BRIEF INTRODUCTION AND PROSPECT FOR PERSON RE-IDENTIFICATION

Liqian Ma

Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, China
maliqian@sz.pku.edu.cn

ABSTRACT

Person re-identification (re-id) is crucial for intelligent video surveillance and has drawn wide attention. It is a popular research topic and still faces many challenges in real application scenarios. This paper is to introduce re-id including basic concepts, development history and research statement. We summary the traditional re-id methods from different points of view and discuss some new re-id trends for future research.

***Index Terms*—** Person re-identification, Deep learning, Robot, Open-world, Domain adaptation

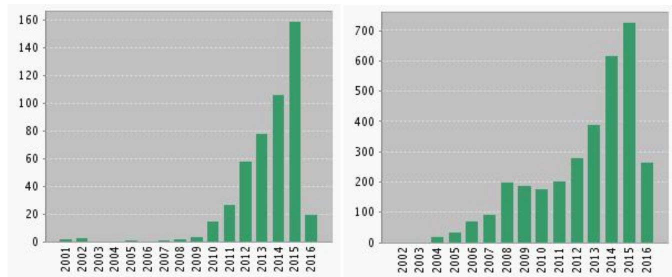
1. INTRODUCTION

Person re-identification (re-id) consists of associating an individual appearing under different cameras or at different time purely based on appearance information without spatio-temporal reasoning. Generally, re-id is formulated as a similarity measure problem, namely given the probe image of an individual re-id compares it with the gallery images and outputs the images with the most similar appearances. It is a fundamental technology which can be used for many applications such as tracking recapture, human-robot-interaction, multi-scene identification and criminal investigation.

Re-id is a new research topic. Its early related work can be tracked back to the literature about multi-camera tracking [1] in 1996. However, the concept of re-id was first formally proposed in 2006 [2]. Different from the multi-camera tracking problem, re-id only focuses on the human appearance information without the topology of camera network. Afterwards, in 2007, Gray $et\ al.$ constructed VIPeR [3] dataset, the first dataset specifically for re-id problem. Since then, re-id attracts more and more attention and the first re-id workshop is hosted by ECCV 2012 in Italy Florence. Recently, the first treatise on re-id, *Person re-identification*, is published in 2014. The trend of re-id in recent several years can be seen from the number of publications and citations shown in Fig. 1. Totally, 476 papers about person re-identification are retrieved from *Web of Science*, and about one-third of these papers are published in 2015 indicating the popularity of re-id.

Although re-id is making some process, it still faces many difficulties in both feature extraction and model learning.

[1]This trend is from Web of Science `http://apps.webofknowledge.com`



(a) Number of publications (b) Number of citations

Fig. 1. The trend of Person Re-identification. [1]

- For feature extraction, complex backgrounds, occlusions, low resolution, various illumination and viewpoints are the main challenges.
- For model learning, the inter-class similarity and inner-class dissimilarity results in significant ambiguity and make it hard to learn a discriminative model. Besides, the limited data size and different parameters of different cameras always constrain the generalization ability of models.

2. RESEARCH STATUS

The general pipeline of re-id can be divided into four basic steps shown in Fig. 2. The four steps are image input, feature extraction, feature representation and matching. Each step covers different technics.

2.1. Traditional Methods

The re-id methods can be divided into four main groups according to the image input which are single-shot re-id, multi-shot re-id, video-based re-id and zero-shot re-id. Most of the literatures are about single-shot re-id [**?**, 4–10], since it focuses on the single image description and matching which is the basis for other kinds of re-id. While multi-shot re-id [5, 11–14] focuses on information fusion to make full use of appearance information from multiple images. Furthermore, some works utilize the motion information to improve

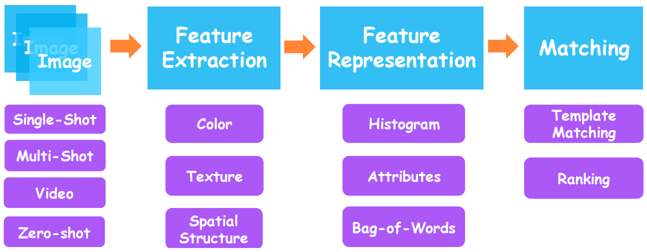


Fig. 2. The general pipeline of person re-identification.

the performance and propose the video based re-id [15–19]. Different from the previous ones, zero-shot re-id [20–24] uses a linguistic description of an individuals profile instead of appearance information.

The two steps of feature extraction and feature representation mainly focus on how to describe human appearance information effectively. According to the appearance description levels, re-id methods can be divided into two groups. The straightforward way is using low-level feature directly [5,7,12,25], such as color histogram and texture filters. However, low-level features are usually sensitive to complex background and space misalignment. In contrast, mid-level feature could be more robust to serious space misalignment and can capture more discriminative vision information [26]. Therefore, they could be robust to variations of poses and viewpoints [8–10, 27].

For the matching step, there are two common ways, namely template matching and ranking. First, template matching methods always measure the similarity between probe and gallery images with a distance metric function. Some literatures focus on the unsupervised methods [4, 5, 14, 28, 29]. While most literatures learn a Mahalanobis distance metric function in a supervised way, namely metric learning methods [25, 30–32]. Second, ranking based methods convert the re-id problem from an absolute scoring problem to a learning to rank problem which explores the pairwise relative information. Some works solve the ranking problem based on RankSVM [33, 34] to translate the ranking problem into a classification problem. Besides, Paisitkriangkrai *et al*. took the advantages of both metric learning and learning to rank with an ensemble framework [35, 36].

2.2. New Trend

Deep Learning Re-id: Compared to the traditional re-id pipeline, deep learning can combine the feature extraction, feature representation and matching steps into a unified model which could utilize end-to-end optimization and pre-training procedure to train a better model. In 2014, two methods are first proposed to apply deep learning to re-id [37, 38]. And in 2015, Ahmed *et al*. [39] proposed an improved deep learning architecture for re-id which significantly outperforms the state-of-the-arts on both a large dataset (CUHK03) and a

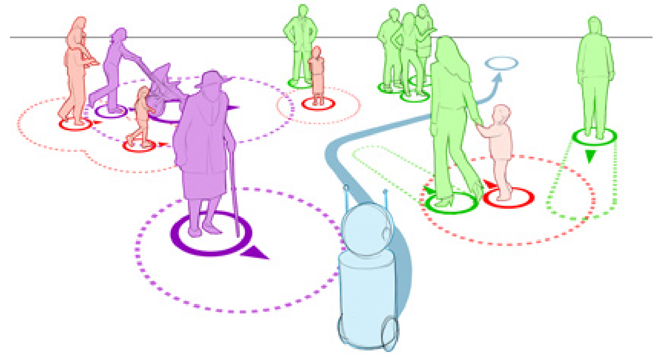


Fig. 3. The robot re-id. [2]

medium-sized dataset (CUHK01). However, due to the deficiency of training data and model complexity, the performance of deep model is limited on small datasets even with pre-training. To address the overfitting problem, Ding *et al*. designed a simplified network architecture with a simple L_2 distance metric to learn features with triplet samples, which achieves good performance on small datasets [40]. In 2016, deep learning is wildly applied to re-id and many researchers try to train end-to-end models [41–46]. Chen *et al*. combined CNN model and learning to rank model into a unified framework to perform an end-to-end optimization [41]. To construct a complete end-to-end re-id system, Xiao *et al*. proposed an end-to-end deep learning framework to jointly handle pedestrian detection and person re-id [43].

Open-World and Domain Adaptation Re-id: A central limitation of existing re-id approaches is that they are more suited to closed-world benchmark problems rather than realistic open-world scenarios. Closed-world benchmark assumes that the gallery contains the matching images for the probe image, and the methods require many pairs of person images annotated by same/different for each camera pair. However, in realistic open-world scenarios, gallery images detected by pedestrian detector may contain many distractor and junk images instead of true matches [47], and the number of camera pairs is quadratic of the number of cameras [48]. Therefore, there is a mismatch of efficiency and scalability between traditional re-id and real application scenarios. To address the problem, Liao *et al*. proposed a database for open-set re-id [49]. Zheng *et al*. treated re-id as an image retrieval problem, and utilize TF-IDF and reranking technics to improve the performance [47, 50]. Besides, to make the model domain adaptive, some researchers relax the practically unrealistic assumption of exhaustive training data within each domain and propose an effective cross-domain learning method [48,51,52]. And Xiao *et al*. propose a domain guided dropout strategy to jointly learn generic and robust deep feature representations [44].

[2] This picture is from SPENCER project `http://www.spencer.eu/`

Re-id on Robot Platform Compared to the surveillance re-id, robot re-id is much different. Robot platform (Fig. 3) has three important characteristics, namely mobility, activity and interactivity, which lead to different advantages and difficulties. Generally, there are few cameras instead of a large camera network on a robot, and the image quality is much better to obtain some high-level semantics features, such as face, gender, race, etc. Equipped with RGB-D cameras, we could segment and track the human body more easily, and anthropometric measures, such as arm length and height, can also be obtained [53–55]. Besides, the activity and interactivity can help the robot collect more useful data. These advantages can provide more and precise information for robot re-id. However, the mobility of robot leads to constantly changing scenarios for re-id, and there are more occlusions for robot re-id because of the lower camera view. In addition, the real-time is a key to robot application. So that robot re-id needs to balance the robustness, adaptivity and computation complexity of the model. As the robot re-id is still in an early stage, few works and datasets have been proposed. The earliest related work can be tracked back to 2005, which is to help robot keep track of humans who leave the field of view and re-enter later. Recently, the works for robot re-id are mainly based on RGB-D sensors which could provide more information and are suitable for close-range indoor scenes [53–58].

3. CONCLUSIONS

This paper introduces the basic concepts, development history and research statement for re-id. We summary the traditional re-id pipeline, and discuss some new trends for future research. Since large datasets are constructed, deep learning becomes more and more popular in re-id, especially the end-to-end deep model. And to address the gap between traditional re-id and real application scenarios, open-world and domain adaption re-id is necessary to study. Besides, robot re-id as a key component for human-robot interaction may become more and more popular. These new trends will guide our further research in re-id.

4. REFERENCES

[1] Q. Cai and J. K. Aggarwal, “Tracking human motion using multiple cameras,” in *Proc. of ICPR*, pp. 68–72, 1996.

[2] N. Gheissari, T. B. Sebastian, and R. Hartley, “Person reidentification using spatiotemporal appearance,” in *Proc. of CVPR*, vol. 2, pp. 1528–1535, IEEE, 2006.

[3] D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” in *Proc. of PETS workshops*, 2007.

[4] B. Ma, Y. Su, and F. Jurie, “Bicov: a novel image representation for person re-identification and face verification,” in *Proc. of BMVC*, pp. 57.1–57.11, 2012.

[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven accumulation of local features,” in *Proc. of CVPR*, pp. 2360–2367, 2010.

[6] R. Zhao, W. Ouyang, and X. Wang, “Person re-identification by salience matching,” in *in Proceedings of ICCV*, pp. 2528–2535, 2013.

[7] C. Liu, S. Gong, C. C. Loy, and X. Lin, “Person re-identification: what features are important?,” in *in Proceedings of ECCV Workshops*, pp. 391–401, 2012.

[8] Z. Yang, L. Jin, and D. Tao, “A comparative study of several feature extraction methods for person re-identification,” in *in Proceedings of Biometric Recognition*, pp. 268–277, 2012.

[9] “Learning Mid-level Filters for Person Re-identifiation,”

[10] L. An, M. Kafai, S. Yang, and B. Bhanu, “Person re-identification with reference descriptor,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–12, 2015.

[11] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, “Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences,” in *Proc. of ICDSC*, pp. 1–6, IEEE, 2008.

[12] L. Bazzani, M. Cristani, A. Perina, and V. Murino, “Multiple-shot person re-identification by chromatic and epitomic analyses,” vol. 33, no. 7, pp. 898–903, 2012.

[13] D. Baltieri, R. Vezzani, and R. Cucchiara, “Sarc3d: a new 3d body model for people tracking and re-identification,” in *Proc. of ICIAP*, pp. 197–206, 2011.

[14] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom Pictorial Structures for Re-identification.," in *in Proceedings of BMVC*, vol. 2, p. 6, 2011.

[15] D. Simonnet, M. Lewandowski, S. A. Velastin, J. Orwell, and E. Turkbeyler, "Re-identification of pedestrians in crowds using dynamic time warping," in *Proc. of ECCV Workshops*, pp. 423–432, 2012.

[16] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. of ECCV*, pp. 688–703, 2014.

[17] K. Liu, B. Ma, W. Zhang, and R. Huang, "A spatio-temporal appearance representation for viceo-based pedestrian re-identification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3810–3818, 2015.

[18] M. Gou, X. Zhang, A. Rates-Borras, S. Asghari-Esfeden, M. Sznaier, and O. Camps, "Person re-identification in appearance impaired scenarios," *arXiv preprint arXiv:1604.00367*, 2016.

[19] J. You, A. Wu, X. Li, and W. Zheng, "Top-push video-based person re-identification," *arXiv preprint arXiv:1604.08683*, 2016.

[20] R. Layne, T. M. Hospedales, S. Gong, *et al.*, "Person Re-identification by Attributes," in *in Proceedings of BMVC*, vol. 2, p. 3, 2012.

[21] R. Layne, T. M. Hospedales, and S. Gong, "Towards person identification and re-identification with attributes," in *Proc. of ECCV Workshops*, pp. 402–412, Springer, 2012.

[22] R. Satta, F. Pala, G. Fumera, and F. Roli, "People search with textual queries about clothing appearance attributes," in *Person Re-Identification*, pp. 371–389, Springer, 2014.

[23] R. Layne, T. M. Hospedales, and S. Gong, "Re-id: Hunting attributes in the wild," in *Proc. of BMVC*, British Machine Vision Association, BMVA, 2014.

[24] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao, "Multi-task learning with low rank attribute embedding for person re-identification," in *Proc. of ICCV*, pp. 3739–3747, 2015.

[25] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. of CVPR*, pp. 2288–2295, 2012.

[26] Y. L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proceedings of CVPR*, pp. 2559–2566, 2010.

[27] H. Liu, L. Ma, and C. Wang, "Body-structure based feature representation for person re-identification," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1389–1393, 2015.

[28] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *Proc. of ECCV Workshops*, pp. 413–422, Springer, 2012.

[29] K. E. Aziz, D. Merad, and B. Fertil, "People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation," in *Proc. of AVSS*, pp. 303–308, IEEE, 2011.

[30] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. of ICCV*, pp. 498–505, IEEE, 2009.

[31] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. of ICML*, pp. 209–216, 2007.

[32] K. Q. Weinberger and L. K. Saul, "Fast solvers and efficient implementations for distance metric learning," in *Proc. of ICML*, pp. 1160–1167, ACM, 2008.

[33] B. Prosser, W. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person Re-identification by Support Vector Ranking," in *Proc. of BMVC*, vol. 1, pp. 5–15, 2010.

[34] S. Shi, C. Guo, J. Lai, S. Chen, and X. Hu, "Person re-identification with multi-level adaptive correspondence models," *Neurocomputing*, vol. 168, pp. 550–559, 2015.

[35] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. of CVPR*, pp. 1846–1855, 2015.

[36] S. Paisitkriangkrai, L. Wu, C. Shen, and A. v. d. Hengel, "Structured learning of metric ensembles with application to person re-identification," *arXiv preprint arXiv:1511.08531*, 2015.

[37] D. Yi, Z. Lei, and S. Z. Li, "Deep metric learning for practical person re-identification," in *Proc. of ICPR*, pp. 34–39, 2014.

[38] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proc. of CVPR*, pp. 152–159, 2014.

[39] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. of CVPR*, pp. 3908–3916, 2015.

[40] S. Ding, L. Lin, G. Wang, and H. Chao, “Deep feature learning with relative distance comparison for person re-identification,” *Pattern Recognition*, vol. 48, no. 10, pp. 2993–3003, 2015.

[41] S. Chen, C. Guo, and J. Lai, “Deep ranking for person re-identification via joint representation learning,” *IEEE Transactions on Image Processing*, pp. 1–15, 2016.

[42] L. Wu, C. Shen, and A. v. d. Hengel, “Personnet: Person re-identification with deep convolutional neural networks,” *arXiv preprint arXiv:1601.07255*, 2016.

[43] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “End-to-end deep learning for person search,” *arXiv preprint arXiv:1604.01850*, 2016.

[44] T. Xiao, H. Li, W. Ouyang, and X. Wang, “Learning deep feature representations with domain guided dropout for person re-identification,” *arXiv preprint arXiv:1604.07528*, 2016.

[45] L. Zheng, H. Zhang, S. Sun, M. Chandraker, and Q. Tian, “Person re-identification in the wild,” *arXiv preprint arXiv:1604.02531*, 2016.

[46] S. Wu, Y. Chen, X. Li, A. Wu, J. You, and W. Zheng, “An enhanced deep feature representation for person re-identification,” in *Proc. of WACV*, pp. 1–8, IEEE, 2016.

[47] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in *Proc. of ICCV*, pp. 1116–1124, 2015.

[48] R. Layne, T. M. Hospedales, and S. Gong, “Domain transfer for person re-identification,” in *Proc. of ACM/IEEE international workshop on Analysis and retrieval of tracked events and motion in imagery stream*, pp. 25–32, ACM, 2013.

[49] S. Liao, Z. Mo, J. Zhu, Y. Hu, and S. Z. Li, “Open-set person re-identification,” *arXiv preprint arXiv:1408.0872*, 2014.

[50] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, “Query-adaptive late fusion for image search and person re-identification,” in *Proc. of CVPR*, pp. 1741–1750, 2015.

[51] A. Ma, P. Yuen, and J. Li, “Domain transfer support vector ranking for person re-identification without target camera label information,” in *Proc. of ICCV*, pp. 3567–3574, 2013.

[52] X. Wang, W. Zheng, X. Li, and J. Zhang, “Cross-scenario transfer person re-identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.

[53] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, “Re-identification with rgb-d sensors,” in *Proc. of ECCV Workshops and Demonstrations*, pp. 433–442, Springer, 2012.

[54] A. Mogelmose, C. Bahnsen, T. Moeslund, A. Clapés, and S. Escalera, “Tri-modal person re-identification with rgb, depth and thermal features,” in *Proc. of CVPR Workshops*, pp. 301–307, 2013.

[55] F. Pala, R. Satta, G. Fumera, and F. Roli, “Multi-modal person re-identification using rgb-d cameras,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.

[56] M. Munaro, A. Basso, A. Fossati, L. Van Gool, and E. Menegatti, “3d reconstruction of freely moving persons for re-identification with a depth sensor,” in *Proc. of ICRA*, pp. 4512–4519, IEEE, 2014.

[57] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, “A feature-based approach to people re-identification using skeleton keypoints,” in *Proc. of ICRA*, pp. 5644–5651, IEEE, 2014.

[58] G. Cushen, “A person re-identification system for mobile devices,” *arXiv preprint arXiv:1512.04133*, 2015.