

A Verify-Correct Approach to Person Re-identification Based on Partial Least Squares Signatures

Gabriel Lorencetti Prado Helio Pedrini
Institute of Computing, University of Campinas
Campinas-SP, Brazil, 13083-852

William Robson Schwartz
Department of Computer Science, Universidade Federal de Minas Gerais
Belo Horizonte-MG, Brazil, 31270-901

Abstract

In the surveillance field, it is very common to have camera networks covering large crowded areas. Not rarely, cameras in these networks do not share the same field of view and they are not always calibrated. In these cases, common problems such as tracking cannot be directly applied as the information from one camera must be also consistent with the others. This is the most common scenario for the person re-identification problem, where there is the need to detect, track and keep a consistent identification of people across a network of cameras. Many approaches have been developed to solve this problem in different manners. However, person re-identification is still an open problem due to many challenges required to be addressed to build a robust system. To tackle the re-identification problem and improve the accuracy, we propose a novel approach based on Partial Least Squares signatures, which is based on the visual appearance of people. We demonstrate the method performance with experiments conducted on three public available data sets. Results show that our method overcome the chosen baseline on all data sets.

1. Introduction

Person re-identification is a problem of interest in video surveillance. It emerges from some classic problems, such as person detection, tracking, identification and recognition, being described as the task of assigning the same identifier to all samples of the same person throughout a video sequence across different cameras. This definition leads to many challenges to efficiently accomplish this task. For instance, considering only a single camera, the most common issues are occlusion, pose variation and lighting conditions. When more cameras are considered, the main difficulty be-

comes the task of keeping track of a person across the whole camera network.

Another common problem faced in re-identification is due to the large number of people that might be in the scene. The similarity among people leads to errors when labeling each people with only one of the available options. This problem is often treated by changing the evaluation metric. Instead of assigning only one label to a person, many labels are given with different degrees of certainty. This approach transforms the re-identification problem into a ranking one [9].

In this paper, we propose a different approach to solving the re-identification problem. We do not make a strong assumption regarding the camera setting, so that our method can be easily applied on disjoint non-calibrated cameras. A verify-correct schema based on Partial Least Squares (PLS) [27] signatures is proposed to evaluate re-identification results after the final assignment and correct people that were misclassified. The PLS signatures model the behavior of each person against some generic patterns. Samples that do not fit the person model are considered as errors and are, therefore, assigned to another person that fits them better.

We compare our approach with the method proposed in [23] on three challenging video data sets: CAT, PETS2006 and Techgate and experimental results demonstrate that the proposed technique is able to further improve the accuracy of the baseline method for all data sets considering all evaluated metrics.

2. Related Work

Person re-identification has been extensively addressed by several methods in the literature. Due to its novelty, different views of the problem are given, with different solutions for each of them. Some of the most common ap-

proaches differ with respect to the camera network configuration, where cameras can be disjoint, moving, have different quality and not be synchronized. Some recent approaches still rely on some information on the network topology [20].

Another important aspect regards to the person samples. The available methods are usually classified into single-shot methods, when only one sample is analyzed at a time; or multiple-shot methods, where a collection of samples, temporally related or not, are available to build the person model. Other characteristics include full body or face based person models, learned or direct methods and real-time approaches.

In this section, we present some recent approaches to the problem and briefly describe them in terms of different points of view of the person re-identification problem. More detailed reviews are given in [6, 7, 19, 26].

In [3], a person descriptor called Symmetry-Driven Accumulation of Local Features (SDALF) is described, which exploits person symmetry and gives higher weights to features extracted from its neighborhood. Another approach given in [1] present a different person feature based on Mean Riemannian Covariance (MRC) matrices. In a posterior work, both alternatives are combined by a linear weighting in order to create a more powerful descriptor called SDALF+C [22].

The method proposed in [18] uses a joint Gabor filter and a covariance descriptor to address the problem in terms of both full body and face models. In [28], the person is modeled through a body part decomposition, which creates a person template. The person matching is then performed on the template subspace.

In the work of [32], the person signature is built upon mid-level filters computed on person patches. The partial Area Under Curve (AUC) score is then used as a metric for selecting the most discriminative patches. Other patch-based approaches are given in [30, 31], where human saliencies are used to describe person patches.

A multi-level approach is described in [11], where well separated samples in the high dimensional space are ignored (impostors) and an optimization problem is efficiently solved in the feature space. In [12], a metric learning method is proposed to relax the learning phase to a two class problem, where samples can be either from a same person or from different ones. The approach given in [14] learns different metric weights for each person, yielding better results especially when there is a large training data available.

In [16], a dictionary approach to approximating the input feature vectors as a linear combination of anchor points by using local coordinate coding (LCC) is described. This dictionary is used in the testing phase to match targets in the LCC representation. In [29], an energy-based loss function is described to compute the similarity among people in a

fast manner, paving the way for online methods.

A different approach is provided in [15], as the system is semi-supervised by a user that can enter his/her own feedback on positive and negative person samples. The method then learns this new information and update its ranking to help the user to select person matchings. An approach considering 3D person models is proposed in [8], where person descriptors are mapped into a cylindrical representation of the person. Also considering 3D models, real-time results are achieved in [2], however, relying on a calibrated set of cameras. Finally, the method proposed in [17] handles the scenario where some cameras have person label information and others do not by a domain transfer approach.

In our method, we propose the PLS signature descriptor to refine re-identification results in an iterative manner without the need of a person gallery and with weak requirements on the camera setting.

3. Proposed Method

Our proposed approach to addressing the re-identification problem is based on the work described in [23], where full body models are used to describe the person appearance. Such modeling is performed via PLS regression [27]. The proposed approach extends the work in [23] by employing a verify-correct schema to refine the final person assignment. To perform this task, we propose the use of PLS signatures to model the person tracklets.

The proposed method consists on five sequential stages to reach the first results and a final verify-correct cycle to refine them. We first describe the Partial Least Squares technique, which is used on multiple stages of the method. Then, we review the stages required to obtain the first results. Finally, we detail the verify-correct schema based on PLS signatures. Figure 1 illustrates the architecture of the proposed re-identification method.

3.1. Partial Least Squares

The PLS is a dimensionality reduction technique that uses a collection of high dimensional feature vectors and their classes. The main purpose of this method is to maximize the covariance between the classes and their feature vectors during the dimension reduction. This is performed by the creation of latent variables as a linear combination of the input ones.

The result is a weight vector that can be applied on new high dimension feature vectors to reduce their dimension. They can then be classified into classes in the low-dimensional space using a regression-based approach [25].

3.2. Re-identification Technique

The first stage consists in the training of a person detector based on PLS appearance models. For this usage of PLS,

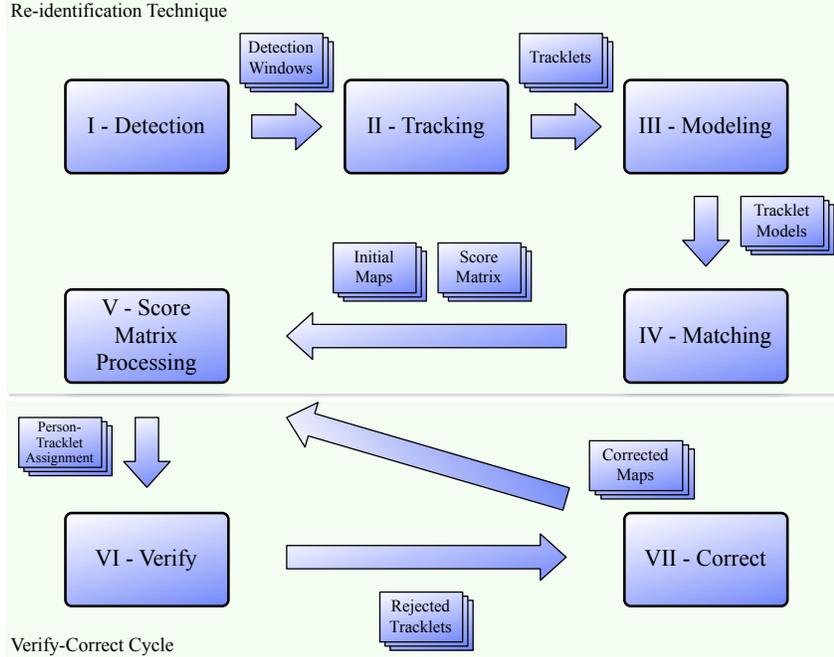


Figure 1. Architecture of the proposed method. At first, PLS models for person bodies are created using samples from the training data. These are used in the detection stage to extract detection windows from the test frames (Stage I - Detection). Person tracklets are then built by grouping them with Kalman filters (Stage II - Tracking). Then, a PLS model is created for each tracklet (Stage III - Modeling). A score matrix is built by evaluating the matching between each tracklet pair (Stage IV - Matching). Maximum values are extracted to discover the best correspondence among the detected people (Stage V - Score Matrix Processing). In the final cycle, PLS signatures are used to verify the tracklet-person assignment (Stage VI - Verify) and correct it, generating refined results (Stage VII - Correct).

only two classes are considered for classification: positive and negative person samples [24]. The detector is trained by using low-level features obtained from overlapping blocks of person samples.

The following stage is responsible for the detection of the people present in the video frames. The output of this stage is a set of detection windows found in each of the frames. To obtain them, we first decompose the video frames into samples of different sizes and extract the same low-level features used in the training from them. Each sample is then classified into positive or negative person sample. This approach leads to many detection of the same person in different scales. To overcome this issue, a non-maximum suppression is performed to eliminate redundant detections. A full description of this stage is given in [24].

The next stage is a temporal tracking of each person throughout the video. Each detection window is tracked with a Kalman filter, resulting in a set of person tracklets. Additionally, missing detections are bridged by the tracking when they happen for a short time, that is, during only a few frames.

The obtained tracklets are organized by the method into tracklet partitions, where each partition contains tracklets that correspond to different people. This information is learned by the method using the fact that different detec-

tions in the same frame must correspond to distinct people.

These partitions are updated in the following manner. In the first video frame where a person is detected, a new partition is created and all tracklets from the people detected in this frame are assigned to it. In the other frames, one of the following situations can occur:

- *All the detections are tracked people:* only the tracklets are updated, no further work is required on the partitions.
- *A new person is found:* a new tracklet is created for this person and it is added to the current partition.
- *A person is lost:* this situation can happen when there are many missing detections or the person walks off the camera range. In this scenario, a new partition is created and all tracklets from this frame are added to it. If the same partition was used in the next frames, the lost person could return to the camera range and would be incorrectly marked as distinct from itself.

In the modeling stage, a PLS appearance model is created for each tracklet in each of the tracklet partitions, this time in a one-against-all approach. Within a partition, when building the model for a given tracklet, the remaining tracklets of the same partition are used as counter-examples.

All tracklets are then pairwise matched against each

other using the generated models in the last sequential stage. The matching results are organized in a matrix of matching scores between each tracklet pair.

3.3. Tracklet Maps

For the final stages, we use two types of auxiliary tracklet maps to keep track of similar and distinct tracklets. Each tracklet has its own similarity (S) and distinctivity (D) maps. The first holds the other tracklets that correspond to the same person, whereas the second holds tracklets that correspond to people known to be distinct.

These maps are initialized according to the following equations. As soon as the tracklet is created, it is inserted into its own similarity map (Equation 1a). All other tracklets in the same partition (P) are inserted into its distinctivity map (Equation 1b). Map initialization equations:

$$S_i \leftarrow i, \quad \forall i \in P_p \quad (1a)$$

$$D_i \leftarrow j, \quad \forall i, j \in P_p \mid i \neq j \quad (1b)$$

3.4. Score Matrix Processing

The score matrix processing stage consists of successive removals of the maximum value from the score matrix, marking the corresponding tracklets as belonging to the same person. The distinctivity maps of each tracklets are used to make sure they are not known to belong to different people.

When a matching is found, their maps are updated according to the following equations. Both maps for the two involved tracklets are merged (Equations 2a and 2b). The new similarity map is used to update each of the distinctivity maps for the tracklets found in the new distinctivity map (Equation 2c). Map update equations:

$$S_{\text{new}} \leftarrow S_i \cup S_j \quad (2a)$$

$$D_{\text{new}} \leftarrow D_i \cup D_j \quad (2b)$$

$$\begin{aligned} S_k &\leftarrow S_{\text{new}}, \quad \forall k \in S_{\text{new}} \\ D_k &\leftarrow D_{\text{new}}, \quad \forall k \in S_{\text{new}} \\ D_k &\leftarrow D_k \cup S_{\text{new}}, \quad \forall k \in D_{\text{new}} \end{aligned} \quad (2c)$$

When no more matrix removals are possible, unique person identifiers are assigned to each tracklet according to the information held in the similarity maps, in order to obtain a tracklet-person assignment.

3.5. Verify-Correct Cycle

The verification step makes use of PLS signatures to determine whether a tracklet assigned to a person should be rejected or not. In sequence, the correction step updates the tracklet maps according to the rejection information.

After all people are checked, the cycle is reinitiated and the score matrix processing stage is re-executed with the corrected maps. The method finishes when no more rejections are found.

We first describe how the PLS signatures are built for each tracklet. Then, we describe how the verification step uses them to reject person tracklets. Finally, we detail how the tracklet maps are corrected.

3.6. Partial Least Squares Signatures

After the tracklets are found in the tracking stage, we create a set of generic partitions, different from the person partitions already described. These are called generic due to the fact that they consist of person tracklets randomly selected from the whole video sequence. The random selection only makes sure that the same tracklet is not selected twice. Therefore, within a generic partition there are no requirements that tracklets belongs to the same people, they are only generic people partitions.

We then create a PLS model for each tracklet within a generic partition in a one-against-all approach, using the other tracklets as counter-examples. These models represent generic person models.

Finally, we match the prior person models from the modeling stage against all the generic people models. The matching against each generic model gives the result for this person tracklet, which should be the similar for all tracklets that belong to the same people.

A set of generic models is used because, given that real person tracklets are used to build the generic partitions, we cannot match a tracklet against itself. We then need another model to match these tracklets used to create the generic partitions.

We name the collection of matching results for a tracklet as its signature. As some tracklets are not matched against all generic models (the ones that were built using them), the effective signature dimension may vary among the tracklets.

3.7. Verification Step

In this step, both tracklet-person assignment and tracklet signatures are available. The verification consists in checking for each person if their tracklets have similar signatures.

As already mentioned, effective signature dimensions may vary among tracklets. So, we evaluate the similarity among tracklets within a person in a dimension-by-dimension schema, so that the rejection does not get biased for the tracklets with fewer dimensions.

For each person, a one dimensional Gaussian is fit for each signature dimension, as the response distribution for the tracklets in the same person are considered to be approximately the same. This Gaussian is built using all tracklets in the person, except those that were not matched against the generic model responsible for this dimension. Then, for each dimension, a tracklet is considered rejected for this dimension if its response value lies outside the region of acceptance, defined as a percentage of the Gaussian.

The final rejection decision is taken considering the total rejection count for all dimensions. If a tracklet is considered to be rejected by at least a determined fraction of the total number of dimensions, then it is also considered rejected for the person it was assigned to.

3.8. Correction Step

The tracklets rejected on the previous step should be considered as distinct from the other tracklets assigned to the same person. So, in the next iteration they will get assigned to another person that should fit it better.

For each person p , the tracklets that are found to not belong to them are put in a rejected set (R_p). This set is used to correct the maps according to the following equations. The distinctivity map for each of the rejected tracklets are updated with the non-rejected tracklets (Equation 3a). The distinctivity map for each of the non-rejected tracklets are updated with the rejected tracklets (Equation 3b). Map correct equations:

$$D_i \leftarrow j, \quad \forall i \in R_p, \quad \forall j \in S_i, \quad \forall p \mid j \notin R_p \quad (3a)$$

$$D_j \leftarrow i, \quad \forall i \in R_p, \quad \forall j \in S_i, \quad \forall p \mid j \notin R_p \quad (3b)$$

4. Experimental Results

In this section, we describe the experiment results and compare our method with the baseline described in [23].

In the detection stage, the same parameter settings of the baseline method are used. The person detector was trained by using cropped person images from the INRIA Person Dataset [5]. All PLS models use the same low-level features: histogram of oriented gradients (HOG) [5] for edge information and gray-level co-occurrence matrices (GLCM) [10] as a texture descriptor.

In the verification stage, we adopted a threshold of 95% for the single dimension rejection and iterate over a range from 5% to 30% for the final rejection decision. Values higher than 30% do not significantly alter the results.

The comparison between both approaches are given in the same three video data sets used in the baseline: CAT, PETS2006 and Techgate. The first contains 4 cameras and 4 people are present. The second is sequence S7 of PETS2006 [4], with 56 people viewed by 4 cameras. The last one has 4 people in a 6 camera environment.

The evaluating metrics are the same multiple camera metrics reported in the baseline method: Crossing Fragments (X-Frag), Crossing ID Switches (X-IDS), Returning Fragments (R-Frag) and Returning ID Switches (R-IDS) [13]. The single camera metrics are not considered, as they do not change in comparison with the baseline since no changes are performed at this level.

The results for the Crossing ID Switches metric are shown in Figure 2. It can be seen that the proposed achieve

superior results when running for the rejection parameter between 5% and 10%.

The results for the Return ID Switch metric are shown in Figure 3. Again, the method achieves the best results with the rejection parameter varying from 5% to 10%.

The results for the Crossing Fragment metric are shown in Figure 4. For this metric, the method performs better with the rejection parameter a little higher than the previous metrics, between 20% and 25%.

The results for the Returning Fragment metric are shown in Figure 5. Similarly to the Crossing Fragment metric, in this case a rejection parameter of roughly 25% is the best choice.

It is clearly observed from the results that our method is able to improve the baseline results for all data sets by considering all metrics. It is also clear that, although the best parameter choice vary depending on the adopted case, the parameters can be adjusted to better fit some desired property, as within the same metric it does not vary significantly.

5. Conclusions And Future Work

In this work, a novel approach to person re-identification based on PLS signatures is proposed to improve accuracy results in a verify-correct schema. Our method can be used in any camera setting, so it does not depend on camera calibration, FOV intersection or frame synchronization. Furthermore, it does not require a person gallery, which allows a wide range of applications.

We also demonstrate through three different video data sets that the results of the proposed method are superior than the baseline for all data sets and all considered metrics.

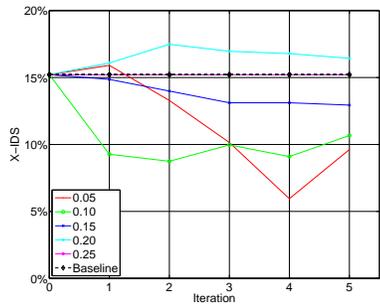
As future work, different data sets can be used to evaluate the performance of the method using videos with higher resolutions. The proposed method can also be adapted to other applications, such as face-based person re-identification.

6. Acknowledgements

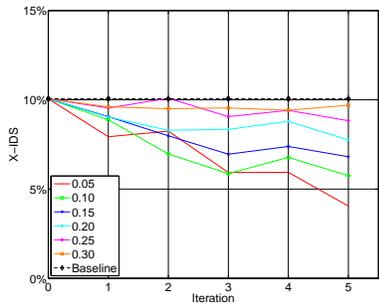
The authors are grateful to FAPESP, CAPES and CNPq (Grants #307113/2012-4, #477457/2013-4 and #487529/2013-8) for the financial support. The authors are also grateful to the organizers of the ICPR 2012 Contest - *People Tracking in Wide Baseline Camera Networks* [21] for providing the data sets.

References

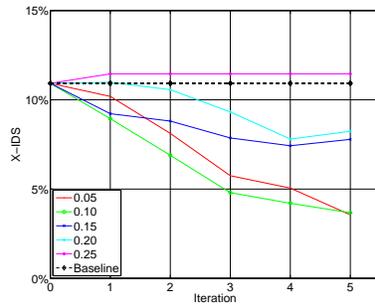
- [1] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Multiple-shot human re-identification by Mean Riemannian Covariance Grid. In *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pages 179–184, Aug. 2011.



(a) CAT data set.

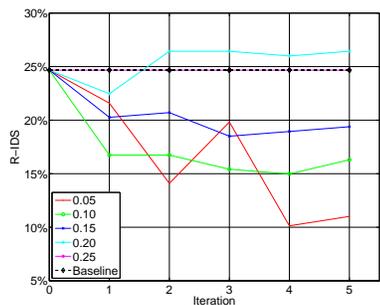


(b) PETS2006 data set.

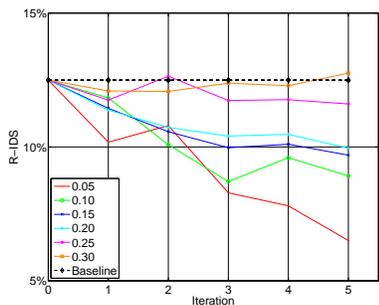


(c) Techgate data set.

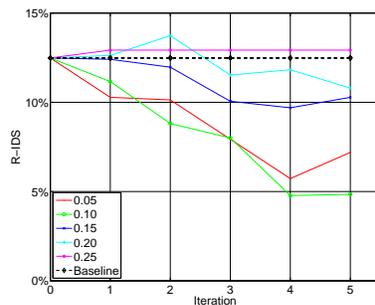
Figure 2. Results for Crossing ID Switches.



(a) CAT data set.

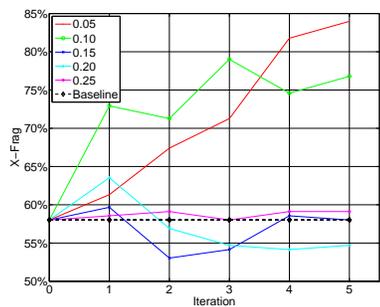


(b) PETS2006 data set.

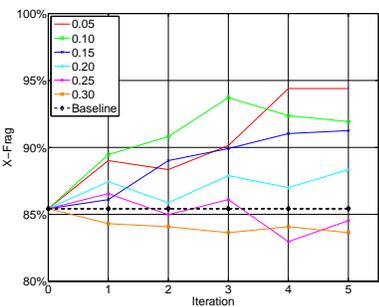


(c) Techgate data set.

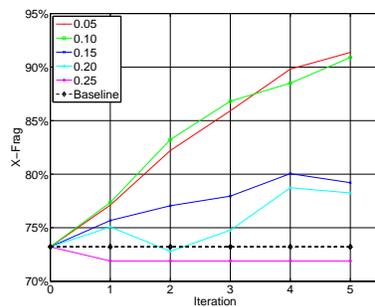
Figure 3. Results for Returning ID Switches.



(a) CAT data set.

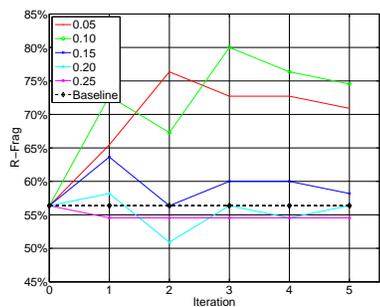


(b) PETS2006 data set.

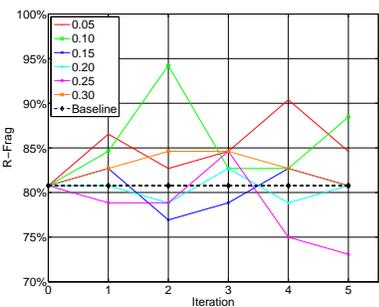


(c) Techgate data set.

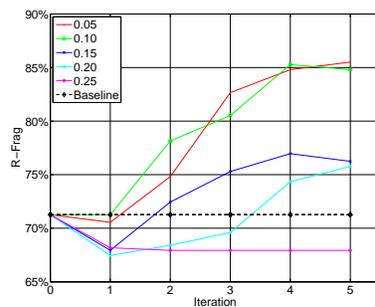
Figure 4. Results for Crossing Fragments.



(a) CAT data set.



(b) PETS2006 data set.



(c) Techgate data set.

Figure 5. Results for Returning Fragments.

- [2] D. Baltieri, R. Vezzani, and R. Cucchiara. SARC3D: A New 3D Body Model for People Tracking and Re-identification. In *International Conference on Image Analysis and Processing*, volume 6978 of *LNCS*, pages 197–206. 2011.
- [3] L. Bazzani, M. Cristani, and V. Murino. SDALF: Modeling Human Appearance with Symmetry-Driven Accumulation of Local Features. In *Person Re-Identification, Advances in Computer Vision and Pattern Recognition*, pages 43–69. 2014.
- [4] L. L. D. Thirde and J. Ferryman. Overview of the PETS2006 Challenge. In *9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2006)*, New York, USA, June 2006.
- [5] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, San Diego, CA, USA, June 2005.
- [6] T. D’Orazio and G. Cicirelli. People Re-Identification and Tracking from Multiple Cameras: A Review. In *IEEE International Conference on Image Processing*, Orlando, FL, USA, Sept./Oct. 2012.
- [7] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher. Appearance-based Person Reidentification in Camera Networks: Problem Overview and Current Approaches. *Journal of Ambient Intelligence and Humanized Computing*, 2:127–151, 2011.
- [8] T. Gandhi and M. Trivedi. Person Tracking and Reidentification: Introducing Panoramic Appearance Map (PAM) for Feature Representation. *Machine Vision and Applications*, 18:207–220, 2007.
- [9] D. Gray, S. Brennan, and H. Tao. Evaluating Appearance Models for Recognition, Reacquisition, and Tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007.
- [10] R. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. In *IEEE Transactions on Systems, Man and Cybernetics*, volume SMC-3, pages 610–621, Nov. 1973.
- [11] M. Hirzer, P. M. Roth, and H. Bischof. Person Re-identification by Efficient Impostor-Based Metric Learning. In *Advanced Video and Signal-Based Surveillance*, pages 203–208, 2012.
- [12] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed Pairwise Learned Metric for Person Re-identification. In *European Conference on Computer Vision*, volume 7577 of *LNCS*, pages 780–793. 2012.
- [13] C. Kuo, C. Huang, and R. Nevatia. Inter-Camera Association of Multi-Target Tracks by On-Line Learned Appearance Affinity Models. In *11th European Conference on Computer Vision*, pages 383–396, Crete, Greece, Sept. 2010.
- [14] W. Li, R. Zhao, and X. Wang. Human Reidentification with Transferred Metric Learning. In *Asian Conference on Computer Vision*, 2012.
- [15] C. Liu, C. Loy, S. Gong, and G. Wang. POP: Person Re-identification Post-rank Optimisation. In *IEEE International Conference on Computer Vision*, pages 441–448, Dec. 2013.
- [16] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-Supervised Coupled Dictionary Learning for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014.
- [17] A. J. Ma, P. C. Yuen, and J. Li. Domain Transfer Support Vector Ranking for Person Re-identification without Target Camera Label Information. In *IEEE International Conference on Computer Vision*, Dec. 2013.
- [18] B. Ma, Y. Su, and F. Jurie. BiCov: A Novel Image Representation for Person Re-identification and Face Verification. In *British Machine Vision Conference*, page 11, 2012.
- [19] R. Mazzon, S. Tahir, and A. Cavallaro. Person Re-identification in Crowd. *Pattern Recognition Letters*, 33(14):1828–1837, Oct. 2012.
- [20] B. Meden, F. Lerasle, and P. Sayd. MCMC Supervision for People Re-identification in Nonoverlapping Cameras. In *British Machine Vision Conference*, pages 66.1–66.11, 2012.
- [21] People Tracking in Wide Baseline Camera Networks. 21st International Conference on Pattern Recognition Contest, 2012. <http://www.wide-baseline-camera-network-contest.org>.
- [22] S. J. Poletti, V. Murino, and M. Cristani. SDALF+C: Augmenting the SDALF Descriptor by Relation-Based Information for Multi-shot Re-identification. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8259 of *LNCS*, pages 423–430. 2013.
- [23] G. L. Prado, W. R. Schwartz, and H. Pedrini. Person Re-identification Using Partial Least Squares Appearance Modeling. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 8259 of *LNCS*, pages 382–390. 2013.
- [24] W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis. Human Detection Using Partial Least Squares Analysis. In *12th IEEE International Conference on Computer Vision*, pages 24–31, Kyoto, Japan, Sept./Oct. 2009.
- [25] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis. Face Identification Using Large Feature Sets. *IEEE Transactions on Image Processing*, 21(4):2245–2255, 2012.
- [26] R. Vezzani, D. Baltieri, and R. Cucchiara. People Re-identification in Surveillance and Forensics a Survey. *ACM Computing Surveys*, 46(2):29:1–29:37, Dec. 2013.
- [27] H. Wold. Partial Least Squares. *Encyclopedia of Statistical Sciences*, 6:581–591, 1985.
- [28] Y. Xu, L. Lin, W.-S. Zheng, and X. Liu. Human Re-identification by Matching Compositional Template with Cluster Sampling. In *IEEE International Conference on Computer Vision*, Dec. 2013.
- [29] G. Zhang, Y. Wang, J. Kato, T. Marutani, and K. Mase. Local Distance Comparison for Multiple-shot People Re-identification. In *Asian Conference Computer Vision*, volume 7726 of *LNCS*, pages 677–690. 2013.
- [30] R. Zhao, W. Ouyang, and X. Wang. Person Re-identification by Saliency Matching. In *IEEE International Conference on Computer Vision*, Dec. 2013.
- [31] R. Zhao, W. Ouyang, and X. Wang. Unsupervised Saliency Learning for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2013.
- [32] R. Zhao, W. Ouyang, and X. Wang. Learning Mid-level Filters for Person Re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014.