Online Domain Adaptation for Person Re-Identification with a Human in the Loop

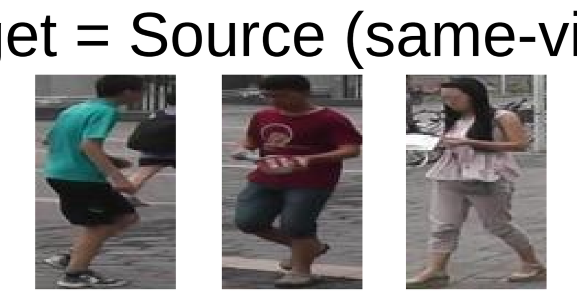
Rita Delussu, Lorenzo Putzu, Giorgio Fumera and Fabio Roli
Pattern Recognition and Applications Lab, University of Cagliari (Italy)
https://pralab.diee.unica.it
{rita.delussu, lorenzo.putzu, fumera, roli}@unica.it

University of Cagliari, Italy — Department of Electrical and Electronic Engineering

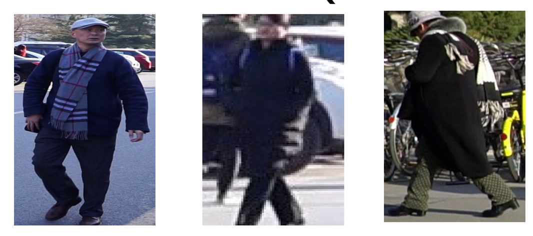
Cross-view person re-identification

Target camera

Target = Source (same-view)



Target ≠ Source (cross-view)



Source camera(s)

- Supervised approaches

- Domain Adaptation (DA)*
- Unsupervised DA (UDA)*
- Domain Generalisation

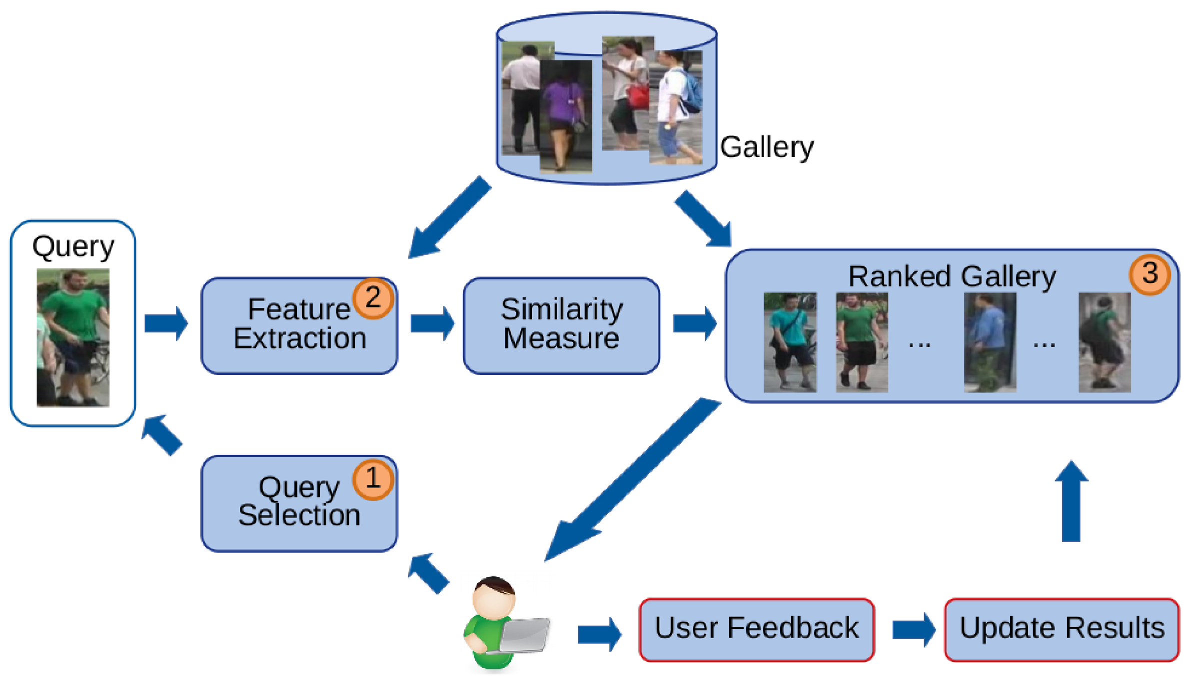
*requires images of the target camera for fine-tuning

Real application scenarios

- Target camera **is usually different** from source cameras
- Often target camera images are **not available** for training or fine-tuning

We focus on human-in-the-loop approach

Human-in-the-loop (HITL)



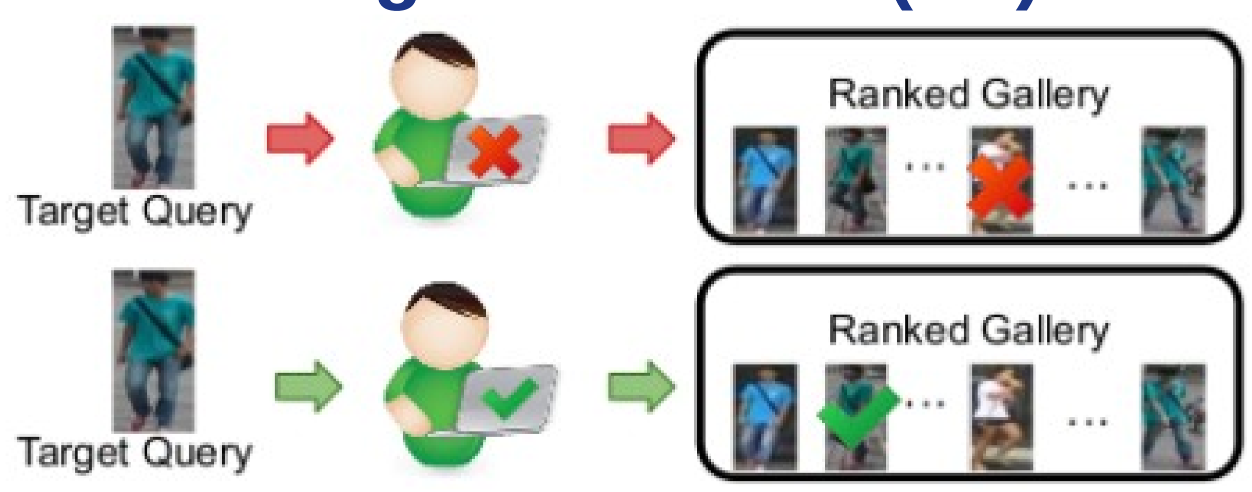
- HITL does not require images during training, but exploits operator's feedback during operation ⇨ can be seen as online DA
- Existing HITL methods for person re-identification are relatively complex
- The user interaction in HITL is similar to the one used in relavance feedback (RF) in context-based image retrieval
- RF algorithms have not been investigated in this task so far

Existing HITL person re-identification methods:
- Liu, Chunxiao, et al. "Pop: Person re-identification post-rank optimisation." Proceedings of the IEEE International Conference on Computer Vision. 2013.
- Wang, Hanxiao, et al. "Human-in-the-loop person re-identification." European conference on computer vision. Springer, Cham, 2016.

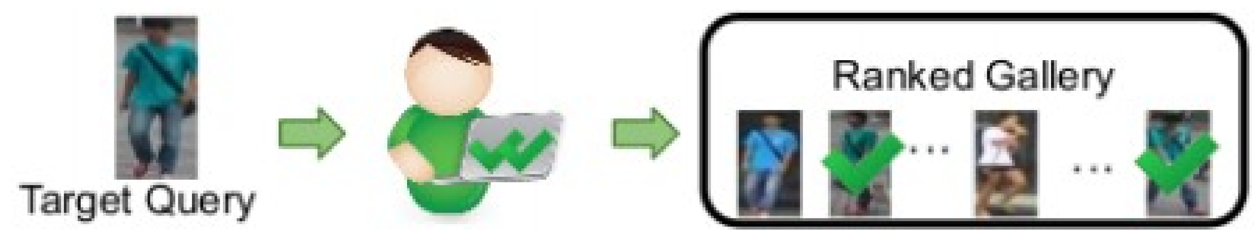
Feedback protocol

Existing HITL methods: single-feedback (SF)

The user selects a *single* image in the top-*k* ranks.
- ✔ a true match (if any)
- ✖ a strong negative



Protocol proposed in this work: multi-feedback (MF)



The user selects **all true matches** in top-*k* ranks. The other top-*k* gallery images are considered as negatives

Goal of this work

We investigate the effectiveness of RF algortithms by using the multi-feedback protocol

Experimental settings

- UDA methods: Mutual Mean Teaching (MMT)[1]
- RF methods: Query Shift (QS), Relevance Score (RS)
- SF vs MF protocol
- Cross-data set experiments (target cameras ≠ source)
- Data sets: Market-1501 (Market), DukeMTMC-reID (Duke)
- Query set: 300 identities
- RF methods: 3 feedback rounds
- User feedback on top-50 gallery images

Method	Market ⇨ Duke					Duke ⇨ Market				
	mAP	Rank-1	Rank-5	Rank-10	Rank-20	mAP	Rank-1	Rank-5	Rank-10	Rank-20
Source model	29.1	47.7	61.3	66.0	72.0	25.5	54.3	72.0	79.0	81.7
MMT	60.8	76.0	85.3	88.0	90.3	69.4	87.0	**95.3**	**97.0**	**97.7**
QS - SF	42.71	68.67	74.33	76.33	78.0	33.9	71.0	78.0	82.0	84.0
RS - SF	56.6	82.33	82.67	83.0	83.67	41.69	77.0	80.33	81.33	85.0
QS - MF	51.74	73.67	82.67	83.67	85.0	47.64	80.67	87.33	88.0	88.0
RS - MF	74.67	**92.0**	**92.67**	**92.67**	**93.0**	75.09	**92.67**	92.67	93.33	93.67

Conclusions

- The proposed MF protocol outperformes SF in both RF methods
- RS with the proposed MF protocol achieved better or similar performances than the UDA method MMT
- RF methods require user feedback on a much smaller amount of target data than the amount used by UDA methods

[1] Ge, Yixiao, Dapeng Chen, and Hongsheng Li. "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification." arXiv preprint arXiv:2001.01526 (2020).

25th International Conference on Pattern Recognition — 10 – 15 January 2021