Deep Convolutional Neural Network for Person Re-identification: A Comprehensive Review

Harendra Chahar and Neeta Nain

Abstract In video surveillance, person re-identification (re-id) is a popular technique to automatically finding whether a person has been already seen in a group of cameras. In the recent years, availability of large-scale datasets, the deep learning-based approaches have made significant improvement in the accuracy over the years as compared to hand-crafted approaches. In this paper, we have distinguished the person re-id approaches into two categories, i.e., image-based and video-based approaches; deep learning approaches are reviewed in both categories. This paper contains the brief survey of deep learning approaches on both image and video person re-id datasets. We have also presented the current ongoing works, issues, and future directions in large-scale datasets.

Keywords Person re-identification · Convolutional neural network · Open-world person re-identification

1 Introduction

The definition of re-identification is introduced in [1] as follows: "To re-identify a particular subject, then, is to identify it as numerically the same particular subject as one encountered on a previous instance". In video surveillance, person identification is defined as whether the same person has been already observed at another place by different cameras. This person re-identification task is used for the safety purpose at public place, distributed large networks of cameras in public-parks, streets and university campuses, etc. It is very strenuous for human to manually monitor video surveillance systems to accurately and efficiently finding a probe or to track a person

H. Chahar (✉) · N. Nain
Department of Computer Science and Engineering,
Malaviya National Institute of Technology, Jaipur, India
e-mail: hchahar616@gmail.com

N. Nain
e-mail: nnain.cse@mnit.ac.in

B. B. Chaudhuri et al. (eds.), *Proceedings of 2nd International Conference on Computer Vision & Image Processing*, Advances in Intelligent Systems and Computing 703, https://doi.org/10.1007/978-981-10-7895-8_22

Fig. 1 Typical examples of pedestrians shot by different cameras. Each column corresponds to one person. Huge variations due to the light, pose, and viewpoint changes

across a group of cameras. A person re-id system can be divided into three parts, i.e., person detection, person tracking, and person retrieval. In this survey, person retrieval part is defined as person re-identification. In computer vision field, matching accurately two images of the same person under intensive appearance changes, such as lighting, pose, occlusion, background clutter, and viewpoint, is the most challenging problems for re-id system depicted in Fig. 1. Given its significance in research and real-world application problem, the re-id community is growing rapidly in recent years.

Few person re-id surveys already exist [2–5]. In this survey, we mainly discuss the vision part, which is also a focus in the computer vision community, another difference from previous surveys is that we focus on different re-id subtasks currently available or likely to be visible in the future, and special emphasis has been given to deep learning methods for person re-identification and issues on very large-scale person re-id datasets, which are currently popular topics or will be reflected in future trends.

This paper is organized as follows: Sect. 2 introduces a brief history of person re-id; Sect. 3 describes different kinds of deep learning approaches in image-based person re-id systems. Section 4 presents deep learning approaches in video-based person re-id systems. In Sect. 5, we present different open ongoing issues and future directions on large-scale datasets. Conclusions have drawn in Sect. 6.

2 History of Person Re-id Systems

Person re-id research problem has started with multi-camera tracking in [6]. Later, Huang and Russell [7] have proposed a Bayesian formulation to estimate the posterior of predicting the appearance of objects in one camera given evidence observed in other camera views. This appearance model combines multiple spatial-temporal features like color, velocity, vehicle length, height, and width. More details of multi-camera tracking are presented in [6].

In 2005, Wojciech Zajdel et al. [8] have proposed a method to re-identify, wherein a unique latent label is used for each person, and a dynamic Bayesian network is defined to encode the probabilistic relationship between the labels and features from the tracklets. Bayesian inference algorithm is used for determining the Id of incoming person by computing posterior label distributions.

In 2010, authors in [9, 10] have proposed technique for multi-shot person re-id. Color is a common feature used in both works, and in [10] authors additionally employ a segmentation model to detect the foreground. Minimum distance among bounding boxes in two image sets has been used for distance measurement, and authors in [9] also use the Bhattacharyya distance for the color and generic epitome features.

In 2014, Yi et al. [11] and Li et al. [12] have proposed a siamese neural network, which is used to find whether a pair of input images belong to same subject. Since then, this deep learning becomes a popular option in computer vision community for person re-id.

3 Deep Learning-Based Person Re-identification on Image Datasets

In 2006, Gheissari et al. [13] have proposed a method for person re-id based on using single images. Consider a closed-world model scenario, where G is set of N images, denoted as $\{g_i\}_{i=1}^{N}$ belongs to N different identities $1, 2, ..., N$. For a query image q, its identity is determined by:

$$i^* = argmax_{i \in 1,2,...,N} sim(q, g_i), \tag{1}$$

where sim (,) is a similarity function and i^* is the identity of query image q.

In 2012, Krizhevsky et al. [14] won the ILSVRC'12 competition with a large margin by using convolutional neural network (CNN)-based deep learning model, since then CNN-based deep learning models have been becoming popular. Two kinds of CNN model, i.e., the classification model used in image classification [14] and object detection [15], have been employed in the vision community. Since, these deep learning based CNN architecture requires the large number of training data. Therefore, currently most of the CNN-based re-id methods are using the siamese model [11]. In [12], authors have proposed a CNN model to jointly handle misalignment, photometric and geometric transforms, occlusions, and background clutter. In this model, a patch matching layer is added which multiplies the convolution responses of two images in different horizontal stripes and uses product to compute patch similarity in similar latitude.

Improved siamese model has proposed by Ahmed et al. [16], wherein the cross-input neighborhood dissimilarity features have computed, which are used to compare the features from one input image to features in neighboring locations of the other

Table 1 Statistics of image-based benchmark datasets for person re-id

Dataset	Time	#ID	#Image	#Camera	Label
VIPeR [27]	2007	632	1264	2	Hand
iLIDS [28]	2009	119	476	2	Hand
GRID [29]	2009	250	1275	8	Hand
CUHK01 [30]	2012	971	3884	2	Hand
CUHK02 [31]	2013	1816	7264	10	Hand
CUHK03 [12]	2014	1467	13164	2	Hand/DPM
PRID 450S [32]	2014	450	900	2	Hand
Market-1501 [33]	2015	1501	32668	6	Hand/DPM

image. Varior et al. [17] have proposed a system based on a siamese network, which uses long short-term memory (LSTM) modules. This module is used to store spatial connection to enhance the discriminative ability of the deep features by sequential access of image parts. In [18], authors have proposed a method to find effective subtle patterns in testing of paired images fedded into the network by inserting a gating function after each convolutional layer. In [19], siamese network has been integrated with a soft attention-based model to adaptively focus on the important local parts of paired input images. Cheng et al. [20] have proposed a triplet loss function, which takes three images as input. After the first convolutional layer, each image is partitioned into four overlapping body parts and fused with a global one in the fully connected layer.

In [21], authors have been proposed a three-stage learning process for attribute prediction based on an independent dataset and an attributes triplet loss function has trained on datasets with id labels. In [22], training set consists of identities from multiple datasets and a softmax loss is used in the classification. This method provides good accuracy on large datasets, such as PRW [23] and MARS [24] without careful training sample selection. In [25], authors have proposed a method, wherein a single Fisher vector [26] for each image has been constructed by using SIFT and color histograms aggregation. Based on the input Fisher vectors, a fully connected network has been build and linear discriminative analysis is used as an objective function which provides high inter-class variance and low intra-class variance.

3.1 Accuracy on Different Datasets Over the Years

Different kinds of datasets have been released for image-based person re-id such as VIPeR [27], GRID [29], iLIDS [28], CUHK01 [30], CUHK02 [31], CUHK03 [12], and Market-1501 [33]. The statistics about these datasets have been provided in Table 1. From this table, we have observed that the size of datasets has been increased over the years. As compared to earlier datasets, recent datasets, such as CUHK03

and Market-1501, have over the 1000 subjects which is good amount for training the deep learning models. Still, computer vision community is looking for large amount of datasets to train the models because deep learning models fully depend on datasets and provide good performance on larger datasets.

For the evaluation, the cumulative matching characteristics (CMC) curve and mean average precision (mAP) are usually used in both image and video datasets for person re-identification methods. CMC calculate the probability that a query image appears in gallery datasets. No matter how many ground truth matches in the gallery, only the first match is counted in the CMC calculation. If there exist multiple ground truths in the gallery, then mean average precision (mAP) is used for evaluation, which provides all the true matches belong in the gallery datasets to the query image.

From Table 2, we have observed that improvement in rank-1 accuracy on the different datasets VIPeR [27], CUHK01 [30], CUHK03 [12], PRID [32], iLIDS [28], and Market-1501 [33] over the years. We have observed highest rank-1 accuracy on

Table 2 Rank-1 accuracy of different image-based person re-identification approaches based on deep learning architecture on different datasets, i.e., (VIPeR, CUHK-01, CUHK-03, PRID, iLIDS, and Market-1501)

Authors/year	Evolution	VIPeR (%)	CUHK-01 (%)	CUHK-03 (%)	PRID (%)	iLIDS (%)	Market-1501 (%)
D Y [34] (2014)	CMC	28.23	–	–	–	–	–
Wei Li [12] (2014)	CMC	–	27.87	20.65	–	–	–
Ahmed [16] (2015)	CMC	34.81	65.0	54.74	–	–	–
Shi-Zhe Chen [35] (2016)	CMC	38.37	50.41	–	–	–	–
Lin Wu [36] (2016)	CMC/mAP	–	71.14	64.80	–	–	37.21
Xiao [22] (2016)	CMC	38.6	66.6	**75.33**	**64.0**	**64.6**	–
Chi-Su [21] (2016)	CMC/mAP	43.5	–	–	22.6	–	39.4
Cheng [20] (2016)	CMC	**47.8**	53.7	–	22.0	60.4	–
Hao Liu [19] (2016)	CMC/mAP	–	**81.04**	65.65	–	–	48.24
Varior [17] (2016)	CMC/mAP	42.4	–	57.3	–	–	61.6
Varior [18] (2016)	CMC/mAP	37.8	–	68.1	–	–	**65.88**
Wang [37] (2016)	CMC	35.76	71.80	52.17	–	–	–

these datasets 47.8%, 81.04%, 75.33%, 64.0%, 64.6%, and 65.88% from these works [18–20, 22], respectively. Except the VIPeR dataset, from the literature, we have observed that deep learning methods provided new state of the art on remaining five datasets as compared to hand-crafted person re-id systems. We have also observed overwhelming advantage of deep learning [18, 22] on largest datasets CUHK03 and Market-1501 so far. The improvement in object detection and image classification methods using deep learning in the next few years will also continuously dominate person re-id community. We have also observed that rank-1 accuracy is 65.88% and mAP is 39.55% , which is quite low, on Market-1501 dataset. This indicates that although it is relatively easy to find rank-1 accuracy, it is not trivial to locate the hard positives and thus achieve a high recall (mAP). Therefore, there is still much room for further improvement, especially when larger datasets are to be released and important breakthroughs are to be expected in image-based person re-id.

4 Deep Learning-Based Person Re-identification on Video Datasets

In recent years, video-based person re-id has become popular due to the increased data richness which induces more research possibilities. It shares a similar formulation to image-based person re-id as Eq. 1. Video-based person re-id replaces images q and g with two sets of bounding boxes $\{q_i\}_{i=1}^{n_q}$ and $\{g_j\}_{j=1}^{n_g}$, where n_q and n_g are the number of bounding boxes within each video sequence, respectively.

The common difference between video-based and image-based person re-id is that there are multiple images for each video sequence. Therefore, either a multi-match strategy or a single-match strategy should be employed after video pooling. In the previous works [9, 10], multi-match strategy has been used which requires higher computational cost. This may lead to be problematic on large datasets. Alternatively, a global vector has been constructed by aggregates frame-level features, which has better scalability called as pooling-based methods. As a consequence, recent video-based re-id methods generally use the pooling step. It can be either max/average pooling as [24, 38] or learned by a fully connected layer [39].

In [24], authors have proposed a system which does not require to capture the temporal information explicitly, wherein the images of subjects are used as its training samples to train a classification CNN model with softmax loss. Max pooling has been used to aggregate the frame features which provided the competitive accuracy on three datasets. Hence, these methods have been proven to be effective. Still, there is room for improvement at this stage, and the person re-id community is looking to take ideas from community of action/event recognition.

Fernando et al. [40] have proposed model which is used to capture frame features generated over the time in a video sequence. Wang et al. [41] have proposed a model,

wherein CNN model is embedded with a multi-level encoding layer and provides video descriptors of different sequence lengths.

In recent works of [38, 39, 42], where appearance features such as color and LBP are used as the starting point into recurrent neural networks to capture the time flow between frames. In [38], authors have been proposed a model, wherein CNN is used to extract features from consecutive video frames, after that these features are fedded through a recurrent final layer. Max or average pooling is used to combine features to produce an appearance feature for the video. In [42], authors have used the similar architecture as [38] with miner difference. The special kind of recurrent neural network, the gated recurrent unit, and an identification loss are used, which provide loss convergence and improve the performance. Yan et al. [39] and Zheng et al. [24] have proposed models which use the identification model to classify each input video into their respective subjects, and hand-crafted low-level features (i.e., color and local binary pattern) are fed into many LSTMs. The output of these is connected to a softmax layer. Wu et al. [43] have proposed a model to extract both spatial-temporal and appearance features from a video. A hybrid network is build by fusing these two types of features. From this survey, we may conclude that spatial-temporal models and discriminative combination of appearance are efficient solution in future video person re-id research community.

There exist many video-based person re-id datasets such as ETHZ [44], PRID-2011 [46], 3DPES [45], iLIDS-VID [47], MARS [24]. The statistics about these datasets have been provided in Table 3. The MARS dataset [24] was recently released which is a large-scale video re-id dataset containing 1,261 identities in over 20,000 video sequences. From Table 4, we have observed highest Rank-1 accuracy on iLIDS-VID and PRID-2011 datasets 58%, 70% respectively. Deep learning

Table 3 Statistics of video-based benchmark datasets for person re-id

Dataset	Time	#ID	#Track	#Bbox	#Camera	Label
ETHZ [44]	2007	148	148	8580	1	Hand
3DPES [45]	2011	200	1000	200 k	8	Hand
PRID-2011 [46]	2011	200	400	40 k	2	Hand
iLIDS-VID [47]	2014	300	600	44 k	2	Hand
MARS [24]	2016	1261	20715	1 M	6	DPM&GMMCP

Table 4 Rank-1 accuracy of different video-based person re-identification approaches based on deep learning architecture on different datasets, i.e., iLIDS-VID and PRIQ-2011

Authors/year	Evaluation	iLIDS-VID (%)	PRIQ-2011 (%)
Wu [42]	CMC	46.1	69.0
Yan [39]	CMC	49.3	58.2
McLaughlin [38]	CMC	**58**	**70**

methods are producing overwhelmingly superior accuracy in video-based person re-id. On both the iLIDS-VID and PRID-2011 datasets, the best performing methods are based on the convolutional neural network with optional insertion of a recurrent neural network [38].

5 Currently Ongoing Underdeveloped Issues and Future Directions

Annotating large-scale datasets has always been a focus in the computer vision community. This problem is even more challenging in person re-id, because apart from drawing a bounding box of a pedestrian, one has to assign him an ID. ID assignment is not trivial since a pedestrian may reenter the fields of view (FOV) or enter another observation camera a long time after the pedestrians first appearance. In this survey, we believe two alternative strategies can help bypass the data issue.

First, how to use annotations from tracking and detection datasets remains underexplored. The second strategy is transfer learning that transfers a trained model from the source to the target domain. Transferring CNN models to other re-id datasets can be more difficult because the deep model provides a good fit to the source. Xiao et al. [22] gather a number of source re-id datasets and jointly train a recognition model for the target dataset. Hence, unsupervised transfer learning is still an open issue for the deeply learned models.

The re-identification process can be viewed as a retrieval task, in which re-ranking is an important step to improve the retrieval accuracy. It refers to the reordering of the initial ranking result from which re-ranking knowledge can be discovered. For a detailed survey of search re-ranking methods, re-ranking is still an open direction in person re-id, while it has been extensively studied in instance retrieval.

We have observed that existing re-id works can be viewed as an identification task described in Eq. 1. In the identification task, query subjects are assumed to exist in the dataset and our aim is to determine the id of the query subject. On the other side, study of open-world person re-id systems is person verification task. This verification task is based on identification task described in Eq. 1 with one more constraint $sim(q, g_i) > h$, where h is the threshold. If this condition satisfies, then query subject q belongs to identity i^*; otherwise, subject q is determined as an outlier subject which is not presented in the dataset, although i^* is the first ranked subject in the identification phase.

Only few works have been done on open-world person re-id systems. Zheng et al. [48] have been designed a system which has dataset of several known subjects and a number of probes. The aim of this work is to achieve low false target recognition rate and high true target recognition. Liao et al. [49] have proposed a method which has two phases, i.e., detection and identification. In the first phase, it finds whether a probe subject is present in the dataset or not. In the second phase, it assigns an id to the accepted probe subject.

Open-world re-id still remains a challenging task as evidenced by the low recognition rate under low false accept rate, as shown in [48, 49].

5.1 Person Re-id in Very Large Datasets

In recent years, the size of data has increased significantly in the re-id community, which gives rise to community for use of deep learning approaches. However, it is evident that available datasets are still far from a real-world problem. We have observed that the largest dataset used in survey is 500 k [33], and evidence suggests that mAP drops over 7% compared to Market-1501 with a 19 k dataset. Moreover, in [33], approximate nearest neighbor search has used for fast retrieval with low accuracy.

From both a research and an application perspective, person re-id in very large datasets should be a critical direction in the future. There is also a need to design a person re-id systems for highly crowded scenes, e.g., in a public rally or a traffic jam. Therefore, there is a need to design an efficient method to improve both accuracy and efficiency of the person re-id systems. We also have to design a person re-id system which is robust and large-scale learning of descriptors and distance metrics. As a consequence, training a global person re-id model with adaptation to various illumination condition and camera location is a priority.

6 Conclusion

Person re-identification is gaining extensive interest in the modern scientific community. We have presented a history of person re-id systems. Then, deep learning approaches have been discussed in both images and video-based datasets. We also highlight some important open issues that may attract further attention from the community. They include solving the data volume issue, re-id re-ranking methods, and open-world person re-id systems. The integration of discriminative feature learning, detector/tracking optimization, and efficient data structures will lead to a successful person re-identification system which we believe are necessary steps toward practical systems.

References

1. Plantinga, A.: Things and persons. The Review of Metaphysics, pp. 493–519 (1961)
2. D'Orazio, T., Grazia C.: People re-identification and tracking from multiple cameras: A review., In 19th IEEE International Conference on Image Processing (ICIP), pp. 1601–1604 (2012)

3. Bedagkar, G., Apurva, Shishir K.S.: A survey of approaches and trends in person re-identification, Image and Vision Computing, Vol. 32 no. 4, pp. 270–286 (2014)
4. Gong, S., Cristani, M., Yan, S., Loy, C. C.: Person re-identification, Springer, Vol. 1 (2014)
5. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review, arXiv preprint arXiv:1307.5748 (2013)
6. Wang, X.: Intelligent multi-camera video surveillance: A review, Pattern recognition letters, Vol. 34 no. 1, pp. 3–19 (2013)
7. Huang, T., Russell, S.: Object identification in a bayesian context, In IJCAI, Vol. 97, pp. 1276–1282 (1997)
8. Zajdel, W., Zivkovic, Z., Krose, B. J. A.: Keeping track of humans: Have I seen this person before?, In Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2081–2086, IEEE (2005)
9. Bazzani, L., Cristani, M., Perina, A., Farenzena, M., Murino, V.: Multiple-shot person re-identification by hpe signature, In 20th International Conference on Pattern Recognition (ICPR), pp. 1413–1416, IEEE (2010)
10. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2360–2367, IEEE (2010)
11. Yi, D., Lei, Z., Liao, S., Li, S. Z.: Deep metric learning for person re-identification, In 22nd International Conference on Pattern Recognition (ICPR), pp. 34–39, IEEE (2014)
12. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159, (2014)
13. Gheissari, N., Sebastian, T. B., Hartley, R.: Person reidentification using spatiotemporal appearance. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 1528–1535, IEEE (2006)
14. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks, In Advances in neural information processing systems, pp. 1097–1105 (2012)
15. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587 (2014)
16. Ahmed, E., Jones, M., Marks, T. K.: An improved deep learning architecture for person re-identification, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3908–3916 (2015)
17. Varior, R. R., Shuai, B., Lu, J., Xu, D., Wang, G.: A siamese long short-term memory architecture for human re-identification, In European Conference on Computer Vision, Springer International Publishing, pp. 135–153 (2016)
18. Varior, R. R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification, In European Conference on Computer Vision, Springer International Publishing, pp. 791–808 (2016)
19. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification, arXiv preprint arXiv:1606.04404 (2016)
20. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1335–1344 (2016)
21. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep attributes driven multi-camera person re-identification, In European Conference on Computer Vision, Springer International Publishing, pp. 475–491 (2016)
22. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1249–1258 (2016)
23. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Tian, Q.: Person re-identification in the wild, arXiv preprint arXiv:1604.02531 (2016)

24. Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., Tian, Q.: Mars: A video benchmark for large-scale person re-identification, In European Conference on Computer Vision, Springer International Publishing, pp. 868–884 (2016)
25. Wu, L., Shen, C., van den Hengel, A.: Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification, Pattern Recognition (2016)
26. Perronnin, F., Snchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification, In European conference on computer vision, Springer Berlin Heidelberg, pp. 143–156 (2010)
27. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features, In European conference on computer vision, Springer Berlin Heidelberg, pp. 262–275 (2008)
28. Wei-Shi, Z., Shaogang, G., Tao, X.,: Associating groups of people, In Proceedings of the British Machine Vision Conference, pp. 23.1–23.11 (2009)
29. Loy, C. C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis, In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1988–1995, IEEE (2009)
30. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning, In Asian Conference on Computer Vision, Springer Berlin Heidelberg, pp. 31–44 (2012)
31. Li, W., Wang, X.: Locally aligned feature transforms across views, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3594–3601 (2013)
32. Roth, P. M., Hirzer, M., Kstinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification, In Person Re-Identification, pp. 247–267, Springer (2014)
33. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark, In Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
34. Yi, D., Lei, Z., Liao, S., Li, S. Z.: Deep metric learning for person re-identification, in Proceedings of International Conference on Pattern Recognition, pp. 2666–2672 (2014)
35. Chen, S. Z., Guo, C. C., Lai, J. H.: Deep ranking for person re-identification via joint representation learning, IEEE Transactions on Image Processing, Vol. 25 no.5, pp. 2353–2367 (2016)
36. Wu, L., Shen, C., Hengel, A. V. D.: Personnet: person re-identification with deep convolutional neural networks. arXiv preprint arXiv:1601.07255 (2016)
37. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: Joint learning of single-image and cross-image representations for person re-identification, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1288–1296 (2016)
38. McLaughlin, N., Martinez del Rincon, J., Miller, P.: Recurrent convolutional network for video-based person re-identification, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1325–1334 (2016)
39. Yan, Y., Ni, B., Song, Z., Ma, C., Yan, Y., Yang, X.: Person re-identification via recurrent feature aggregation, In European Conference on Computer Vision, Springer International Publishing, pp. 701–716 (2016)
40. Fernando, B., Gavves, E., Oramas, J., Ghodrati, A., Tuytelaars, T.: Rank pooling for action recognition, IEEE transactions on pattern analysis and machine intelligence (2016)
41. Wang, P., Cao, Y., Shen, C., Liu, L., Shen, H. T.: Temporal pyramid pooling based convolutional neural networks for action recognition, arXiv preprint arXiv:1503.0122 (2015)
42. Wu, L., Shen, C., Hengel, A. V. D.: Deep recurrent convolutional networks for video-based person re-identification: An end-to-end approach, arXiv preprint arXiv:1606.01609 (2016)
43. Wu, Z., Wang, X., Jiang, Y. G., Ye, H., Xue, X.: Modeling spatial-temporal clues in a hybrid deep learning framework for video classification, In Proceedings of the 23rd ACM international conference on Multimedia, pp. 461–470 (2015)
44. Ess, A., Leibe, B., Van Gool, L.: Depth and appearance for mobile scene analysis, In IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007)
45. Baltieri, D., Vezzani, R., Cucchiara, R.: 3dpes: 3d people dataset for surveillance and forensics, In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, pp. 59–64 (2011)

46. Hirzer, M., Beleznai, C., Roth, P. M., Bischof, H.: Person re-identification by descriptive and discriminative classification, In Scandinavian conference on Image analysis, Springer Berlin Heidelberg, pp. 91–102 (2011)
47. Wang, T., Gong, S., Zhu, X., Wang, S.: Person re-identification by video ranking, In European Conference on Computer Vision, Springer International Publishing, pp. 688–703 (2014)
48. Zheng, W. S., Gong, S., Xiang, T.: Towards open-world person re-identification by one-shot group-based verification, IEEE transactions on pattern analysis and machine intelligence, Vol. 38 no. 3, pp. 591–606 (2016)
49. Liao, S., Mo, Z., Zhu, J., Hu, Y., Li, S. Z.: Open-set person re-identification, arXiv preprint arXiv:1408.0872 (2014)