

Supplementary Materials for “Learning 3D Shape Feature for Texture-insensitive Person Re-identification”

Jiaxing Chen^{1,5#}, Xinyang Jiang^{3#}, Fudong Wang³, Jun Zhang³, Feng Zheng⁴, Xing Sun³, Wei-Shi Zheng^{1,2*}

¹School of Computer Science and Engineering, Sun Yat-sen University, China

²Peng Cheng Laboratory, Shenzhen, China

³Youtu Lab, Tencent

⁴CSE, Southern University of Science and Technology

⁵Pazhou Lab, Guangzhou, China

chenjx228@mail2.sysu.edu.cn, xinyangj9024@gmail.com, {fudongwang, bobbyjzhang}@tencent.com
zhengf@sustech.edu.cn, winfredsun@tencent.com, wszheng@ieee.org

Abstract

The supplementary materials of “Learning 3D Shape Feature for Texture-insensitive Person Re-identification” mainly consist of some further evaluations.

1. Evaluation on different 3D features.

We carry out experiments to further demonstrate the effectiveness of our 3D shape feature learning. We make comparison with: (1) the 3D reconstruction based features [2, 3]; (2) the 3D based ReID model OG-Net [5]. In Table A, the “HMR based feature” indicates 3D features learned by directly adding ReID losses to the original HMR model [2]. The “SPIN based feature F_1 ” is the output of *res4* in SPIN [3] without any ReID supervision. The “SPIN based feature F_2 +supervision” is the feature trained by using sub-networks (the *res4* block in Resnet[1]) and identity labels upon the output of *res3* in SPIN [3].

Although the architecture of our 3D branch is similar with the HMR model [2], some important modifications are made to fit the original 3D reconstruction architecture to the ReID feature learning. We decouple the pose and shape branches to thoroughly separate pose- and shape-related features, since the pose-related feature is an interference factor for person ReID and only shape feature is identity-specific. As shown in Table A, directly using HMR [2] as the 3D reconstruction regularizer achieves unsatisfactory performance, which validates that the separate modeling of pose and shape information is important for discriminative shape-related feature learning.

Equal Contribution.

* Corresponding Author.

Table A. Performance (%) of different 3D features.

Models	PRCC		Market1501	
	rank-1	rank-5	rank-1	mAP
Our 3D Shape feature	49.2	84.2	93.2	83.7
HMR [2] based feature	38.8	70.6	88.4	76.5
SPIN based feature F_1 [3]	5.3	14.9	1.8	0.6
SPIN based feature F_2 [3] + supervision	31.7	53.9	73.0	48.9
OG-Net [5]	20.4	45.8	85.9	66.9

We also make comparison with the state-of-the-art 3D reconstruction model SPIN [3]. The performances of “SPIN based feature F_1 ” and “SPIN based feature F_2 +supervision” are significantly lower than our model. This reveals that directly using the pretrained 3D reconstruction weights for ReID might be limited by the task gap and the poor discriminative ability. Consequently, our ReID oriented customized design on the 3D model is essential.

Besides, without the specialized mechanism for clothing texture-insensitive feature learning, the 3D based ReID model [5] achieves poor performance on the data covering clothing change situations.

2. Evaluation on the displacement.

Our motivation to introduce free-form displacements is that the low-dimension shape parameters of SMPL [4] could not represent human body shape intactly and we expect the estimation of displacement to help capture shape representation covering more identity-specific characteristics. To examine the effectiveness of displacement, we use the version without MGS for fair comparison, that are only global features to separately estimate shape parameters and global displacements, which is indicated as “with displacement” in Table B. The version “without displacement” indicates that only a single global feature to estimate shape parameters. However, we also do not want the 3D model

Table B. Performance (%) of model with/without displacements and regularization.

Models	PRCC	
	rank-1	rank-5
without displacement	44.5	82.7
with displacement and without regularization	47.6	84.5
with displacement and regularization	48.0	84.6

with displacements to be too detailed and cover clothing deformations, because only the shape of a “naked” body is required. We add a $L2$ regularization term on δ to keep the 3D model “smooth” in Equation 5 of the manuscript. The versions with/without the $L2$ term are denoted as “with/without regularization” in Table B.

In Table B, we could observe that the introduction of displacements could bring a significant improvement of 3.1% in rank-1, verifying the importance of displacements for capturing identity-specific features. The regularization on the displacement δ could further boost the ReID performance.

References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. [1](#)
- [2] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *CVPR*, pages 7122–7131, 2018. [1](#)
- [3] Nikos Kolotouros, Georgios Pavlakos, Michael J Black, and Kostas Daniilidis. Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In *ICCV*, pages 2252–2261, 2019. [1](#)
- [4] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015. [1](#)
- [5] Zhedong Zheng and Yi Yang. Parameter-efficient person re-identification in the 3d space. *arXiv preprint arXiv:2006.04569*, 2020. [1](#)