep similarity learning for person re-identification,” in *Proc. ICCV2017*, pp. 385-393, 2017.

[61] M. Hirzer, C. Beleznai and P. M. Roth, “Person re-identification by descriptive and discriminative classification,” in *Proc. SCIA2011*, Ystad Saltsjöbad, Sweden, pp. 91-102, 2011.

[62] T. Q. Wang, S. G. Gong and X. T. Zhu, “Person re-identification by discriminative selection in video ranking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2501-2514, 2016.

[63] L. Zheng, Z. Bie, Y. F. Sun, J. D. Wang, C. Su, et al. “MARS: A Video Benchmark for Large-Scale Person Re-Identification,” in *Proc.ECCV2016*, vol. 9910, pp. 868-884, 2016.

[64] R. M. Bolle, J. H. Connell, S. U. Pankanti and N. K. Ratha, “The relation between the ROC curve and the CMC,” in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies,* pp. 15-20, 2005.

[65] L. Zheng, L. Y. Shen and L. Tian, “Person re-identification meets image search,” *Computer Vision and Pattern Recognition*, 2015.

[66] L. Bazzani, M. Cristani and A. Perina, “Multiple-shot person re-identification by hpe signature,” in *Proc. ICPR2010*, Istanbul, Turkey, pp. 1413-1416, 2010.

[67] M. Farenzena, L. Bazzani and A. Perina, “Person re-identification by symmetry-driven accumulation of local features,” in *Proc. CVPR2010*, San Francisco, California, USA, pp. 2360-2367, 2010.

[68] O. Hamdoun, F. Moutarde and B. Stanciulescu, “Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences,” in *Proc. ICDSC2008*, *Las Vegas,* Nevada, USA, pp. 1-6, 2008.

[69] D. N. T. Cong, C. Achard and L. Khoudour, “Video sequences association for people re-identification across multiple non-overlapping cameras,” in *Proc. ICIAP2009*, pp. 179-189, 2009.

[70] X. Y. Jing, X. K. Zhu and F. Wu, “Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning,” in *Proc. CVPR2015*, pp. 695-704, 2015.

[71] T. Q. Wang, S. G. Gong and X. T. Zhu, “Person re-identification by video ranking,” in *Proc.ECCV2014*, pp. 688-703, 2014.

[72] K. Liu, B. P. Ma and W. Zhang, “A spatio-temporal appearance representation for viceo-based pedestrian re-identification,” in *Proc. ICCV2015,* Columbus, Ohio, pp. 3810-3818, 2015.

[73] C. Gao, J. Wang and L. Liu, “Temporally aligned pooling representation for video-based person re-identification,” in *Proc. ICIP2016*, Phoenix, Ariz, pp. 4284-4288, 2016.

[74] W. S. Zheng, S. G. Gong and T. Xiang, “Transfer re-identification: From person to set-based verification,” in *Proc. CVPR2012*, pp. 2650-2657, 2012.

[75] W. Li and X. G. Wang, “Locally aligned feature transforms across views,” in *Proc. CVPR2013*, pp. 3594-3601, 2013.

[76] X. K. Zhu, X. Y. Jing and F. Wu, “Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics,” *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5683-5695, 2018.

[77] J. J. You, A. C. Wu and X. Li, “Top-push video-based person re-identification,” in *Proc. CVPR2016*, Las Vegas, Nevada, USA, pp. 1345-1353, 2016.

[78] L. Zheng, Z. Bie and Y. F. Sun, "Mars: a video benchmark for large-scale person re-identification," in *Proc. ECCV2016*, Amsterdam, The Netherlands, pp. 868-884, 2016.

[79] L. Wu, C. H. Shen and A. V. D. Hengel, "Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification," *Computer Vision and Pattern Recognition*, 2016.

[80] Y. C. Yan, B. B. Ni and Z. C. Song, "Person re-identification via recurrent feature aggregation," in *Proc. ECCV2016*, pp. 701-716, 2016.

[81] Z. X. Wu, X. Wang and Y. G. Jiang, "Modeling spatial-temporal clues in a hybrid deep learning framework for video classification," in *Proc. ACM Multimedia*, pp. 461-470, 2015.

[82] M. Ye, J. W. Li and A. J. Ma, "Dynamic graph co-matching for unsupervised video-based person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2976-2990, 2019.

[83] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," *Brazilian Symposium on Computer Graphics and Image Processing*, pp. 322-329, 2009.

[84] G. Lisanti, I. Masi and A. D. Bagdanov, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 37, no. 8, pp. 1629-1642, 2014.

[85] T. Xiao, S. Li and B. C. Wang, "End-to-end deep learning for person search," *CVPR*, 2016.

[86] L. Zheng, H. H. Zhang and S. Y. Sun, "Person re-identification in the wild," in *Proc. CVPR2017*, Honolulu, Hawaii, pp. 3346-3355, 2017.

[87] J. Berclaz, F. Fleuret, E. Turetken, "Multiple object tracking using k-shortest paths optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1806-1819, 2011.

[88] L. Zheng, L. Y. Shen, L. Tian, S. J. Wang, J. D. Wang, "Scalable person re-identification: a benchmark," in *Proc. ICCV2015*, Santiago, Chile, pp. 1116-1124, 2015.

[89] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. CVPR2016*, pp. 779-788, 2016.