SCIENCE CHINA
Information Sciences

• LETTER •

June 2024, Vol. 67, Iss. 6, 169101:1–169101:2
https://doi.org/10.1007/s11432-023-4023-y

Pyramid-resolution person restoration for cross-resolution person re-identification

Chunlei PENG[1,2], Bo WANG[1,2], Decheng LIU[1,2], Nannan WANG[1,3*] & Xinbo GAO[4]

[1]*State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an* 710071*, China;*
[2]*School of Cyber Engineering, Xidian University, Xi'an* 710071*, China;*
[3]*School of Telecommunications Engineering, Xidian University, Xi'an* 710071*, China;*
[4]*Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing* 400065*, China*

Received 3 April 2023/Revised 29 August 2023/Accepted 3 April 2024/Published online 16 May 2024

Person images captured by surveillance cameras are typically with various resolutions and are susceptible to pose changes, occlusion and background variations, thus affecting the performance of person ReID. Existing approaches [1, 2] either use an image restoration model to restore the low-resolution images to obtain a corresponding high-resolution image, or learn the resolution-invariant representation of the person images to reduce the impact of resolution differences. However, in unconstrained conditions, person images usually have various resolutions and it is challenging to define a suitable resolution for restoration or find robust features for various resolution situations. Although feature pyramid strategy has been used before, they applied the pyramid strategy of part image, attention region, or layer features in traditional person ReID task, while pyramid resolution restoration has not been exploited for cross-resolution ReID scenario yet. In fact, the usage of pyramid resolution restoration is reasonable in cross-resolution ReID scenario, but it is difficult to set proper fixed pyramid resolutions for restoration, especially when dealing with real-world person ReID conditions with irregular aspect ratios and resolutions of gallery images and query images. Most existing cross-resolution ReID methods are evaluated on simulated datasets with fixed resolution ratios, while unconstrained datasets such as MTA-ReID have rarely been explored. Therefore, we present our pyramid-resolution person restoration method for cross-resolution person ReID, which can fully exploit the valid information of the person images with various resolutions. Experimental results on both real-world cross-resolution datasets and simulated datasets show the superiority of our method. The architecture of the method is illustrated in Figure 1.

Most existing methods for cross-resolution person ReID predefine a suitable target resolution for person images with various resolutions and then use classical image restoration methods to restore the low resolution image to obtain a corresponding high resolution image. In complex application scenarios, however, we are often unable to determine a suitable resolution for restoration. For this reason, we have designed a pyramid-resolution restoration network based on the existing image restoration models. The original images can obtain the high-quality images with pyramid resolutions through this network.

Input an arbitrary image $x_{\rm ori} \in R^{H\times W\times C}$, where H, W, C represent its height, width, and number of channels, respectively. We first use a classical feature extractor to extract features from the original images. The feature extractor can be either convolutional neural network (CNN)-based [3] or transformer-based [4] models. This process can be defined as follows:

$$F_{\rm ori} = E(x_{\rm ori}), \tag{1}$$

where $x_{\rm ori}$ is the original image of the input, E refers to the feature extractor, $F_{\rm ori}$ is the feature of the original image obtained by the feature extractor.

After obtaining the feature, high-quality images with pyramid resolutions are obtained by our pyramid-resolution image reconstructor as designed.

$$X_{h1}, X_{h2}, X_{h3} = {\rm RC}(F_{\rm ori}), \tag{2}$$

where X_{h1}, X_{h2}, and X_{h3} represent the generated pyramid resolution images, and RC refers to our pyramid-resolution image reconstructor.

We can restore the original image through our pyramid-resolution restoration network to obtain three images with different resolutions. Then we use the OSNet [5] as the feature extraction network. In order to better fit the pyramid resolution situation, we train the feature extractors for each branch respectively. The obtained feature extractor can then be used for person re-identification. Finally, the corresponding cosine distance is calculated based on the extracted features of query images and features of gallery images.

In order to combine the information from pyramid resolution branches together, we use feature distance fusion to make full use of the image feature distances of these different branches, as well as to balance the information between the pyramid-resolution images and the original images. This fusion process is defined as follows:

$$D_{\rm final} = w_{\rm ori} D_{\rm ori} + \sum_j w_{hj} D_{hj}, \tag{3}$$

* Corresponding author (email: nnwang@xidian.edu.cn)

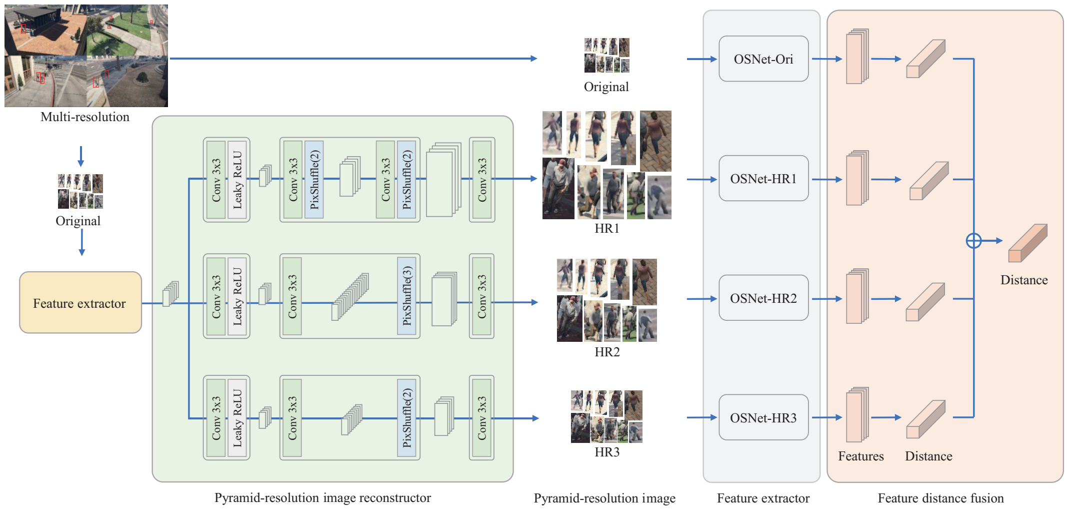

Figure 1 (Color online) Framework of the proposed pyramid-resolution person restoration method.

where D_{final} is the final feature distance for re-identification. D_{ori} is the feature distance of the original images, w_{ori} is its corresponding weight, and D_{hj} is the feature distance corresponding to branch j, where w_{hj} is its corresponding weight.

Experiments. For the image restoration procedure, we trained for 5000 iterations using the mean average error (MAE) loss. In the person ReID phase, we conducted the training process for 250 epochs, employing the Softmax loss and label smoothing. Comprehensive experimental results and analysis can be found in the supplementary material. Our method achieves supeorior performance on both real-world MTA-ReID dataset and a number of simulated datasets.

Conclusion. In this study, we present a simple yet effective pyramid-resolution person restoration method for cross-resolution person re-identification. Our method involves a pyramid resolution restoration network that enhances pyramid resolution images, and utilizes feature distance fusion to leverage valuable and complementary information from these pyramid images. Extensive experiments demonstrate the effectiveness of our method on both real-world cross-resolution datasets and simulated datasets.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62276198, U22A2035, U22A2096, 62306227), Key Research and Development Program of Shaanxi (Grant No. 2023-YBGY-231), Young Elite Scientists Sponsorship Program by CAST (Grant No. 2022QNRC001), Guangxi Natural Science Foundation Program (Grant No. 2021GXNSFDA075011), Natural Science Basic Research Plan in Shaanxi Province of China (Grant No. 2022JQ-696), Xi'an Science and Technology Plan Project (Grant No. 23GJSY0004), Open Research Project of Key Laboratory of Artificial Intelligence Ministry of Education (Grant No. AI202401), and Fundamental Research Funds for the Central Universities (Grant Nos. QTZX23083, QTZX23042, ZYTS24142).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Jing X Y, Zhu X, Wu F, et al. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. 695–704

2 Chen Y C, Li Y J, Du X, et al. Learning resolution-invariant deep representations for person re-identification. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019. 8215–8222

3 Niu B, Wen W, Ren W, et al. Single image super-resolution via a holistic attention network. In: Proceedings of the European Conference on Computer Vision, 2020. 191–207

4 Liang J, Cao J, Sun G, et al. SwinIR: image restoration using Swin transformer. In: Proceedings of the IEEE International Conference on Computer Vision, 2021. 1833–1844

5 Zhou K Y, Yang Y X, Cavallaro A, et al. Learning generalisable omni-scale representations for person re-identification. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 5056–5069