Progressive Domain Adaptation for Robot Vision Person Re-identification

Conference Paper · October 2020

DOI: 10.1145/3394171.3414358

Zijun Sha
Honda R&D, Japan

Zelong Zeng
The University of Tokyo

Shin'ichi Satoh
National Institute of Informatics

Progressive Domain Adaptation for Robot Vision Person Re-identification

Zijun Sha
HONDA R&D Co.,Ltd
zijun_sha@jp.honda

Zelong Zeng
The University of Tokyo
zzlbz@nii.ac.jp

Zheng Wang*
National Institute of Informatics
wangz@nii.ac.jp

Yoichi Natori
HONDA R&D Co.,Ltd
yoichi_natori@jp.honda

Yasuhiro Taniguchi
HONDA R&D Co.,Ltd
yasuhiro_taniguchi@jp.honda

Shin'ichi Satoh
National Institute of Informatics
satoh@nii.ac.jp

ABSTRACT

Person re-identification has received much attention in the last few years, as it enhances the retrieval effectiveness in the video surveillance networks and video archive management. In this paper, we demonstrate a guiding robot with person followers system, which recognizes the follower using a person re-identification technology. It first adopts existing face recognition and person tracking methods to generate person tracklets with different IDs. Then, a classic person re-identification model, pre-trained on the surveillance dataset, is adapted to the new robot vision condition incrementally. The demonstration showcases the quality of robot follower focusing.

CCS CONCEPTS

• **Information systems** → **Evaluation of retrieval results**; • **Human-centered computing** → **Visualization systems and tools**; • **Computing methodologies** → Computer vision.

KEYWORDS

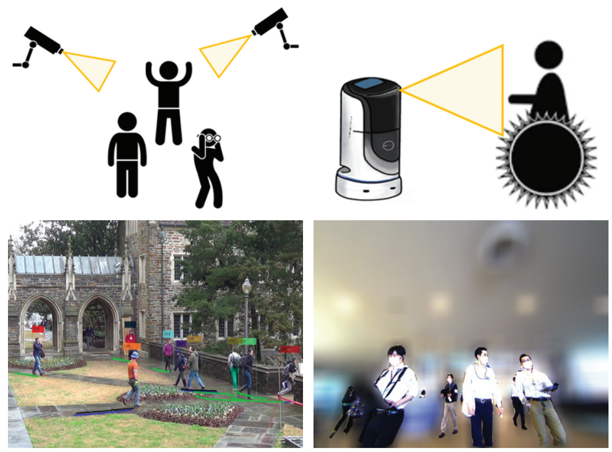
Robot Vision; Domain Adaptation; Person Re-identification

ACM Reference Format:
Zijun Sha, Zelong Zeng, Zheng Wang, Yoichi Natori, Yasuhiro Taniguchi, and Shin'ichi Satoh. 2020. Progressive Domain Adaptation for Robot Vision Person Re-identification. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3394171.3414358

1 INTRODUCTION

In the coming years, tour-guide robots will be widely used in museums, exhibitions, airports and shopping malls [6]. When the robot serves the clients, it not only plans the route and pilots the way, but also focuses the client, so that the client could catch up with the robot and keep in service.

In this demo, we pay our attention to the function of follower recognition. It will address the issue that the follower does not always appear in the camera area of the back-view camera. Generally, the robot system can use face recognition [1] and person tracking [4] to focus the follower. However, since the follower's face may not be always in front of the camera, the follower may be blocked by the others, or the follower may go to another place in a short time, the system based on face recognition and tracking does not always work well. An intuitive idea is to introduce person re-identification technology [2, 3, 7–14] to improve the results.



(a) Security Vision Re-ID (b) Robot Vision Re-ID

Figure 1: Security Vision Re-ID *v.s.* Robot Vision Re-ID.

Person re-identification (Re-ID) exploiting the person's appearance to distinguish different person images, has been widely used in video surveillance network applications [8]. To the best of our knowledge, this kind of technology has not been used in robot vision community, and no annotated images can be utilized to benefit the follower recognition. Therefore, directly exploiting existing Re-ID models to robot vision applications is not suitable.

As Fig. 1 shows, Security Vision Re-ID[1] and Robot Vision Re-ID are different. The differences lie in three aspects: 1) the setting of cameras in security vision is fixed, so the background of camera view remains almost changeless. Each camera will work under a stable illumination condition. However, the robot's moving will lead to some influence such as blur, illumination change and background change on the captured image in robot vision. 2) The camera view of security vision is from top to bottom, while the view of robot

*Zijun Sha, Zelong Zeng, Zheng Wang and Yoichi Natori have equal contributions. Zheng Wang is the corresponding author.

[1]Fig. 1 (a) is obtained from an open dataset. https://megapixels.cc/duke_mtmc/

ACM ISBN 978-1-4503-7988-5/20/10.
https://doi.org/10.1145/3394171.3414358

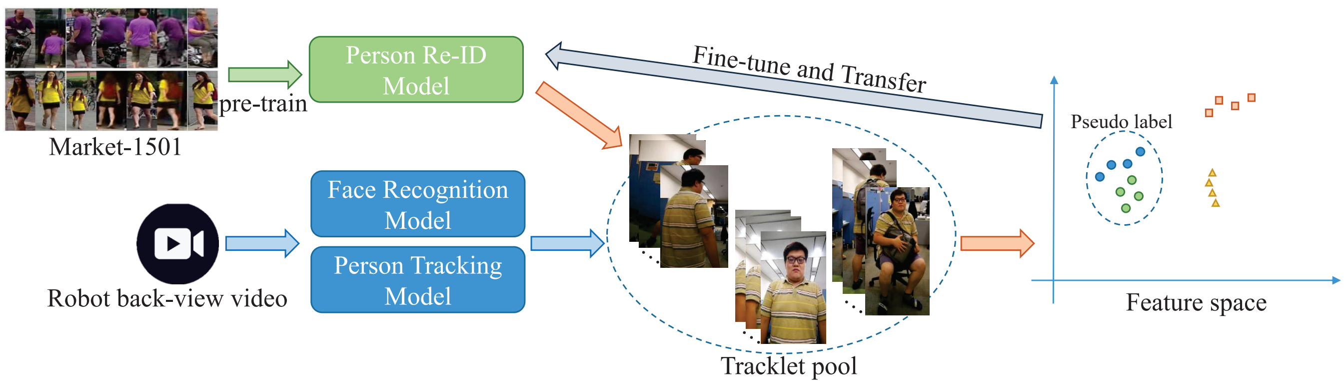


Figure 2: The framework of our system. 1) A Re-ID model is pre-trained on the Market-1501 dataset. 2) The robot back-view video generates a batch of tracklets by face recognition and person tracking models. 3) The Re-ID model extracts features of tracklets, and our system assigns pseudo labels to tracklets based on their similarities. 4) Selected tracklets and their pseudo labels are used to fine-tune the Re-ID model.

vision is from bottom to top. This makes a large gap between the appearances of these two kinds of applications. 3) Persons in robot vision have more variations. For example, the persons' clothes and face/body directions may change, and the body will also be blocked under severe occlusion much more frequently. In this demo, we propose to use the classic Re-ID model and adapt it to the robot vision application.

2 METHOD

Fig. 2 shows the framework of our system. The system consists of four key steps. In the first step, classic Re-ID models (PCB [5] and IID [15]) are pre-trained on a camera network dataset (Market-1501 [16]). In the second step, the robot back-view video is input into the existing face recognition and person tracking models, then a batch of tracklets will be generated. Each tracklet has a sequence of images and a unique ID.

After this step, there are still a lot of tracklets with different IDs even for tracklets of the same person. Our target is to assign the same ID to the tracklets of the same person. In the third step, the pre-trained Re-ID models are used to extract features from the images of the tracklet pool. Then, in the feature space, we calculate the similarities of images from different tracklets. Given a reference tracklet, if any similarity of one image from the reference tracklet $\mathcal{T}_{ref}$ and the image from another tracklet $\mathcal{T}_i$ is larger than the threshold λ_1, we will assign the same ID to the corresponding tracklet $\mathcal{T}_i$ as the reference tracklet. If all similarities of the image from the reference tracklet $\mathcal{T}_{ref}$ and the image from the tracklet $\mathcal{T}_i$ are smaller than the threshold λ_2, we will assign a new ID to the tracklet[2]. After the system checks all the images and makes the assignment. We randomly select the tracklets from the pool, and thus generate pseudo labels. In the fourth step, the selected tracklets (some batches of images) and their pseudo labels are used to fine-tune the Re-ID model. To this end, the Re-ID model adapts to the robot vision condition a bit.

[2]In our system, all similarities are normalized. And we set $\lambda_1 = 0.55$ and $\lambda_2 = 0.2$.

Actually, the third and fourth steps can be run in an iterative way, hence the model will adapt to the new domain incrementally.

3 EVALUATION

To quantitatively evaluate our new robot follower focus system with person re-identification, we have tested our approach at a public space and also collected about 3TB video data with different challenges, including heavy occlusions, illumination changes, multiple followers, faces with masks, different poses, different walking speeds, far away from the robot, disappear and reappear again, and so on. Moreover, the collected data is unlabeled. We first use existing face recognition and tracking models to obtain the results, and then adopts the proposed Progressive Domain Adaptation Re-ID model to acquire another result. In this way, a new robot vision Person Re-ID dataset with 1200 IDs has been generated. Then, we observe whether the result has been improved by our new Re-ID model. The results show that the model performs much more stable recognition after domain adaptation. This proves the effectiveness of our method.

4 CONCLUSION

In this paper, we constructed a guiding robot with person followers system by progressively adapting the classic Re-ID model. We made a comparison of security vision Re-ID and robot vision Re-ID, and demonstrated the framework of the system, in particular, how to adapt the classic Re-ID model to the robot vision condition. Several points could be further explored in the future to improve our system, including collaborating with the front-view vision and building a large scale dataset to benefit the community. A demo video is available at https://www.youtube.com/watch?v=W8W_N0vrHsQ.

ACKNOWLEDGMENTS

The research was supported by supported by the Honda-NII Collaborative Research Project, JST CREST Grant (JPMJCR1686), and Grant-in-Aid for JSPS Fellows (18F18378).

REFERENCES

[1] Jin Chen, Jun Chen, Zheng Wang, Chao Liang, and Chia-Wen Lin. 2020. Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition. *IEEE Signal Processing Letters* (2020).

[2] Lingxiao He, Yinggang Wang, Wu Liu, He Zhao, Zhenan Sun, and Jiashi Feng. 2019. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In *ICCV*.

[3] Haolin Ren, Zheng Wang, Zhixiang Wang, Lixiong Chen, Shin'ichi Satoh, and Daning Hu. 2020. An Interactive Design for Visualizable Person Re-Identification. In *ACM Multimedia*.

[4] Weijian Ruan, Wu Liu, Qian Bao, Jun Chen, Yuhao Cheng, and Tao Mei. 2019. Poinet: pose-guided ovonic insight network for multi-person pose tracking. In *ACM Multimedia*.

[5] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*.

[6] Anna-Maria Velentza, Dietmar Heinke, and Jeremy Wyatt. 2019. Human Interaction and Improving Knowledge through Collaborative Tour Guide Robots. In *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*.

[7] Zheng Wang, Xiang Bai, Mang Ye, and Shin'ichi Satoh. 2018. Incremental deep hidden attribute learning. In *ACM Multimedia*.

[8] Zheng Wang, Ruimin Hu, Chao Liang, Yi Yu, Junjun Jiang, Mang Ye, Jun Chen, and Qingming Leng. 2016. Zero-shot person re-identification via cross-view consistency. *IEEE Transactions on Multimedia* (2016).

[9] Zheng Wang, Ruimin Hu, Yi Yu, Junjun Jiang, Jiayi Ma, and Shin'ichi Satoh. 2017. Statistical inference of gaussian-laplace distribution for person verification. In *ACM Multimedia*.

[10] Zheng Wang, Ruimin Hu, Yi Yu, Chao Liang, and Wenxin Huang. 2015. Multi-level fusion for person re-identification with incomplete marks. In *ACM Multimedia*.

[11] Zheng Wang, Zhixiang Wang, Yinqiang Zheng, Yang Wu, Wenjun Zeng, and Shin'ichi Satoh. 2020. Beyond intra-modality: A survey of heterogeneous person re-identification. In *IJCAI*.

[12] Mang Ye, Xiangyuan Lan, and Qingming Leng. 2019. Modality-aware collaborative learning for visible thermal person re-identification. In *ACM Multimedia*.

[13] Mang Ye, Chao Liang, Zheng Wang, Qingming Leng, and Jun Chen. 2015. Ranking optimization for person re-identification via similarity and dissimilarity. In *ACM Multimedia*.

[14] Mang Ye, Chao Liang, Yi Yu, Zheng Wang, Qingming Leng, Chunxia Xiao, Jun Chen, and Ruimin Hu. 2016. Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing. *IEEE Transactions on Multimedia* (2016).

[15] Zelong Zeng, Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. 2020. Illumination-adaptive person re-identification. *IEEE Transactions on Multimedia* (2020).

[16] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *ICCV*.