Rank Persistence: Assessing the Temporal Performance of Real-World Person Re-Identification*

Srikrishna Karanam
Siemens Corporate Technology
Princeton, NJ
srikrishna.karanam@siemens.com

Eric Lam
Rensselaer Polytechnic Institute
Troy, NY
eric@ericblam.com

Richard J. Radke
Rensselaer Polytechnic Institute
Troy, NY
rjradke@ecse.rpi.edu

ABSTRACT

Designing useful person re-identification systems for real-world applications requires attention to operational aspects not typically considered in academic research. Here, we focus on the temporal aspect of re-identification; that is, instead of finding a match to a probe person of interest in a fixed candidate gallery, we consider the more realistic scenario in which the gallery is continuously populated by new candidates over a long time period. A key question of interest for an operator of such a system is: how long is a correct match to a probe likely to remain in a rank-k shortlist of candidates? We propose to distill this information into a Rank Persistence Curve (RPC), which allows different algorithms' temporal performance to be directly compared. We present examples to illustrate the RPC using a new dataset with multiple candidate reappearances, and discuss considerations for future research that explicitly involves temporal aspects.

CCS CONCEPTS

• **Computing methodologies → Temporal reasoning**; **Object identification**; **Visual content-based indexing and retrieval**;

1 INTRODUCTION

Research in the area of automatic human re-identification, or re-id, has exploded in the past ten years. The re-id problem is generally stated as: given an image of a person of interest as seen in a "probe" camera view, how can we find the same person among a set of candidate people seen in a "gallery" camera view? Re-id research to date typically falls into one or more of the following categories:

- Appearance modeling, in which the goal is to design or learn a feature representation for re-id candidates that is invariant to challenges like viewpoint and illumination variation (e.g., [1, 12, 15, 17])
- Metric learning, in which the goal is to learn, in a supervised fashion, a distance metric that is used to search for the person of interest in the gallery set (e.g., [10, 16, 21, 22]).
- Multi-shot re-id, in which both the probe and gallery candidates are represented as short image sequences instead of single frames (e.g., [8, 9, 11, 20]).

While critical to the success of a real-world deployment of a re-identification algorithm, research in these areas generally over-simplifies the problem that would face a real-world user of a re-id system. In particular, the *temporal* aspect of the re-id problem is totally ignored in most academic re-id research. That is, in the real world, candidates would be constantly added to the gallery as new subjects are automatically tracked, as opposed to presented to an algorithm all at once. Even if a correct match to the probe appears in a rank-ordered shortlist shortly after they appear in a gallery camera, this isn't helpful to a user if the candidate is immediately "pushed off" the list after a few minutes by a new wave of incoming candidates. To the user, a natural question is, how long can a correct match be expected to stay in the shortlist under typical circumstances? Proposed re-id algorithms should be judged based on this notion of *persistence* in time, not just raw batch performance as presented in a cumulative match characteristic (CMC) curve. Figure 1 illustrates the problem; an incoming correct match arrives at a low rank upon its first appearance, but is quickly pushed down the list as more candidates arrive. In this example, the correct candidate would have only stayed on a shortlist of the top 10 candidates for 10 minutes (about the same duration as a typical coffee break!).

In this paper, we explore several temporal aspects of re-id and propose new evaluation methodologies that allow different re-id algorithms to be compared based on the concept of rank persistence. We discuss strategies for evaluating algorithms in circumstances when the same person of interest can appear multiple times in the gallery (which could occur if the gallery spans several hours or days) as well as when the performance on multiple persons of interest should be aggregated. The key concept we propose is called the *Rank Persistence Curve (RPC)*, which quantifies the percentage of candidates that remain at a certain rank for a given duration. RPCs for different algorithms can be directly compared on the same dataset to allow a user to make informed choices about expected performance in real-world deployments.
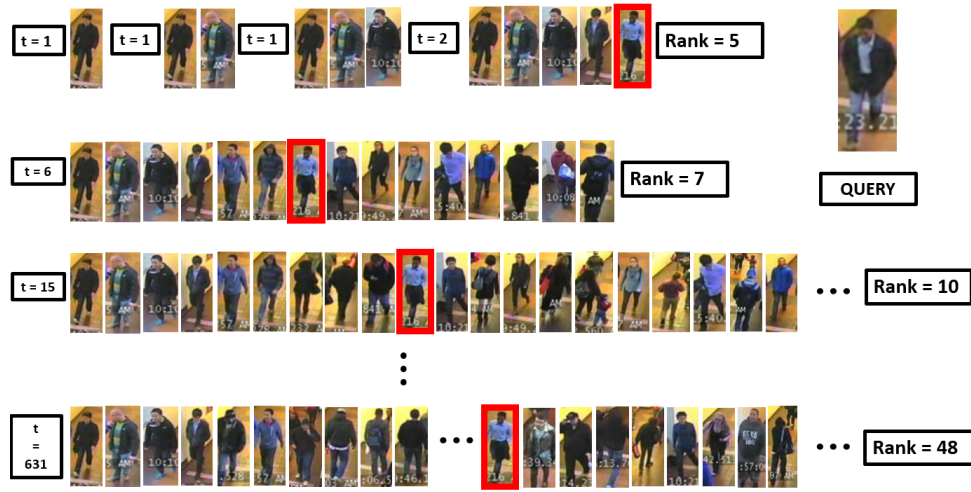
Figure 1: An example showing the temporal evolution of a gallery set captured in a real surveillance system, and the corresponding time-varying rank for a given query achieved by a re-id algorithm. The candidate arrives at $t = 2$ at rank 5, and drops in rank over time as more candidates arrive in the target camera.

We illustrate our proposed methodology with examples and experiments drawn from a dataset we collected over many hours at a light rail station, which contains multiple time-stamped reappearances of multiple actors. We conclude the paper with further discussion about the temporal implications of re-id, and suggestions for future research in this area.

2 PRELIMINARIES

Although the focus of this paper is not the proposal of a new re-id dataset or algorithm, we need instances of each to illustrate our proposed methodology. In particular, since existing re-id benchmarking datasets lack temporal annotation, we collected a new dataset more suitable for this study.

The dataset is composed of 10 hours of surveillance video data from a wall-mounted camera located in an indoor light rail station in the United States. We had 7 known actors participate in the data collection activity. Each actor appeared once in a morning session, and subsequently re-appeared in the same camera view in multiple afternoon and evening sessions. In each re-appearance, the actor wore slightly different clothes or accessories than in the previous session, to make the re-id problem more realistic and challenging. We used each actor's appearance in the morning session as the "probe" for subsequent re-id queries. In total, we ground-truthed 3 known re-appearances of each actor, each randomly spread across the 10-hour duration.

To generate candidate tracks for the gallery (i.e., for both the actors and the distractor pedestrians that entered the camera view over the 10-hour video), we used an off-the-shelf person detector, based on the aggregated channel features (ACF) algorithm of Dollar *et al.*[3, 4], to crop out person images. Sample images from the different appearances of the actors in our dataset are shown in Figure 2. A statistical summary of our dataset is provided in Table 1.

Table 1: Statistical summary of our dataset

Property	Total/Type
Bounding Boxes	4,639
Known Actors	7
Known Actor Re-Appearances	3
Candidates (incl. actors)	535
Video Hours	10
Detection algorithm	ACF [3, 4]

As a baseline re-id algorithm, we used the recently proposed Gaussian of Gaussian (GOG) descriptor [14] to rank gallery candidates. GOG is an unsupervised algorithm that constructs an appearance model given image data and uses Euclidean distance to rank candidates. The baseline algorithm is multi-shot in that it uses multiple frames of each candidate's appearance to form feature vectors to be compared, instead of a single rectangle.

Specifically, given a track of images for the probe appearance of the person of interest and each gallery candidate, we extract features for each image using GOG. Let $\mathbf{x}_p^k$, $k = 1, \ldots, n$ and $\mathbf{x}_i^k$, $k = 1, \ldots, m$ denote the n feature vectors of the probe and the m feature vectors of the ith gallery candidate, respectively. We then determine the appearance model for the probe, $\mathbf{f}_p$, and the gallery candidate, $\mathbf{f}_i$, as the mean feature vector of the available feature vectors. Specifically,

$$\mathbf{f}_p = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_p^k$$

$$\mathbf{f}_i = \frac{1}{m} \sum_{k=1}^{m} \mathbf{x}_i^k$$

Figure 2: Examples of actors and their appearances in our dataset.

The similarity score s_i, used to rank gallery candidates, is computed using the Euclidean distance as

$$s_i = \|\mathbf{f}_p - \mathbf{f}_i\|$$

3 RANK PERSISTENCE

In this section, we present the concept of rank persistence in steps, working up to evaluating situations in which multiple persons of interest each appear multiple times over the course of a long video sequence.

3.1 One probe, one reappearance

We first consider the case of a single probe/query that has exactly one reappearance over the course of an entire video. In this case, we can easily graph their rank over time with respect to an ever-increasing gallery, as illustrated in Figure 3.

The horizontal axis is real time (i.e., minutes from the beginning of our 10-hour test sequence). The vertical axis shows the rank of the true reappearance in the gallery over time. Clearly, the rank can only decrease as more candidates arrive in the gallery. Also, the graph does not start at $t = 0$ but indicates the time at which the reappearance occurs. These per-probe curves will form the basis for the aggregate Rank Persistence Curve discussed in Section 3.3.

Figure 4 illustrates three rank curves for probe/gallery pairs of increasing difficulty. In the first example, the gallery reappearance is very similar to the probe, and also dissimilar to other candidates seen over the course of the video, so the rank is fairly constant at a low value. In the last example, the gallery reappearance is quite unlike the probe, so the rank increases quickly as new candidates appear.

3.2 One probe, multiple reappearances

In a real-world scenario, when we are searching for a person of interest in streaming video over many hours or even days, it is possible that s/he may re-appear multiple times at different ranks.
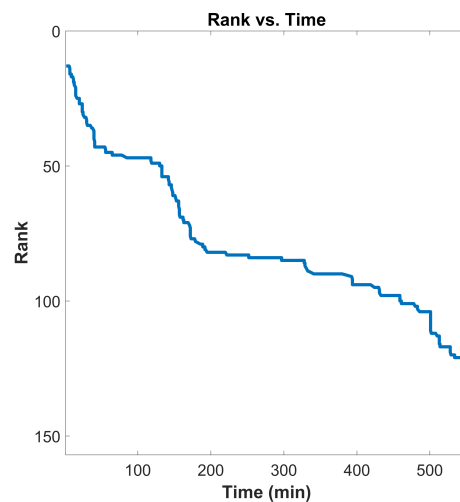

Figure 3: An example of how the rank r of a person of interest changes in a temporally-varying gallery set. This candidate's first re-appearance is at $t = 2$ minutes.

Therefore, we modify the temporal rank curve from the previous section as illustrated in Figure 5. That is, the vertical axis shows the highest instantaneous rank held by *any* reappearance of the probe.

In this example, the first reappearance of the candidate in the gallery looks rather unlike the probe, so the initial rank is relatively high and increases sharply with time. However, in the next reappearance, the gallery image looks very much like the probe; the rank of this new candidate is low, and the candidate persists at this rank for the rest of the long video. A third reappearance of the candidate occurs later in the video, but is not as similar as the second reappearance, so does not affect the temporal rank plot.
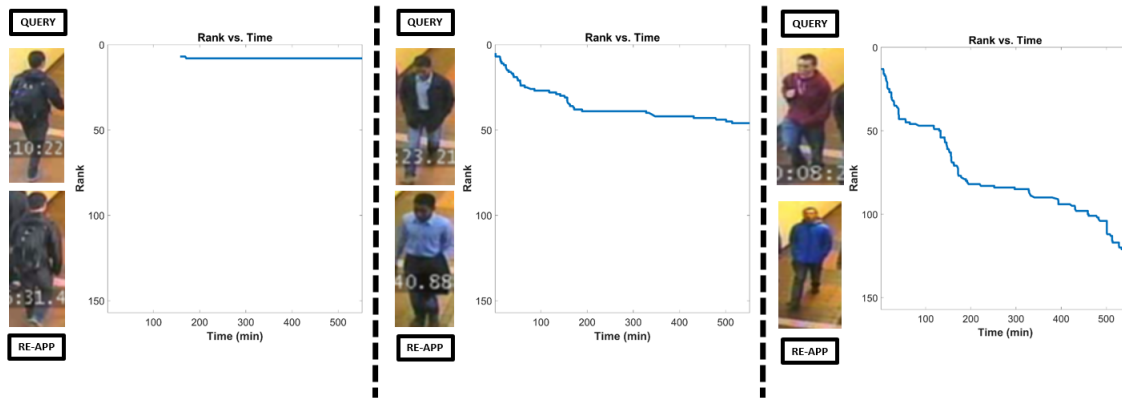
Figure 4: Temporal rank curves for three sample queries. In each case, the image on the top is the probe/query and the image on the bottom is the known reappearance in the gallery. The graph to the right of these images shows the temporal evolution of the rank of the known reappearance.
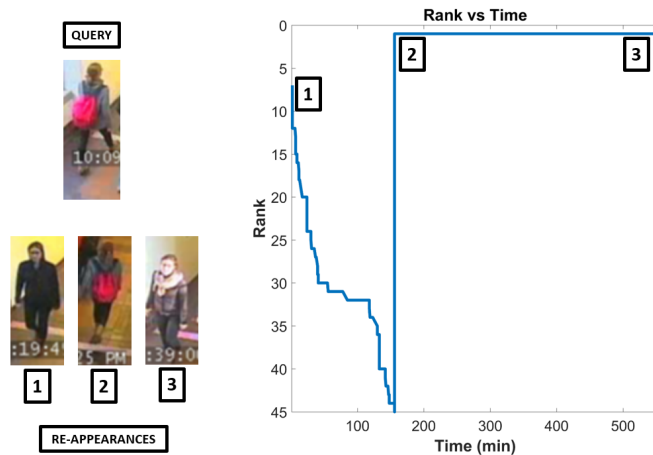

Figure 5: Temporal rank plot for a query with multiple re-appearances.

3.3 Multiple probes, multiple reappearances

Finally, we arrive at the most general situation, in which we characterize the performance of a given re-id algorithm across multiple probes, each of which may have multiple reappearances in the gallery. We define the *Rank Persistence Curve (RPC)* to evaluate performance as follows. First, we fix a specific rank r. For each duration d (in real units) on the horizontal axis, we plot the percentage of candidates that appear continuously in the top-r list for at least d units. Thus, the RPC is monotonically decreasing, and RPCs at higher ranks dominate those at lower ranks.

Figure 6a illustrates RPCs for our dataset/baseline re-id algorithm combination. In contrast, a traditional Cumulative Match Characteristic (CMC) curve, also for our dataset/baseline re-id algorithm combination (averaging all available feature vectors for every candidate), is shown in Figure 6b. We can see that the two types of curves are qualitatively different. Since we want to capture the temporal aspect of rank in the RPC, the dependent axis is no longer rank but duration, and we need a third "axis" (in this case color) to indicate rank.

Let's focus on the RPC for $r = 1$, shown in red. This captures our objective of visualizing how likely and how long a candidate is to stay at rank 1 across a long video sequence. Clearly, this is a stringent requirement, and we can see that only 1 of the 7 probes ever had a reappearance at rank 1 at all. If we consider the RPC for $r = 5$, we see that more candidates are likely to persist at rank 5 for longer amounts of time. RPCs can help operators of re-id systems answer questions like: can my re-id algorithm be expected to preserve correct matches in the top-10 shortlist for at least 15 minutes at least 90% of the time? These considerations are important both in terms of the length of the shortlist (real-world end users, typically not computer vision experts, would not want to scroll through pages and pages of candidates to find the person of interest) and the duration of persistence (in real-world scenarios, end users may only get around to checking the output of a re-id surveillance system a few times an hour).
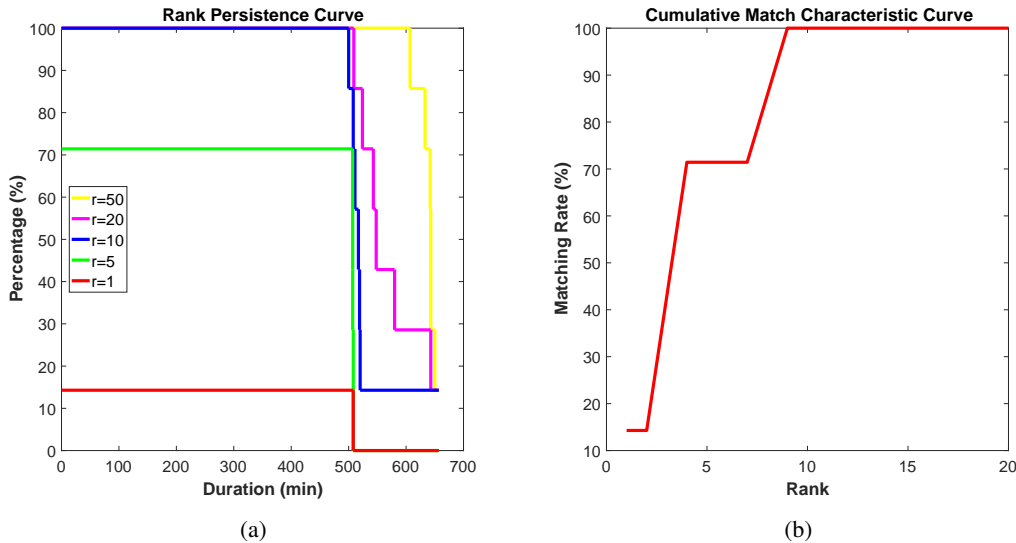
Figure 6: (a) The proposed Rank Persistence Curve (RPC) at different ranks for our dataset/baseline re-id algorithm combination. (b) The traditional Cumulative Match Characteristic (CMC) curve for our dataset/baseline re-id algorithm combination.

Thus, RPCs can be used to quickly visualize the performance of competing re-id systems applied in the same environment. For example, Figure 7 shows the RPCs for our baseline algorithm using GOG vs. the ColorTexture algorithm proposed by Gray and Tao [6], which uses an ensemble of localized color and texture features. We can see that the GOG algorithm dominates the older algorithm with respect to temporal persistence, and that the RPCs capture the essence of both algorithms' performance in an easy-to-read graph.

3.4 Video Flow Density

We conclude this section with a brief discussion on video flow density, a topic we have not considered in this paper but has a direct impact on the behavior of the rank of the person of interest. We define video flow density as the instantaneous number of people seen by the gallery camera per unit time. To understand how this might impact rank persistence, consider two nonoverlapping time blocks with the same duration. In the first block, the gallery camera sees x people walking by. In the second block, the gallery camera sees y people walking by. If $x \gg y$, we add many more people in the gallery in the first time block, likely resulting in a steeper increase in a candidate's re-id rank compared to the second time block. This suggests we should pay close attention to the rate at which new people in the video are added to the gallery.

4 DISCUSSION AND FUTURE WORK

Now that the underlying computer vision and machine learning technologies for re-id have matured [7, 25], we contend that researchers should begin to take a broader view of evaluating how re-id algorithms should integrate into functional real-world systems. For example, as noted in Camps et al. [2], the problem of comparing one candidate rectangle of pixels to another is only a small part of a fully automated re-id system. Instead, we must take into account that the candidate rectangles are likely generated by an automatic
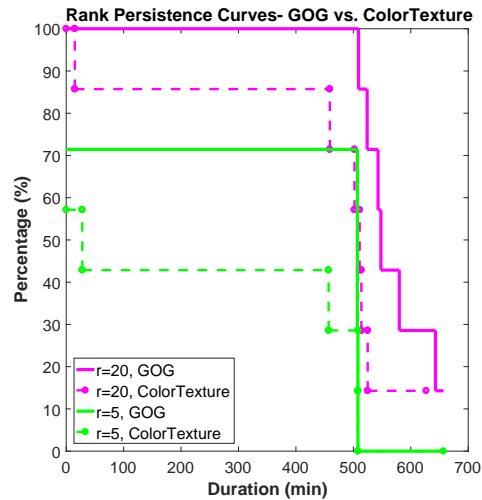

Figure 7: RPCs can be used to easily compare the temporal performance of competing re-id algorithms. In this example, the GOG algorithm's RPC (solid curve) dominates the ColorTexture's algorithm's RPC (dashed curve) at $r = 5$ and $r = 20$, demonstrating the superiority of GOG from a temporal persistence perspective.

(and possibly inaccurate) human detection and tracking subsystem, that the overall system needs to operate in real time, and that the system may be in operation for very long periods of time. Instead of benchmarking datasets in which the gallery images are acquired only a few moments after the probe images, we must consider crime prevention applications in which a perpetrator may return to the scene of the crime days after their initial detection. In such cases,

the gallery of candidates is ever-expanding, and for long periods of time may not contain the person of interest at all.

These real-world considerations present several challenges to the re-id research community, several of which we discuss below.

- While they are derived from real-world surveillance video, current re-id benchmarking datasets such as VIPeR [6], iLIDS-VID [19], and MARS [23] lack time stamps for the gallery sets, rendering them inappropriate for the type of temporal research discussed here. While it might be possible to repurpose them for temporal re-id research, adding artificial time stamps seems suboptimal. It would be better to generate a new temporal re-id dataset for this purpose, ideally spanning several days and thousands of tracked candidates. For validation, multiple reappearances of a substantial number of actors would need to be included. The dataset discussed in this paper is a step in this direction, but has too few unique actors to be broadly useful.

- The concept of "splits" is critical for fair comparisons of re-id algorithms using current benchmarking datasets. That is, the provider of the dataset typically specifies a random subset of the data to be used for training and another to be used for testing. A similar concept would need to be developed for temporal re-id datasets, i.e., not only specifying which candidates should be used for training/testing, but also which temporal spans should be used in the time-evolving galleries.

- As discussed in Section 3.4, the x-axis of the proposed Rank Persistence Curves is critically coupled to the video flow density of candidates in the underlying camera. Operationally, it would be more useful to characterize rank persistence in real units of minutes or hours; however, this ties the RPC strongly to a particular dataset and possibly a time of day. For example, different rail stations may have different levels of traffic, and the traffic would vary during rush hour vs. off-peak times. Temporal re-id benchmarking protocols would need to distinguish subsets of data with respect to video flow rate.

- The temporal dimension, in terms of both the ever-increasing gallery size and multiple probe appearances, discussed here leads to natural challenges from a feature and metric learning point of view. These can potentially be addressed by using temporally incremental approaches to learning re-id models [13, 18] where the model can be temporally adapted over time using either automated or human-in-the-loop feedback.

- In our experience with integrating academic re-id algorithms into operational surveillance command centers, we found the issue of user interfaces to be extremely important. The similarity between a rank-k shortlist and a police lineup was an effective analogy. However, the potentially very long time scales for crime prevention applications requires the re-evaluation of an operationally meaningful shortlist. Should candidates "age out" of the ranked list using some sort of forgetting factor or re-ranking scheme [5, 24]? Should extremely promising candidates from long ago be kept alongside less-certain but more timely recent candidates? Should the time-varying gallery contain all candidates ever seen or only those from the last N minutes? These considerations require close consultation

with the potential users of the system to understand and set expectations and corresponding interface choices.

REFERENCES

[1] Loris Bazzani, Marco Cristani, and Vittorio Murino. 2013. Symmetry-driven accumulation of local features for human characterization and re-identification. *Comput. Vision and Image Understanding (CVIU)* 117, 2 (2013), 130–144.

[2] O. Camps, M. Gou, T. Hebble, S. Karanam, O. Lehmann, Y. Li, R. Radke, Z. Wu, and F. Xiong. 2017. From the Lab to the Real World: Re-Identification in an Airport Camera Network. *IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)* 27, 3 (2017), 540–553.

[3] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. 2014. Fast feature pyramids for object detection. *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)* 36, 8 (2014), 1532–1545.

[4] Piotr Dollár, Serge J Belongie, and Pietro Perona. 2010. The Fastest Pedestrian Detector in the West.. In *Proc. Brit. Mach. Vision Conf. (BMVC)*.

[5] Jorge Garcia, Niki Martinel, Christian Micheloni, and Alfredo Gardel. 2015. Person re-identification ranking optimisation by discriminant context information analysis. In *IEEE Int. Conf. Comput. Vision (ICCV)*.

[6] Douglas Gray and Hai Tao. 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Eur. Conf. Comput. Vision (ECCV)*.

[7] Srikrishna Karanam, Mengran Gou, Ziyan Wu, Angels Rates-Borras, Octavia I. Camps, and Richard J. Radke. 2016. A Comprehensive Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets. *arXiv preprint arXiv:1605.09653* (2016).

[8] Srikrishna Karanam, Yang Li, and Richard J Radke. 2017. Person re-identification with block sparse recovery. *Image and Vision Computing* 60 (2017), 75–90.

[9] Yang Li, Ziyan Wu, Srikrishna Karanam, and Richard J Radke. 2015. Multi-Shot Human Re-Identification Using Adaptive Fisher Discriminant Analysis.. In *Proc. Brit. Mach. Vision Conf. (BMVC)*.

[10] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. 2015. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. In *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*.

[11] Giuseppe Lisanti, Iacopo Masi, Andrew D Bagdanov, and Alberto Del Bimbo. 2015. Person re-identification by iterative re-weighted sparse ranking. *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)* 37, 8 (2015), 1629–1642.

[12] Bingpeng Ma, Yu Su, and Frédéric Jurie. 2012. Local descriptors encoded by Fisher vectors for person re-identification. In *ECCV Workshops*.

[13] Niki Martinel, Abir Das, Christian MIcheloni, and Amit K. Roy-Chowdhury. 2016. Temporal Model Adaptation for Person Re-Identification. In *Eur. Conf. Comput. Vision (ECCV)*.

[14] Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato. 2016. Hierarchical Gaussian descriptor for person re-identification. In *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*.

[15] N McLaughlin, J Martinez del Rincon, and P Miller. 2016. Recurrent Convolutional Network for Video-based Person Re-Identification. In *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*.

[16] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. 2010. Person Re-Identification by Support Vector Ranking.. In *Proc. Brit. Mach. Vision Conf. (BMVC)*.

[17] Riccardo Satta. 2013. Appearance descriptors for person re-identification: a comprehensive review. *arXiv preprint arXiv:1307.5748* (2013).

[18] Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang. 2016. Human-In-The-Loop Person Re-Identification. In *Eur. Conf. Comput. Vision (ECCV)*.

[19] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2014. Person Re-identification by Video Ranking. In *Eur. Conf. Comput. Vision (ECCV)*.

[20] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2016. Person re-identification by discriminative selection in video ranking. *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)* 38, 12 (2016), 2501–2514.

[21] Kilian Q Weinberger and Lawrence K Saul. 2009. Distance metric learning for large margin nearest neighbor classification. 10 (2009), 207–244.

[22] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. 2014. Person re-identification using kernel-based metric learning methods. In *Eur. Conf. Comput. Vision (ECCV)*.

[23] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. 2016. MARS: A Video Benchmark for Large-Scale Person Re-identification. In *Eur. Conf. Comput. Vision (ECCV)*.

[24] Liang Zheng, Shengjin Wang, Lu Tian, Fei He, Ziqiong Liu, and Qi Tian. 2015. Query-adaptive late fusion for image search and person re-identification. In *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*.

[25] Liang Zheng, Yi Yang, and Alexander G Hauptmann. 2016. Person Re-identification: Past, Present and Future. *arXiv preprint arXiv:1610.02984* (2016).