RPIfield: A New Dataset for Temporally Evaluating Person Re-Identification*

Meng Zheng[1], Srikrishna Karanam[2], and Richard J. Radke[1]
[1]Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy NY
[2]Siemens Corporate Technology, Princeton NJ
zhengm3@rpi.edu, srikrishna.karanam@siemens.com, rjradke@ecse.rpi.edu

Abstract

The operational aspects of real-world human re-identification are typically oversimplified in academic research. Specifically, re-id algorithms are evaluated by matching probe images to candidates from a fixed gallery collected at the end of a video, ignoring the arrival time of each candidate. However, in real-world applications like crime prevention, a re-id system would likely operate in real time, and might be in continuous operation for several days. It would be natural to provide the user of such a system with instantaneous ranked lists from the current gallery candidates rather than waiting for a collective list after processing the whole video sequence. Re-id algorithms thus need to be evaluated based on their temporal performance on a dynamic gallery populated by an increasing number of candidates (some of whom may return several times over a long duration). This aspect of the problem is difficult to study with current benchmarking re-id datasets since they lack time-stamp information. In this paper, we introduce a new multi-shot re-id dataset, called RPIfield, which provides explicit time-stamp information for each candidate. The RPIfield dataset is comprised of 12 outdoor camera videos, with 112 known actors walking along pre-specified paths among about 4000 distractors. Each actor in RPIfield has multiple reappearances in one or more camera views, which allows the study of re-id algorithms in a more general context, especially with respect to temporal aspects.

1. Introduction

Person re-identification, or re-id, which has extensive applications in video surveillance and forensics, has attracted increasing attention in the past ten years [7]. It can be generally described as the task of re-identifying a person of interest, previously identified in a camera view, in another camera with a non-overlapping field of view. Previous re-id research has been primarily focused on key algorithmic aspects such as (a) feature extraction, i.e., the design of robust, invariant, and descriptive representation vectors for person images, (b) metric learning, i.e., learning distance metrics so that feature vectors belonging to the same person are closer than those belonging to different persons, and (c) end-to-end learning, where features and distance metrics are jointly learned [11, 13, 16]. We refer the reader to Karanam *et al.* [7] and Zheng *et al.* [16] for recent experimental and algorithmic surveys.

While these problems are all crucial for re-id algorithm development, design choices are typically made under the assumption that a collection of candidate images is accessible before we run the algorithm, which is unlikely to happen in a real-world re-id scenario. For a real-world system that operates in real time [12], it is suboptimal to provide the user with selected candidate list after finishing the collection of all candidate images. Instead, the user would like to check the output of the re-id system frequently during operation, and make an informative choice about which re-id algorithm can produce a more "stable" candidate list, in that correct matches to the probe are able to stay within a rank-k shortlist for longer times.

To study re-id algorithms in this way, we introduce a new multi-shot multi-camera re-id dataset in this paper, called *RPIfield*. It preserves time-stamp information for every detected person in each video sequence, in order to better simulate real-world re-id operational scenarios. With explicit time-stamp information of candidates' reappearances, re-id algorithms can then be applied to provide instantaneous rank lists for probes that simulate a real-time re-id system. Current multi-shot multi-camera re-id benchmarking datasets such as DukeMTMC4ReID [6], MARS [14], and Market1501 [15] all lack this kind of time-stamp information, which is crucial for real-world re-id system analysis. The RPIfield dataset is publicly available at this link: https://drive.google.com/file/d/1GO1zm7vCAJwXgJtoFyUs367_Knz8Ev0A/view?usp=sharing.

*This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2013-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

2. The RPIfield Dataset

We placed 12 synchronized 1440×1080 HD surveillance cameras on poles around an outdoor field on the Rensselaer Polytechnic Institute campus. A snapshot of the camera network layout is shown in Figure 1. As can be observed from the figure, 6 poles are positioned around the field, one each at points A through F, with red arrows representing a camera and its viewing direction.

Before recording, we predetermined 112 unique paths around the points A to F, each containing at least 3 different points to ensure multiple reappearances of the same identity at different camera views. For example, if a participant is walking along the path $A \rightarrow B \rightarrow C \rightarrow F \rightarrow B \rightarrow A$, s/he will appear in camera views 2, 3, 5, 6, 7, 11, 4, 3 and 2, respectively. In order to simulate a mass-transit type re-id environment (i.e., a relatively large portion of distractors), the data collection time was arranged at noon on a weekday, when a large amount of unknown pedestrians were walking across the field between classes.

To simulate a fully automatic re-id system, we used the off-the-shelf ACF [3, 4] person detector to automatically generate person images. We then used the Intersection over Union (IoU) measure to gather all cropped images of the same identity. Specifically, for each cropped person image, we calculated the IoUs of the current rectangle with all person images detected in the past 90 frames, and classified the current rectangle as the category with the largest IoU value. For a bounding box with no surviving IoUs (we use a threshold of 0.4), we assign a new label. We manually corrected errors due to false alarms and broken tracklets (e.g., multiple tracklets of the same person) to have one image sequence corresponding to each appearance of a person. The corresponding frame number is used to preserve the time-stamp information for each image.

A summary of the statistics for RPIfield is given in Table 1. Each column from left to right gives the number of cropped bounding boxes (#BBx), unique participants (#Id), reappearances of participants (#Reapp), unique distractors (#Ped), image sequences (#Seq) and video length in minutes (Len) for each camera video. Note that each image sequence given in column (#Seq) belongs to one unique person. If the same person reappears in the same camera view at different times, we treat each reappearance as a separate sequence, but assign it a single identity label in order to ensure the uniqueness of the identities included in our dataset.

We compare several existing benchmarking datasets with RPIfield in Table 2, in which we summarize the attributes (the right column) of each dataset as multi-shot (MS), or multi-camera (MC). From the table, we can see that RPIfield is constructed from the most number of cameras in the camera network. Though the number of identities (112) is relatively small compared to DukeMTMC4ReID [6], Market-1501 [15], and MARS [14], it is important to note that the number "112" in our dataset corresponds to **known actors**, who were provided specific instructions with regard to the walking path and re-appearances. To our knowledge, no other benchmarking dataset was constructed in this way. While most datasets shown in Table 2 have no distractors except DukeMTMC4ReID [6], Market-1501 [15], and MARS [14], RPIfield preserves all image sequences for all detected pedestrians. Most importantly, actors could reappear in one or multiple cameras multiple times with explicit time-stamp information in RPIfield, which distinguishes it from other benchmarking datasets.

3. Discussion and Future Work

As we mentioned above, in real-world re-id applications, we need to consider a fully automated re-id system that automatically detects and generates candidate images. Re-id algorithms applied to such a system need to be evaluated temporally, such that the gallery size consistently increases with more candidates showing up in the video over time. With the explicit time-stamp information provided in RPIfield, we now are able to consider the performance of re-id algorithms on a simulated real-time re-id system, acquiring the instantaneous rank of correct matches to the probes over time during system operation, rather than matching to a fixed gallery with all candidates collected at the end of the video.

Specifically, consider a time-evolving gallery consistently populated with new candidates showing up in the video sequence. For a person of interest who reappears with rank k calculated by a re-id algorithm (in the gallery of candidates collected before its reappearance) at a certain time T, the instantaneous rank of the probe could only increase after T, since the gallery is continually filled with new candidates. An interesting question to the user is: how long can a correct match persist at rank-k over time? For the overall re-id system, the user also would like to know the percentage of the probes whose rank persists at rank-k, to get a sense of the performance of such a real-world re-id system and help them make informative choices about sub-system design. These issues are further discussed in Karanam *et al.* [8].

Table 1: Statistics of the RPIfield dataset.

Cam No.	#BBx	#Id	#Reapp	#Ped	#Seq	Len (m)
Cam 1	59,230	78	107	297	485	152
Cam 2	112,523	83	94	653	830	152
Cam 3	72,005	74	68	822	964	156
Cam 4	53,986	70	72	393	536	157
Cam 5	67,672	63	52	781	896	158
Cam 6	56,472	64	38	865	967	156
Cam 7	17,809	48	36	93	177	155
Cam 8	36,338	55	40	266	361	149
Cam 9	3,910	40	16	32	88	149
Cam 10	10,492	62	51	105	218	151
Cam 11	73,601	76	149	448	673	145
Cam 12	37,543	70	79	233	382	146
Total	601,581	112	802	3,996	6,577	1,826

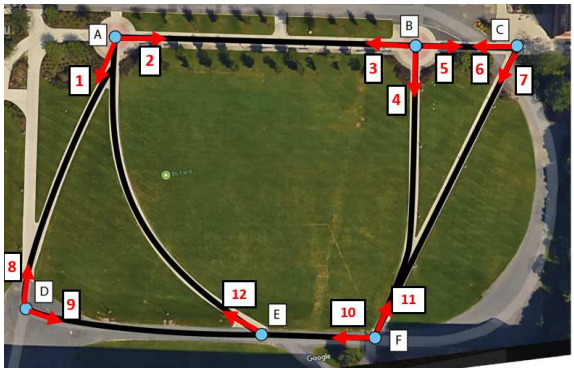

Figure 1: Overhead view of the RPIfield camera locations and orientations, superimposed on a map of the RPI '86 Field.

Table 2: Comparison of existing multi-shot, multi-camera re-id datasets.

Dataset	#BBx	#Id	#Ped	#Cam	Attributes
RPIfield	**601,581**	**112**	**3,996**	**12**	**MS,MC**
DukeMTMC4ReID [6]	4,6261	1,852	21,551	8	MS,MC
Market-1501 [15]	2,668	1,501	2,793	6	MS,MC
MARS [14]	1,067,516	1,261	3,248	6	MS,MC
SAVIT-Softbio [2]	64,472	152	0	8	MS,MC
3DPeS [1]	1,011	192	0	8	MC
CUHK02 [9]	7,264	1816	0	10	MC
CUHK03 [10]	13,164	1360	0	10	MC
HDA+ [5]	2,976	74	0	12	MC

References

[1] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *Proc. of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding*. 3

[2] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey. A database for person re-identification in multi-camera surveillance networks. In *DICTA*, 2012. 3

[3] P. Dollar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 2014. 2

[4] P. Dollar, S. Belongie, and P. Perona. The fastest pedestrian detector in the west. In *BMVC*, 2010. 2

[5] D. Figueira, M. Taiana, A. Nambiar, J. Nascimento, and A. Bernardino. The HDA+ data set for research on fully automated re-identification systems. In *ECCV*, 2014. 3

[6] M. Gou, S. Karanam, W. Liu, O. Camps, and R. J. Radke. DukeMTMC4ReID: A large-scale multi-camera person re-identification dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 1, 2, 3

[7] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke. A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 1

[8] S. Karanam, E. Lam, and R. J. Radke. Rank persistence: Assessing the temporal performance of real-world person re-identification. In *ICDSC*, 2017. 2

[9] W. Li and X. Wang. Locally aligned feature transforms across views. In *CVPR*, 2013. 3

[10] W. Li, R. Zhao, T. Xiao, and X. Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 3

[11] W. Li, X. Zhu, and S. Gong. Person re-identification by deep joint learning of multi-loss classification. *arXiv preprint arXiv:1705.04724*, 2017. 1

[12] O. Camps, M. Gou, T. Hebble, S. Karanam, O. Lehmann, Y. Li, R. J. Radke, Z. Wu, F. Xiong. From the lab to the real world: Re-identification in an airport camera network. *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 27, Issue: 3, 2017. 1

[13] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Deep metric learning for person re-identification. In *Proc. of the 2014 22Nd International Conference on Pattern Recognition*, 2014. 1

[14] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, 2016. 1, 2, 3

[15] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 1, 2, 3

[16] L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: Past, present and future. *ArXiv:1610.02984*, 2016. 1