Person Re-identification Using Robust Brightness Transfer Functions Based on Multiple Detections

Amran Bhuiyan(✉), Behzad Mirmahboub, Alessandro Perina, and Vittorio Murino

Pattern Analysis and Computer Vision (PAVIS),
Istituto Italiano di Tecnologia, Genova, Italy
amran.bhuiyan@iit.it

Abstract. Re-identification systems aim at recognizing the same individuals in multiple cameras and one of the most relevant problems is that the appearance of same individual varies across cameras due to illumination and viewpoint changes. This paper proposes the use of *Minimum Multiple Cumulative Brightness Transfer Functions* to model this appearance variations. It is multiple frame-based learning approach which leverages consecutive detections of each individual to transfer the appearance, rather than learning brightness transfer function from pairs of images. We tested our approach on standard multi-camera surveillance datasets showing consistent and significant improvements over existing methods on two different datasets without any other additional cost. Our approach is general and can be applied to any appearance-based method.

Keywords: Re-identification · Brightness transfer function · Video surveillance

1 Introduction

Person re-identification (ReID) refers to the problem of recognizing individuals at different times and locations. A schematic illustration of the problem is given in Fig. 1a, where the task is to match detections of the same person acquired by the two cameras. Re-identification involves different cameras, views, poses and illuminations and it has recently drawn a lot of attention due to its significant role in visual surveillance systems, including person search and tracking across disjoint cameras.

The core assumption in re-identification is that individuals do not change their clothing so that appearances in the several views are similar, nevertheless it still consists in a very challenging task due to the non-rigid structure of the human body, the different perspectives with which a pedestrian can be observed, and the highly variable illumination conditions (as an example see the images of the same lady on the top of Fig. 1a).

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 449–459, 2015.
DOI: 10.1007/978-3-319-23234-8_42

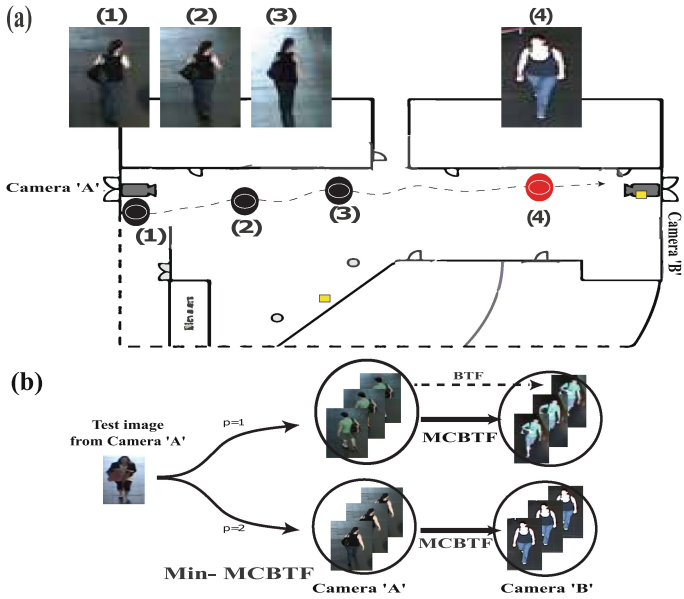

Fig. 1. (a) Typical indoor system; (b) Overview of our approach.

Re-identification approaches can mainly be organized in two classes of algorithms: direct and learning-based methods. In the former group, algorithms search for the most discriminant features to form a powerful descriptor for each individual [1–6]. In contrast, learning-based methods have techniques that learn metric spaces where to compare pedestrians, in order to guarantee a high re-identification rates [7–12]. Finally, we find methods that learn the transformation that the appearance of a person undergoes when passing from one domain to another [13–17].

This work lies in the latter, camera-specific, category, which is very relevant in large video surveillance networks where individuals are observed using various cameras across a large environment.

A thorough review of the state-of-the-art shows how these approaches mainly aim to model a function to transfer "appearance" cues between cameras. For example, [13] estimate the brightness transfer function, in short BTF, to transfer the appearance for object tracking. By employing a set of N labeled pairs $\{(I_A^p, I_B^p)\}$, where I_X^p represents an observation of pedestrian p acquired by camera X, they learn multiple BTFs, one for each pedestrian and then rely on a *Mean-BTF*(MBTF) [13]. Porikli [18] used the same setup previously but estimate the transfer function in the form of color. Later, Prosser [14] proposed the *Cumulative-BTF*(CBTF) amalgamating the training pairs before computing the transfer function. In contrast to MBTF and CBTF which end up with a single transfer function, Datta [15] proposed a *Weighted-BTF* (WBTF) that assigns different weights to test observations based on their proximity to training observations. The latter approach showed a remarkable improvement over [14,18] and therefore we will consider it as our main comparison.

This paper makes another step forward and taking inspiration from real scenarios it bridges the works of [13] and [14]. Actually, most of the state-of-the-art methods learn a single BTF from a pair of images, but nowadays we have powerful tools for person tracking and we robustly have access to at least 5-10 detections of the same individual in consecutive frames [19]. A mild criticism of single pair-based method lies in how they choose labeled pairs. Fig. 1a shows 3 detections from camera A and the detection from camera B. It is easy to figure out how a transfer function learned from the pair (1)-(4) would behave differently from the one learned from the pair (3)-(4). The question we pose here is that if and how these very similar sets of images can be exploited to learn more robust and principled transfer functions. Examples of detections for the same pedestrian are shown in Fig. 1b and, although at a first glance it may appear they do not add anything, we will show that considering all of them is indeed useful. More specifically, we propose here the use of the *Minimum Multiple Cumulative Brightness Transfer Function* (Min-MCBTF). Our approach assigns minimum distance for ReID from the distances calculated using all the MCBTFs individually which, exploiting multiple detections, is more robust of the previous approach based on single pairs. To be more specific, our approach showed the cumulative effect of multiple detections for learning BTF on ReID performance. Our technique is general and strongly outperforms previous appearance transfer function based methods [14,15,18] and the basic framework upon which we built it. Unlike previous work, we also considered the effect of increasing the number of pedestrian in the validation set.

The rest of the paper is organized as follows: Sec. 2 describes the proposed algorithm, Sec. 3 illustrates the re-identification framework in which it is used, Sec. 4, we present an exhaustive experimental session and, finally, concluding remarks are drawn in Sec. 5.

2 Minimum Cumulative Brightness Transfer Function

Our goal is to find the correspondence between multiple observations of an pedestrian across a camera pair C_i and C_j. As in previous work, we assume a limited validation set of labeled detections that can be used to calculate an inter-camera MCBTF. Subsequently the MCBTF of each pedestrian in the validation set are used to transform the test pedestrian and consider the minimum distance to form the final Min-MCBTF.

We assume to have $N \leq 10$ subsequent frames for each of the P pedestrians in the validation set which we used to learn the transfer functions. To obtain such images, we assume the reliability of a tracking algorithm able to detect single pedestrians for less then a second[1], or alternatively one could simply propagate the detected bounding box for $\frac{N}{2}$ before and after the "labeled" detection, as illustrated by Fig. 2a. In this sense, our approach does not increase the amount of labeled data needed.

[1] In standard conditions trackers run at 25 FPS.

To compute the Min-MCBTF, it is necessary to understand the extraction procedure of brightness transfer function, proposed by Javed et al. [13]. In principle, it would be necessary to estimate the pixel-to-pixel correspondences between the pedestrian images in the two camera views, however this is not possible due to self-occlusion and pose difference. Thus, to be robust to occlusions and pose differences, normalized histograms of object brightness values are employed for the BTF calculation under the assumption that the percentage of the image pixels on the observed image I_i with brightness less than or equal to ρ_i is equal to the percentage of image points in the observation I_j with brightness less than or equal to ρ_j. Now, let H_i and H_j be the normalized cumulative histograms of observations I_i and I_j respectively. More specifically, for H_i each bin of brightness value $\rho_1, \ldots, \rho_m, \ldots, \rho_M$ related to one of the three color channels is obtained from the color image I_i as follows:

$$H_i(\rho_m) = \sum_{k=1}^{m} I_i(\rho_k) \tag{1}$$

where $I_i(\rho_k)$ is the pixel count of brightness value ρ_k in I_i. $H_i(\rho_i)$ represents the proportion of H_i less than or equal to ρ_i, then $H_i(\rho_i) = H_j(\rho_j)$ and the BTF function $H_{i \to j}$ can be defined:

$$H_{i \to j}(\rho_i) = H_j^{-1}(H_i(\rho_i)) \tag{2}$$

with H^{-1} representing the inverted cumulative histogram.

As the first step of our approach, we compute a cumulative normalized version of the MCBTF. The cumulative histogram cH_i^p considering the N detection of a pedestrian p in camera view i, can be computed from the brightness values as:

$$cH_i^p(\rho_m) = \frac{1}{M \cdot N} \sum_{m=1}^{M} \sum_{n=1}^{N} I_n^p(\rho_m) \tag{3}$$

being M the number of different brightness levels. The normalization is necessary as bounding boxes can have different sizes. Then, similarly to Eq. 2, its MCBTF is computed as follows:

$$CH_{i \to j}^p(\rho_i) = cH_j^{-1}(cH_i(\rho_i)) \tag{4}$$

We use these MCBTFs to map the illumination from C_i to C_j. In this way, a given test image is transformed into a number of same test images based on the number of MCBTFs specified by the pedestrians present in the validation set.

At this point, we extract the features and compute the distances (described in the section 3.2) of all the transformed test images from the gallery and consider the minimum distance to calculate the ReID score. Mathematically,

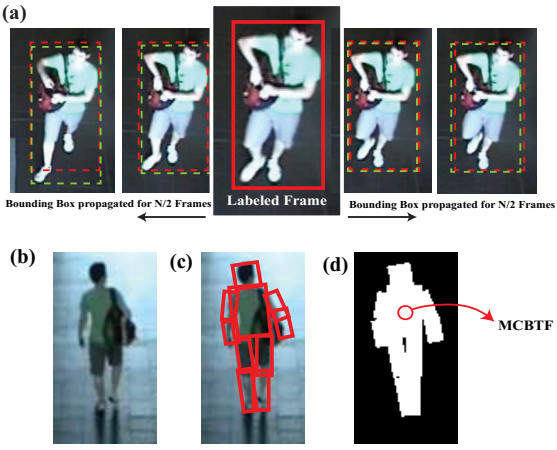

Fig. 2. (a) Bounding Box propagation (red) and actual tracking result (green); (b) One image from the SAIVT-SoftBio; (c) Custom Pictorial Structure (CPS); (d) Segmentation mask M_z derived its CPS.

$$d\left(S_i, S_j\right) = \min_p \{d\left(\tilde{S}_i^p, S_j\right)\}_{p=1}^p \tag{5}$$

where, '$\sim$' indicates the transformed version of the main context after applying MCBTF, S represents the signatures extracted from the respective images and d is the respective feature distances for calculating ReID score.

3 Re-Identification with Min-MCBTF

The aim of this section is to summarize the framework we used for re-identification, nevertheless our method is clearly independent from any appearance direct method employed. The goal of re-identification is to assign to a test image seen in camera C_i an "identity" choosing among the G identities present in the gallery at camera C_j which acts as training set. We summarize our approach by the following three steps: *i)* first, we calculate $CH_{i\rightarrow j}^p$ to transfer the appearance from C_i to C_j using validation set and transformed the test images accordingly as explained in the previous section, *ii)* second, we isolate the actual body appearance from the rest of the scene and we extract a feature signature from its foreground, and *iii)* third, we match the transformed signatures with the gallery and select top matching identities as explained in the previous section. In the following we detail the second and the last steps.

3.1 Pedestrian Segmentation

In previous section, we showed how transfer functions are learned from the whole image, however to increase the robustness, we apply the transfer function to the foreground normalized histogram only.

We performed this separation by exploiting the Custom Pictorial Structure (CPS) [1]. CPS is based on the framework of [20] where the parts are initially located by general part detectors, and then a full body pose is inferred by solving their kinematic constraints.

In CPS [1], the Histogram of Oriented Gradient (HOG) [21] feature based part detector and a linear discriminant analysis (LDA) [22] classifier is used. Moreover, CPS [1] also used the belief propagation algorithm to infer MAP body configurations from the kinematic constraints, isolated as a tree-shaped factor graph. An example of CPS is shown in Fig. 2b-c.

Finally, we generate a segmentation mask M_z employing the following rule $M_z = 1$ if z belongs to at least one of the foreground parts, otherwise $M_z = 0$; this is shown in Fig. 2d.

3.2 Feature Extraction and Matching

The feature extraction stage consists in distilling complementary aspects from each body part in order to encode heterogeneous information, so capturing distinctive characteristics of the individuals. There are many possible cues useful for a fine visual characterization. We use standard ReID features: color histogram and maximally stable color regions (MSCR) already considered in [1,2,4,23].

As for feature matching to calculate re-id score, we use the combination of Bhattacharyya distance for histogram signature and the MSCR distance for MSCR feature, as previously done in [4].

4 Experiment

In this section, we compare the performance of Min-MCBTF with the base framework upon which they are applied, e.g., [1] and the state-of-the-art in transfer functions [14,15,18]. It is important to note that *i)* all the transfer functions are applied to the same framework and *ii)* MCBTF can in principle be applied to any other appearance-based direct approach. So comparison with other methods makes little sense.

Datasets: Two publicly available person re-identification benchmarks datasets were used for our experiments, including SAIVT-SoftBio [24] and PRID 2011 [25].

- ***SAIVT-SoftBio:*** As first dataset, we considered SAIVT-SoftBio [24]. It includes annotated sequences (704×576 pixels, 25 frames per second) of 150 people, each of which is captured by a subset of eight different cameras placed inside an institute, providing various viewing angles and varying illumination conditions. A coarse box indicating the location of the annotated person in each frame is provided. We chose this dataset because it provides consecutive frames of same person which is suitable to evaluate the performance of our approach. We considered one pair of similar view cameras (3-8) and one pair of dissimilar view (5-8) (see [24] for more details).

- ***PRID 2011:*** To further evaluate our approach, we considered PRID 2011 [25] dataset. The dataset consists of images extracted from trajectories recorded from two static outdoor cameras. Images from these cameras contain a viewpoint change and a stark difference in illumination, background and camera characteristics. We considered first 200 persons who appear in both camera views. It also provides consecutive frames of same person like previous one which is suitable to evaluate the performance of our approach.

Evaluation: For each camera pair, we fixed the number of identities in the gallery to $G = 50$. In all our experiments the gallery and the validation set are kept disjoint and we repeated each task 10 times by randomly picking the identities in validation and gallery. We varied the number of pedestrian $P = [0, 2, 3, 4, 5]$ in the validation set, picking $N = 5$ consecutive frames for each pedestrian (we used the bounding boxes provided by a tracking algorithm). When $P = 0$, no transfer function is used and our framework becomes the same as [1].

We quantify re-identification performance using two standard measures, i.e. Cumulative Matching Characteristic (CMC) curve and normalized area under CMC curve (nAUC). Cumulative Matching Characteristic (CMC) curve is a plot of the recognition performance vs. the ReID ranking score. It represents the expectation of finding the correct match in the top k matches. To compare the results numerically at-a-glance, we relied on the normalized area under the CMC ($nAUC$). We report all the CMC curves setting P=5, consistently with [15]. We further quantified ReID performance using the graphs with P-vs-$nAUC$.

As first experiment, we used our approach to compare with the state-of- the-art of transfer functions as well as the base ReID framwork. Fig. 3 and Fig. 4 report the experimental findings for both datasets respectively. In the legend we also report the nAUC for each method, which gives an idea of the trend of the curves across all the ranks. Our approach handily outperforms the base CPS [1] ReID framework, as well as CBTF [14], WBTF [15], and MBTF [13]. The improvement over the state-of-the-art at first rank is particularly noticeable: there is 10-30% differences at the position of rank-1 in the CMC curves between our performance and competing methods. Furthermore, our proposed approach works consistently for all the datasets unlike other methods. Note that considering limited number of pedestrians P in the validation set, we have been able to learn the robust brightness transfer function.

As second test, we evaluated the robustness of our approach to transfer function; in the specific, we considered the *Inter-Camera Color Calibration-* ICC of Porikli et al. [18]. Working in the exact same way of Sec. 2, we adapted (for the first time) [13–15] and we compared with our *Minimum-Multiple-Cumulative-*ICC (Min-MCICC). The results of color calibration for all the dataset is shown in Fig. 5.

Fig. 5 shows the experimental results of the color calibration method for all the datasets. Again our proposed approach yields the best performance. Note that WBTF [15], CBTF [14] and MBTF [13] techniques do not yield the base ReID performance upon which we apply our method consistently, but applying

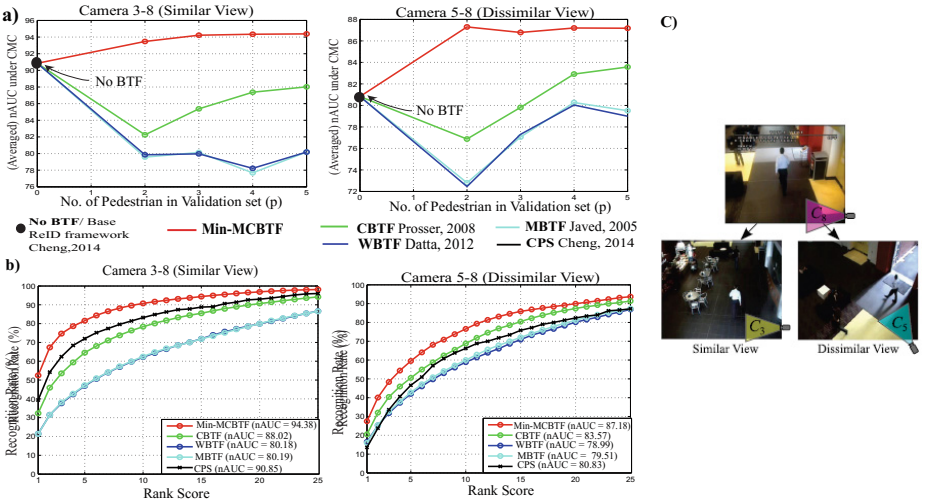

Fig. 3. (a) *P*-vs-*nAUC* for two camera pairs of SAIVT-SoftBio; (b) CMC and nAUC; (c) Example frames of a person in the selected similar and dissimilar cameras [24].

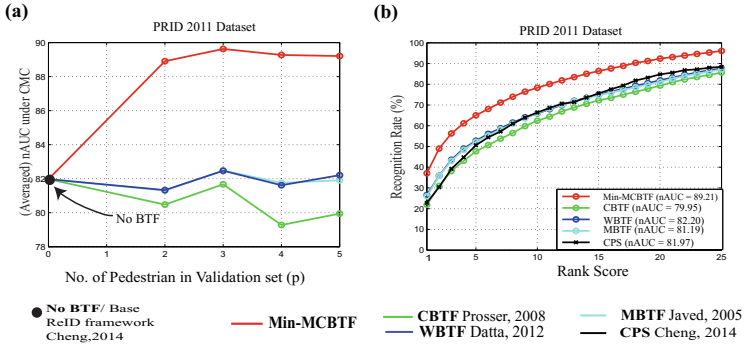

Fig. 4. (a) *P*-vs-*nAUC* PRID 2011 dataset; (b) CMC and nAUC of PRID 2011 dataset.

color calibration method in the same ways gives us WICC, CICC and MICC which work better than the base CPS [1] ReID framework as shown in Fig. 5.

As final test, we tested our approach using the signature that has been proposed by Listanti et al. [6]. The authors designed a descriptor of person appearance for re-identification based on coarse, striped poling local features. It does not require sophisticated background or body part modeling, instead used a central point kernel to approximately segment foreground from background. It should be mentioned that, we did not implement the iterative sparse basis expansion as did on [6], instead we use the Bhattacharyya distance for feature matching to calculate the ReID score.

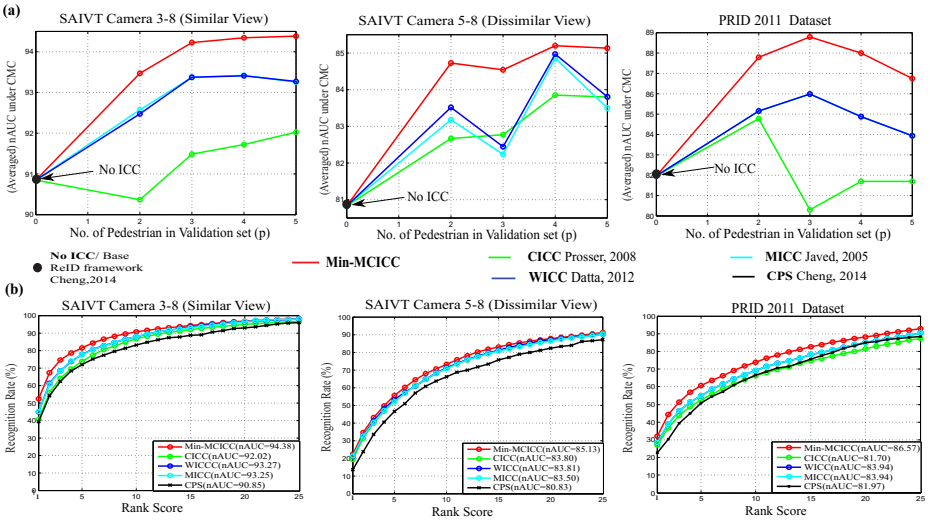

Fig. 5. (a) *P*-vs-*nAUC* of all the datasets; (b) CMC and nAUC of all the datasets.

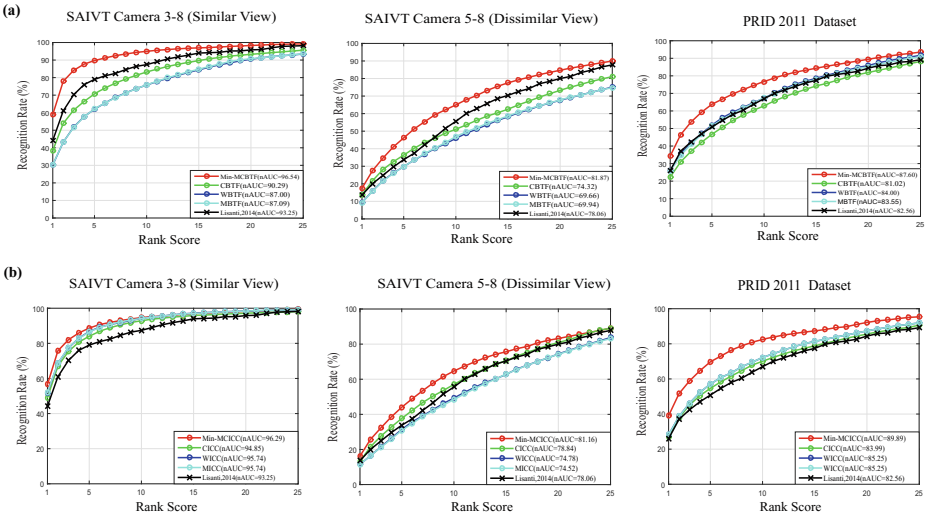

Fig. 6. (a) CMC for brightness transfer function (BTF) of all the datasets ; (b) CMC for inter-camera color calibration (ICC) for all the datasets.

Fig. 6 reports the CMC curves for above mentioned descriptor using our approach. The improvement over the state-of-the-art using our approach for this descriptor is clearly visible both for brightness transfer and color calibration function. In the legend we also report the nAUC for each method, which gives

an idea of the trend of the curves across all the ranks. Again, it has been evident the effectiveness of our method for any appearance-based ReID problems.

All the state-of-the-art use single detection for learning the brightness transfer function. The CBTF [14] also use single detection of each observation and then accumulate all the observations before computing a BTF. In contrast, our approach use multiple consecutive detections of each individual and calculate the cumulative normalized histogram before calculating BTF for each individual. Analyzing all the experimental findings, we can say that our proposed method is able to transfer appearance in the form of brightness more robust than any other methods and works consistently for all the ReID datasets, outperforming the state-of-the-art.

5 Conclusion

This paper proposes the use of cumulative histograms from multiple consecutive detections to learn better and more robust transfer functions. Augmenting the pool of labeled data *within* a camera can be easily carried out by relying on tracking algorithms or simply by propagating the label for few frames. Our results clearly demonstrate a significant improvement over previous ways to model appearance variations. Moreover, our propose approach is general and can be applied to model appearance variation problems beyond person re-identification.

References

1. Cheng, D., Cristani, M.: Person re-identification by articulated appearance matching. in person re-identification, isbn 978-1-4471-6295-7. Springer (2014)
2. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR (2010)
3. Bazzani, L., Cristani, M., Perina, A., Farenzena, M., Murino, V.: Multiple-shot person re-identification by hpe signature. In: ICPR, pp. 1413–1416 (2010)
4. Bhuiyan, A., Perina, A., Murino, V.: Person re-identification by discriminatively selecting parts and features. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops. LNCS, vol. 8927, pp. 147–161. Springer, Heidelberg (2015)
5. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 99 (2012)
6. Lisanti, G., Masi, I., Bagdanov, A., Bimbo, A.: Person re-identification by iterative re-weighted sparse ranking. IEEE Transactions on Pattern Analysis and Machine Intelligence **PP**(99) (2014)
7. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part IV. LNCS, vol. 6495, pp. 501–512. Springer, Heidelberg (2011)
8. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
9. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 391–401. Springer, Heidelberg (2012)

10. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 413–422. Springer, Heidelberg (2012)
11. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVC (2010)
12. Zheng, W., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR, pp. 649–656 (2011)
13. Javed, O., Shafiq, K., Shah, M.: Appearance modeling for tracking in multple non-overlapping cameras. In: CVPR (2005)
14. Prosser, B., Gong, S., Xiang, T.: Multi-camera matching using bi-directional cumulative brightness transfer functions. In: BMVC (2008)
15. Datta, A., Brown, L.M., Feris, R., Pankanti, S.: Appearance modeling for person re-identification using weighted brightness transfer functions. In: ICPR (2012)
16. Brand, Y., Avraham, T., Lindenbaum, M.: Transitive re-identification. In: BMVC (2013)
17. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 381–390. Springer, Heidelberg (2012)
18. Porikli, F.: Inter-camera color calibration using cross correlation model function. In: ICIP (2003)
19. Chen, X., Bhanu, B.: Soft biometrics integrated multi-target tracking. In: ICPR (2014)
20. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: people detection and articulated pose estimation. In: CVPR (2009)
21. Dalal, N., Triggs, B.: Histogram of oriented gradients for human detection. In: CVPR (2005)
22. Forssen, P.E.: Maximally stable color regions for recognition and matching. In: CVPR (2007)
23. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. Computer Vision and Image Understanding **117**, 130–144 (2013)
24. Bialkowski, A., Denman, S., Lucey, P., Sridharan, S., Fookes, C.C.: A database for person re-identification in multi-camera surveillance networks. In: Digital Image Computing: Techniques and Application (DICTA 2012), pp. 1–8 (2012)
25. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011). www.springerlink.com