Article
Exploiting Multiple Detections for Person Re-Identification

Amran Bhuiyan *, Alessandro Perina and Vittorio Murino

Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Via Morego 30,
16163 Genova, Italy; alessandro.perina@gmail.com (A.P.); vittorio.murino@iit.it (V.M.)
* Correspondence: amran.apece@gmail.com; Tel.: +39-331-803-7176

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 cumulative weighted brightness transfer functions (CWBTFs) to model these appearance variations. Different from recently proposed methods which only consider pairs of images to learn a brightness transfer function, we exploit such a multiple-frame-based learning approach that leverages consecutive detections of each individual to transfer the appearance. We first present a CWBTF framework for the task of transforming appearance from one camera to another. We then present a re-identification framework where we segment the pedestrian images into meaningful parts and extract features from such parts, as well as from the whole body. Jointly, both of these frameworks contribute to model the appearance variations more robustly. We tested our approach on standard multi-camera surveillance datasets, showing consistent and significant improvements over existing methods on three different datasets without any other additional cost. Our approach is general and can be applied to any appearance-based method.

Keywords: video surveillance; appearance-based re-identification; brightness transfer function; segmentation

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 Figure 1, 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 disjointed 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 of 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. Illumination variations between non-overlapping camera views can have an immense effect on the appearance of an individual, thus increasing the difficulty of associating people in person re-identification. The variations of illumination can also be observed in the same individual at different instants, even for the same camera. A schematic illustration of the problem is given in Figure 1, where the task is to match detections of the same person acquired by the two cameras.
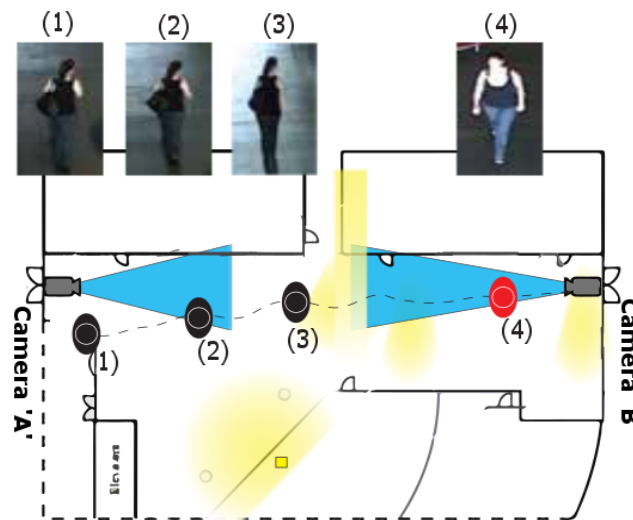
Figure 1. Typical indoor system.

To deal with the appearance variations, modeling the transfer of appearance associated with the two specific cameras [1–4] achieved impressive performance for re-identification by learning a single brightness transfer function (BTF) from a pair of images. However, 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 [5]. A mild criticism of the single-pair-based method lies in how they choose labeled pairs. Figure 1 shows three 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 related to 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 Figure 1; although at a first glance it may appear that they do not add anything, through rigorous experimental analysis we show that considering all of them is indeed useful. The main contributions of our approaches are:

- We propose the use of the cumulative weighted brightness transfer function (CWBTF). Our approach assigns unequal weights to each CBTF which, exploiting multiple detections, is more robust than the previous approaches based on single pairs. Our technique is general and strongly outperforms previous appearance transfer function-based methods.
- We propose an improved stel component analysis (improved SCA) segmentation technique, which is quite effective for pedestrian segmentation.
- A rigorous experimental phase validating the advantages of our approach over existing alternatives on multiple benchmark datasets with variable numbers of cameras.

The rest of the paper is organized as follows: Section 2 presents a comprehensive state of the art of re-identification, Section 3 describes the proposed CWBTF algorithm, Section 4 illustrates the re-identification framework in which CWBTF is used, in Section 5 we present an exhaustive experimental session and, finally, concluding remarks are drawn in Section 6.

2. Related Work

In the last few years, the problem of re-identifying persons across multiple non-overlapping cameras has received increasing attention in the field of computer vision and pattern recognition. Existing systems still suffer from unresolved problems, and they are unable to fulfill the actual demand of surveillance applications. These circumstances led to an increased number of scientific works proposing different strategies to tackle the task of person re-identification. Approaches range from

simple hand-crafted feature-matching to more sophisticated methods that fit structural models on person images or extract discriminative information using unlabeled as well as labeled training data. The community has commonly adopted three different kinds of approaches: direct methods, metric learning methods and transform learning methods.

Direct methods focus more on designing novel features for capturing the most distinguishing aspects of an individual. In [6], a descriptor is proposed by subdividing the person in horizontal stripes, and keeping the median color of each stripe accumulated over different frames. A spatio–temporal local-feature-grouping and matching is proposed in [7], where a decomposable triangular graph is built in order to capture the spatial distribution of the local descriptor over time. The method proposed in [8] segments a pedestrian image into regions, and stores the spatial relationship of the colors into a co-occurrence matrix. This technique proved to work well when pedestrians were seen from similar perspectives. In [9], speeded up robust features (SURF) interest points are collected over short video sequences and used to characterize human bodies. In [10], the authors use a sophisticated appearance model, the symmetry-driven accumulation of local features (SDALF), which models human appearance by the symmetry and asymmetry driven features, based on the idea that features closer to the symmetry axes are more robust against scene clutter than are the features far away from them. In the same pipeline, [11] proposed a pictorial structure (PS)-based ReID method by matching signature features coming from a number of well-localized body parts and manually weighing those parts on the basis of their salience. Following the same idea, [12] devised a method that automatically assigns weights to the body parts on the basis of their discriminative power. Covariance features, originally employed for pedestrian detection [13], are extracted from coarsely located body parts and tailored for ReID purposes in [14]. The extension of this work is proposed in [15] by considering the availability of multiple images of the same individual; the method adopts the manifold mean as surrogate of the different covariances coming from the multiple images. In [16], the authors propose a collaborative sparse approximation method based on multiple-shot that incorporates the discriminative power of collaborative representation into the geometric distance, which increases the robustness to across-camera appearance changes. In [17], an illumination-invariant feature descriptor is proposed based on logchromaticity (log) color space, and demonstrates that color as a single cue has a relatively good performance in identifying persons under greatly varying imaging conditions.

Learning-based approaches are the techniques that use the training data to learn metric spaces to compare pedestrians, in order to guarantee high re-identification rates. The common assumption is that the learned information from the training data could be generalized to the unseen probe. The approach presented in [18] uses boosting to select a combination of spatial and color information for viewpoint invariance. Similarly, in [19], a high-dimensional signature composed by multiple features is projected into a low-dimensional discriminant latent space by partial least-squares (PLS) reduction. Re-identification is cast as a binary classification problem (one vs. all) by [14] using Haar-like features and a part-based MPEG7 dominant color descriptor, while in [3], a relative ranking problem is cast in a higher-dimensional feature space where true and wrong matches become more separable. Pairwise constraint component cnalysis (PCCA) [20] and relaxed pairwise metric learning (RPML) [21] learn a projection metric from a high-dimensional input space to low-dimensional space, where the distance between the pairs of data points relates to the desired constraints. In [22], an adaptive boundary approach is proposed that jointly learns the metric and a locally adaptive thresholding rule. In [23], a probabilistic relative distance comparison (PRDC) model is proposed that attempts to maximize the likelihood of a true match having smaller distance than a false match. In [24], a statistical inference perspective is applied to learn a metric that satisfies equivalence constraints. In [25], a local fisher discriminant analysis (LFDA) is proposed, which is a closed-form solution of a Mahalanobis matrix required to use principal component analysis (PCA) for dimensionality reduction. However, PCA can eliminate discriminant features, defeating the benefits of LFDA. By introducing a regularization term in the methods [20,22,26] and using a series of kernel-based techniques to learn nonlinear feature

transformation functions in [25] to preserve the discriminative features, Ref. [27] reported better performance over the respective methods.

Recently, deep-learning methods have shown significant performance improvement on various computer-vision applications, such as image classification, object detection and face recognition. There has also been a growing number of methods that apply deep-learning methods [28–34] for person ReID in recent years.

In [28–30], Siamese-based deep architectures for person ReID is proposed, which usually employs two or three Siamese-based convolutional neural network (S-CNNs) for deep feature learning. In [31], both feature representation and metric learning have been treated as a joint learning problem employing an end-to-end deep convolutional architecture. It mainly introduces a new cross-input neighborhood layer in the deep architecture that effectively extracts the cross-view relationships of the features. A domain-guided dropout (DGD) [32] learns robust features by selecting the neurons specific to certain domains. However, the above-mentioned deep architectures operate on convolutional filter responses, which capture the local context of a small region that is completely independent of other regions. It is possible to exploit the dependency among local regions by utilizing long short-term memory (LSTM) cells as the constituent components in the conventional Siamese network. Taking this intuition into account, in [33], the authors propose a LSTM-based architecture to process image regions sequentially and enhance the discriminative capability of local feature representation by leveraging contextual information. Following the same pipeline, the authors in [34] proposed a sequential fusion framework that combines the frame-wise appearance information as well as temporal information to generate a robust sequence-level human representation. Nevertheless, similar to many other deep architectures, generating a huge amount of labeled training data is an issue that needs to be addressed carefully.

A third group of work tries to map the appearance from one camera to another by learning a brightness transfer function (BTF) [1–4] between cameras. Since the appearance of a person can vary significantly when viewed from different cameras, modeling appearance transformation in different forms could certainly improve the re-identification performance. This appearance mapping from one view to the other is not unique and depends on various parameters, for example, current illumination, possibly dynamic scene geometry, exposure time and so on. In [35], the re-identification problem was posed as a classification problem in the feature space formed from concatenated features of persons viewed in two different cameras. In [36], transfer learning methods were used in an inference problem setting to reduce the annotation effort inherent in a learning system. In a recent work [37], the authors proposed warp function space (WFS) to learn the transformation of features using a dynamic time-warping-based algorithm from a set of manually labeled data, and used the transformations in a discriminative framework. Reference [1] estimated the brightness transfer function 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) [1]. Porikli [2] used the same setup previously but estimated the transfer function in the form of color. Later, Prosser [3] 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, [4] 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 [2,3] and therefore we will consider it as our main comparison.

Our CWBTF approach lies in the transform learning methods, which is very relevant in large video-surveillance networks where individuals are observed using various cameras across a large environment. The proposed approaches make another step forward, and taking inspiration from real scenarios, it bridges the works of [1,3]. In addition to that, we propose a robust segmentation technique that segments the human appearance into meaningful parts, taking advantages from the multiple detections, different from [10–12].

3. Cumulative Weighted Brightness Transfer Function (CWBTF)

The goal of this section is to figure out a way to find the correspondence between multiple observations of a pedestrian across a camera pair C_i and C_j. Like previous work [4] , a limited validation set of labeled detections is used to calculate an intercamera CBTF. Subsequently, the CBTFs of each pedestrian in the validation set are weighted according to their distance with respect to the test pedestrian to form the final CWBTF.

We assume to have $N \leq 10$ subsequent frames for each of the P pedestrians in the validation set that 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 than a second (in standard conditions trackers run at 25 fps), or alternatively one could simply propagate the detected bounding box for $\frac{N}{2}$ before and after the "labeled" detection. In this sense, our approach does not increase the amount of labeled data needed.

To compute the CWBTF, it is necessary to understand the extraction procedure of the brightness transfer function, proposed by Javed et al. [1]. 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:

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

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

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

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:

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

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

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

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

At this point, given a test image and its color histogram, we compute a weighted average of cumulative-BTFs. To be robust to occlusions, camera noise or tracking errors, typical problems when considering a set of detections, we compute the root-mean-squared distances—ψ—as specified in the

Bhattacharya distance between the test image and the set of N detections of each pedestrian p in the validation set.

$$d_\psi^p = \sqrt{\frac{\sum_{n=1}^{N} \psi(I^{test}, I_n^p)^2}{N}} \tag{5}$$

It is important to note that while in [4], a weighted BTF is calculated using background only, here we are considering the full set of images. Observations in the validation set that are close in the feature space to the test observation must be assigned a higher weight for the purpose of CWBTF calculation, and therefore we compute normalized weights α^p and we average the cumulative BTFs as follows:

$$\alpha^p = \frac{e^{-\gamma \cdot d_\psi^p}}{\sum e^{-\gamma \cdot d_\psi^p}} \qquad H_{i \to j}^{CWBTF} = \sum_{p=1}^{P} \alpha^p \cdot CH_{i \to j}^p. \tag{6}$$

The parameter γ is a weight factor that can be employed to reduce α's peak, thus averaging between fewer images. This is akin to what done in [4], where the estimated transfer function is averaged only using the top K matches.

4. Re-Identification with CWBTF

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; this acts as training set. We summarize our approach by the following three steps (illustrated in Figure 2): (*i*) firstly, we calculate $CH_{i \to j}^p$ to transfer the appearance from C_i to C_j using validation set, and transform the test images accordingly as explained in the previous section; (*ii*) secondly, we isolate the actual body appearance from the rest of the scene by using our proposed improved SCA segmentation technique; (*iii*) thirdly, we extract a feature signature from its foreground, and (*iiii*) finally, 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 last three steps.
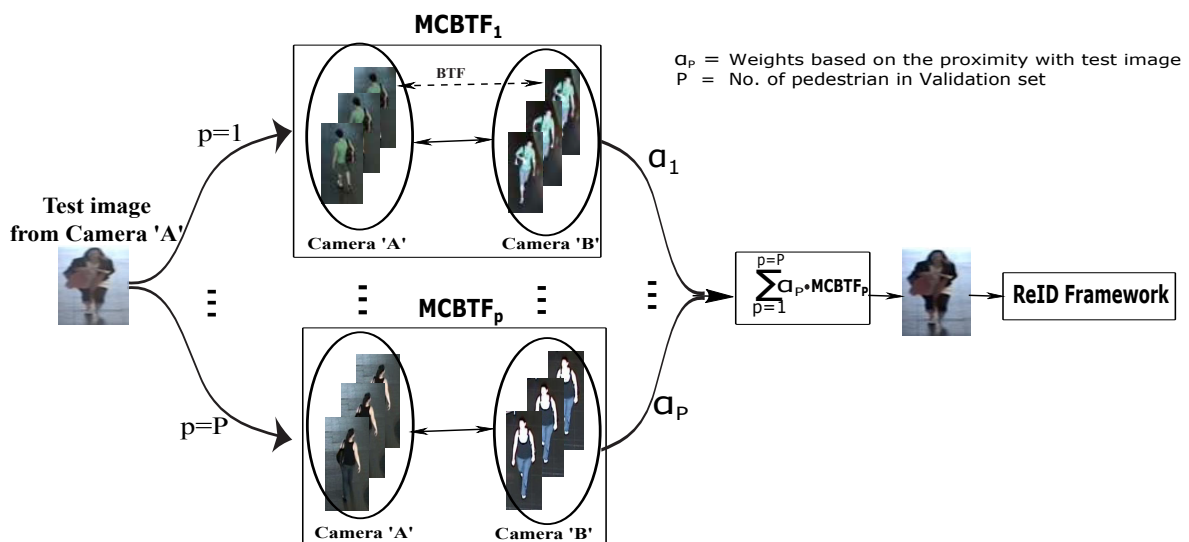

Figure 2. Overview of CWBTF approach.

4.1. Pedestrian Segmentation by Improved SCA

The aim of this phase is to isolate the actual body appearance from the rest of the scene. This allows the proposed approach to focus solely on the individual and its parts, ignoring the context in which it is immersed.

We performed this separation by using stel component analysis (SCA) [38]. This segmentation algorithm is based on the notion of a "structure element", or stel, which can be explained as an image portion (often discontinuous) whose topology is consistent over an image class. A stel often represents a meaningful and semantic part of a class of objects, like an arm or the head for pedestrians. For example, in Figure 3 we show a few images of pedestrians and their related segmentation in stels. Each color indexes a particular stel s_i while maintaining consistent the segmentation (same color, same part of the body) across the whole dataset. This is very important as it allows us to compare consistent feature signatures extracted in the various body parts.
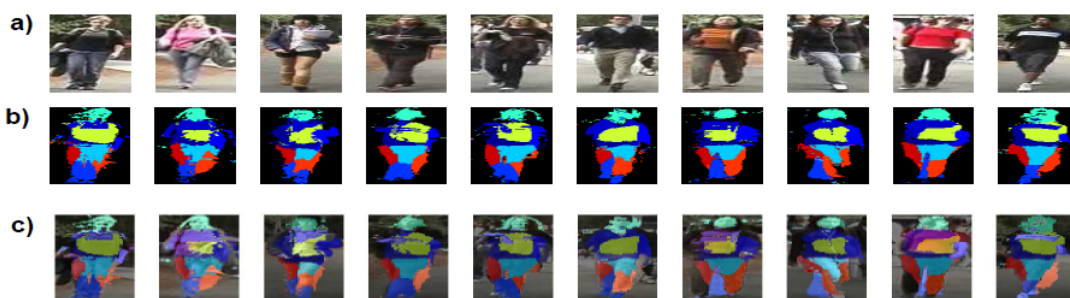

Figure 3. (**a**) Lineups of pedestrians; and (**b**,**c**) corresponding segmentations. (Best viewed in color.)

More formally, SCA captures the common structure of an image class by blending together multiple stels: It assumes that each pixel measurement x_i, with its 2D coordinate i, has an associated discrete variable s_i, which takes a label from the set $\{1, ..., S\}$. Such a labeling is generated from a stel prior $p(s_i)$, which captures the common structure of the set of images.

The model detects the image self-similarity within a segment: Pixels with the same label s are expected to follow a tight distribution over the image measurements. Finally, while the appearance similarity is local (for each image), the model insists on consistent segmentation by means of a stel prior, which is common for all the images.

SCA has been previously considered for ReID, in particular in [10,39], where the authors used it to perform a background–foreground segmentation ($S = 2$): An example of the segmentation prior $p(s_i)$ used is shown in Figure 4a. Here, for the first time, we exploited SCA's segmentation in multiple parts ($S > 2$), discarding background parts and considering the features in each part separately. An example of a learned SCA segmentation prior is shown in Figure 4a, where $S = 20$. After learning the stel prior, we manually filtered out the background stels, as shown in Figure 4b. It is worth noting that the model is learned only once and it is independent of the dataset. Furthermore, the background suppression is performed once and not for each individual image, because all the images have consistent segmentation.

In our experiments, we set the number of segments to $S = 20$ and we modeled the distribution over the image measurements as a mixture of Gaussians. To learn the segmentation prior we set the number of iterations to 50. Segmentation of new images (i.e., probe) consists of fast inference, and is done in real time.

a) SCA prior p(s)

b) Background stels are filtered out

c) Pictorial Models
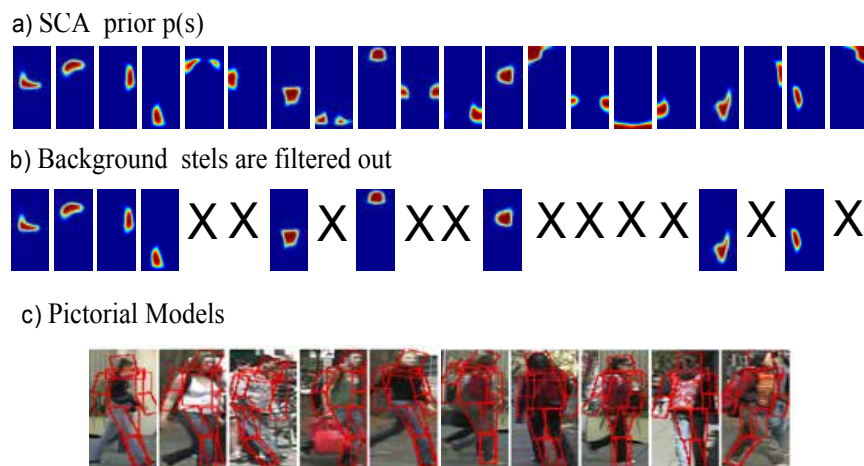
Figure 4. (**a**) Segmented part-masks for whole image; (**b**) foreground extraction procedure; (**c**) lineup of pedestrians and superimposed pictorial structures.

4.2. Feature Extraction and Matching

The feature extraction stage consists of distilling complementary aspects from each body part in order to encode heterogeneous information and to capture distinctive characteristics of the individuals. There are many possible cues useful for a fine visual characterization, and we considered two types of features: color histograms and maximally stable color regions (MSCRs) [40]. This is the same signature already used in [10,11,39,41].

Like [11,41], we employed a color histogram and the MSCR blobs as the image signature. To match two signatures, we use the combination of Bhattacharyya distance for histogram signature and the MSCR distance for MSCR features as in [10,11,39,41].

5. Experiments

In this section, we evaluate the performance of our approach by performing several illustrative experiments: (i) re-identification by the improved SCA segmentation technique in Section 5.2; (ii) re-identification by CWBTF in Section 5.3; (iii) effect of other feature representations in Section 5.4, and (iv) effect of parameter variations in Section 5.5.

5.1. Datasets and Settings

There are a number of publicly available ReID datasets, which are considered to be challenging as well as realistic for evaluating the performance of different ReID methods. The specific characteristics that define the uniqueness of a particular dataset are: the number of subject considered, the number of instances per subject, pose and illumination variations, severity of occlusions, number of viewpoints and the resolution of the captured image. Based on the number of instances of a subject in the dataset, the ReID can be divided into two categories: single-shot and multi-shot modality. In the former case, there are only two instances, one for probe and one for gallery. In the latter case, there are multiple instances per subject, which help to extract the robust information from the subject. Among all the datasets, a few of them are widely used. The characteristics of those datasets are explained as follows:

VIPeR Dataset [18] contains two views of 632 pedestrians. Each pair is made up of images of the same pedestrians taken from arbitrary viewpoints under varying illumination conditions. Each image is 128×48 pixels and presents a centered unoccluded human figure, although cropped short at the feet in some side-views. In this dataset, we compare the performance by computing the normalized area-under-curve (nAUC) value, indicated in the figure legend. Figure 5 reports the recognition accuracy for comparing the aforementioned part-based approaches.
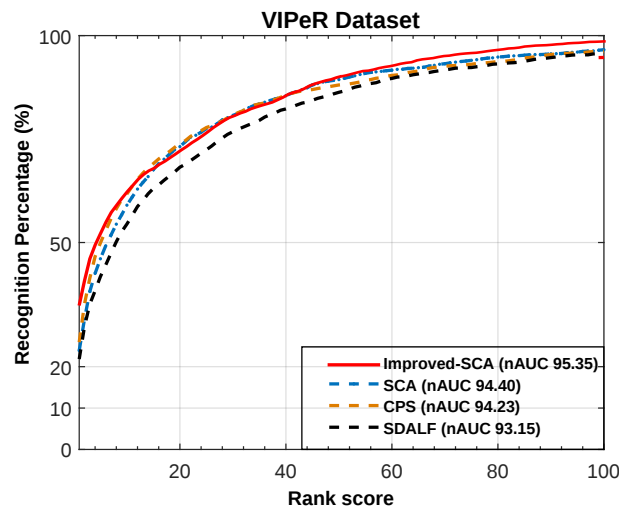
Figure 5. CMC and nAUC for VIPeR dataset.

CAVIAR4REID [11] **Dataset** contains images of pedestrians extracted from the CAVIAR repository [11], providing a challenging real-world setup. This is the only publicly available dataset where the intraperson images vary a lot in terms of resolution, light, pose and so on. The main challenges of this dataset are its broad changes in resolution, as the minimum and maximum sizes of the images are 17×39 and 72×144, respectively. Severe pose variations make it more challenging. From the 72 identified different individuals, 50 are captured by two cameras with 20 images for each of them and 22 from only one camera, with 10 images for each of them. So in short, this dataset accurately selected a set of images for each camera view for each pedestrian in order to maximize the variance with respect to resolution changes, light conditions, occlusions and pose changes so as to make challenging the re-identification task.

PRID2011 [42] **Dataset**. To further evaluate our approach, we considered the PRID 2011 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. One camera view shows 385 persons, which another shows 749 persons. The first 200 persons appear in both camera views. The remaining persons in each camera view complete the gallery set of the corresponding view.

SAIVT-SoftBio [43] **Dataset** 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. Most of the state-of-the-art studies [44,45] evaluated this dataset for two camera pairs: cameras 3–8 from similar viewpoints, and Cameras 5–8 from dissimilar viewpoints with large viewing angles.

Performance Measures. We show results in terms of recognition rate as cumulative matching characteristic (CMC) curves and normalized area-under-curve (nAUC) values, as is common practice in ReID literature [46–53]. The CMC curve is a plot of recognition performance versus re-identification ranking score, and represents the expectation of finding a correct match in the top k matches. nAUC gives an overall score of how well a ReID method performs, irrespective of the dataset size. Higher curves represent better performance. The performance can also be evaluated by computing the ranked matching rates, and these results are shown in the following tables. We report all the CMC curves setting $P = 5$, consistently with [4]. We further quantified ReID performance using the graphs with P vs. *nAUC*. Following the literature [4,24,47,53], in all our experiments the gallery and the validation set are kept disjointed and we repeated each task 10 times by randomly picking the identities in validation

and gallery. For each camera pair, we fixed the number of identities in the gallery to $G = 50$ for PRID 2011 and SAIVT-SoftBio, $G = 25$ for CAVIAR4REID and $G = 316$ for VIPeR.

5.2. Re-Identification by Improved SCA Segmentation Technique

Goal. The goal of this experiment is to analyze the performance of the proposed improved SCA that yields consistent segment-ordering across images, and allows us to discard background regions and focus the analysis on the foreground parts.

Compared Methods. We compare our approach with several re-identification methods including SDALF [10], Custom Pictorial Structures (CPS) [11] and SCA [12]. In this pipeline, SDALF [10] works as a baseline method where SCA is used only for foreground and background segmentation.

Implementation Details. We use publicly available codes for CPS and SDALF, and tested on our experimented datasets. We also implemented SCA for multiple body parts under the same experimental setup.

Results. Figure 5 and Tables 1 and 2 show the results for VIPeR, CAVIAR4REID, PRID 2011 and SAIVT-SoftBio datasets. From these tables, the following observations can be made: (i) overall performance of the proposed improved SCA for re-identification is better than all compared methods on all the datasets; (ii) among the alternatives, the SCA baseline is the most competitive, as we proposed an improved version of this method. The rank-1 performance improvements over SCA are 5.61%, 1.20% , 1.4% and 2.2% on VIPeR, CAVIAR4REID, PRID2011 and SAVIT-SoftBio datasets (similar view), respectively. For the dissimilar-view camera pair of SAIVT-SoftBio, all the ranks except rank-1 perform consistently better than SCA; (iii) it is comparatively easy to model the appearance for similar view where there are not many viewpoint variations. Table 2 shows the results for similar-view camera pairs (i.e., camera 3 and camera 8) in the SAIVT-SoftBio dataset, which suggest improved re-identification accuracy compared to all the other datasets; (iv) finally, the performance gap between our improved SCA method and SCA shows that the proposed segmentation technique is effective in exploiting information from different meaningful body parts while computing re-identification accuracy across camera pairs.

Table 1. Experiment on CAVIAR4REID and PRID2011 datasets reported by top-ranked matching rate (%) and normalized area under curve (nAUC) (%) for different methods compared with our improved SCA approach.

Methods	CAVIAR4REID					PRID2011				
	r = 1	r = 5	r = 10	r = 15	nAUC	r = 1	r = 5	r = 10	r = 15	nAUC
SDALF [10]	11.5	38.5	62	74.5	73.80	13	36.4	49.4	60.2	72.65
CPS [11]	20.25	53	71	83. 25	82.01	22.8	50.6	66.4	75.6	81.97
SCA [12]	22.75	**59.25**	71.5	81.75	82.63	27.2	**59.8**	73.6	82.2	85.28
Improved SCA	**23.95**	54.75	**73.75**	**86**	**83.10**	**28.6**	58.2	**74**	**84.2**	**85.64**

Table 2. Experiment on similar view (cameras 3–8) and dissimilar view (cameras 5–8) of the SAIVT-SoftBio dataset reported by top-ranked matching rate (%) and normalized area under curve (nAUC) (%) for different methods compared with our improved SCA approach.

Methods	SAIVT (cameras 3–8)					SAIVT (cameras 5–8)				
	r = 1	r = 5	r = 10	r = 15	nAUC	r = 1	r = 5	r = 10	r = 15	nAUC
SDALF [10]	22	49.4	65	84.4	83.20	14.8	45.2	65	85.4	81.54
CPS [11]	39.4	72	83.2	93	90.85	13.6	46.6	66.2	82.4	80.83
SCA [12]	43.2	77.2	**89.4**	96.4	93.50	**16**	45.40	66.8	86.2	82.47
Improved SCA	**45.4**	**80**	88.8	**97.2**	**93.92**	15.2	**48.40**	**69**	**87.20**	**83.20**

5.3. Re-Identification by CWBTF

Goal. The goal of this experiment is to analyze the performance of our CWBTF approach for re-identification frame while maintaining the same settings for all experiments.

Compared Methods. We compare our approach with several alternatives, which fall into two categories: (i) direct methods including SDALF [10], CPS [11] and SCA [12]; (ii) transform learning-based methods including Implicit Camera Transfer (ICT) [35], Explicit Camera Transfer (ECT) [35], Mean Brightness Transfer (MBTF) [1],Cumulative Brightness Transfer Function (CBTF) [3], Weighted Brightness Transfer Function (WBTF) [4], CWBTF with CPS [50], and WFS [37].

Implementation Details. We use publicly available codes for ICT and ECT and tested on our experimented datasets. We also implemented MBTF, CBTF, WBTF and CWBTF baselines under the same experimental settings as mentioned earlier to have a fair comparison with our proposed method.

Results. Table 3 and Table 4 shows the results for the VIPeR and CAVIAR4REID datasets, whereas Figures 6 and 7 show the performance on SAIVT-SoftBio and PRID 2011 datasets, respectively. We can make the following observations from these reported results:

(i) Among other alternatives, ICT [35] performs worst while the number of the training samples for mapping the transfer function is lower compared to their corresponding test sets, unlike CWBTF, where all the experiments have been carried out for fewer training samples as well as for the same setup for fair comparisons.

(ii) Since the VIPeR dataset is a single-shot dataset, our proposed CWBTF method cannot show its dominance like other multi-shot datasets. From the experimental results of the VIPeR dataset, it can easily be seen that unreliable single-detection-based BTFs did not perform like multiple-detection-based BTFs. Still, rank-1 of CWBTF is marginally outperforming the state of the art.

(iii) For the CAVIAR4REID dataset, transform learning methods have less effect on the performance since there is not much illumination variation between the cameras. Nevertheless, the brightness appearance transfer methods proposed in WBTF [4], CBTF [3] and MBTF [1] work better than ICT [35], while considering a smaller number of training samples, which is actually more realistic for ReID. Moreover, using multiple detection for learning brightness transfer, as in our CWBTF approach, outperforms all the single-detection-based state-of-the-art methods. The experimental results reported in the Table 4 support all our claims discussed above. So far, the performance of WFS [37] beats our methods as well as all the other considered state-of-the-art methods by a good margin. The possible reasons include the fact that WFS uses high-dimensional features, using color as well as texture-based features, while we use low-dimensional features using only color-based features. Again, for the CAVIAR4REID dataset, applying transform learning methods has less effect on the performance, since there is not much illumination variation across the cameras. So, the use of such texture-based features along with color-based features has resulted in relatively high re-identification performance for this dataset.

(iv) As mentioned earlier, the images in PRID 2011 are captured from outdoor environments and there exists a significant amount of illumination variation between the cameras. Thus, applying brightness transfer methods has a significant effect on the performance reported in Figure 7. The availability of multiple shots for the proposed CWBTF improves the re-identification accuracy, further outperforming all the state-of-the-art methods.

(v) Images in the SAIVT-SoftBio dataset are taken from an indoor environment and are also characterized by varying illumination across cameras, like PRID 2011. Consequently, adopting brightness transfer to map the illumination from one camera to another directly influences the re-identification accuracy, irrespective of the indoor or outdoor condition. The results reported in Figure 6 prove this assertion.

The authors of WFS [37] neither report the results for PRID 2011 and SAIVT-SoftBio datasets nor make the code online-available for implementing the WFS method.
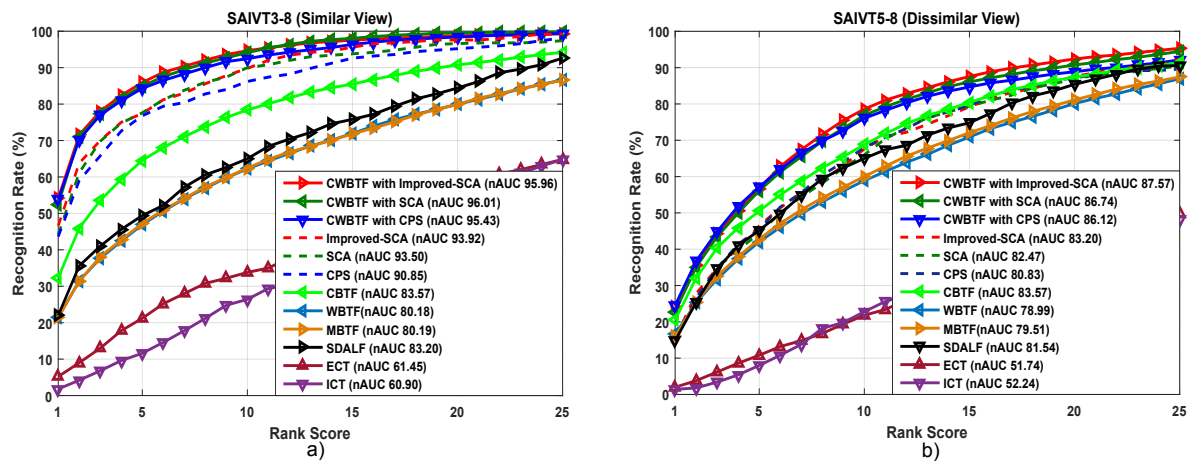
Figure 6. CMC and nAUC for the SAIVT-SoftBio dataset. Plots (**a**,**b**) show the performance of different methods compared with our CWBTF approach.
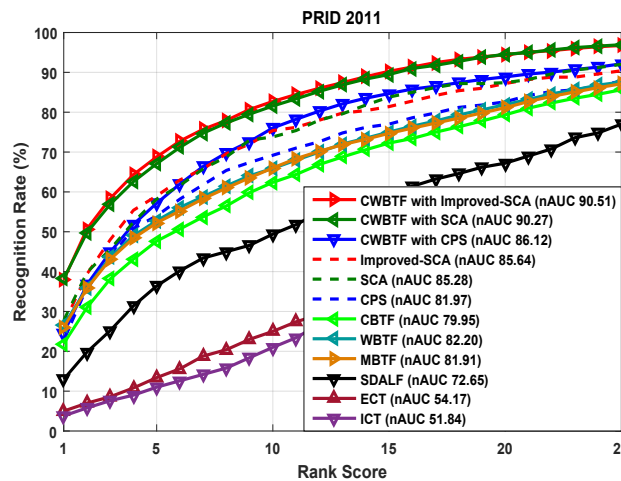

Figure 7. CMC and nAUC for the PRID 2011 dataset that shows the performance of different methods compared with our CWBTF approach.

Table 3. Experiment on the VIPeR dataset reported by top-ranked matching rate (%) and normalized area under curve (nAUC) (%) for different methods compared with our CWBTF approaches.

Methods	r = 1	r = 5	r = 10	r = 15	r = 20	nAUC
ICT [35]	14.4	42.5	60	71.3	76.5	95.3
ECT [35]	16.3	43.2	61.2	70.5	75.7	95.2
MBTF [1]	15.24	44.37	63.28	70.25	74.25	94.8
WBTF [4]	16.17	44.35	61.5	70.75	74.75	94.95
SDALF [10]	21.34	42.73	63.23	70.8	74	93.15
CPS [11]	25.84	43.7	65.51	72	76.2	94.23
SCA [12]	23.24	42.64	63.85	71.8	76	94.4
CBTF [3]	25.32	44.3	67.28	71.28	77.32	94.23
CWBTF with CPS [50]	26.15	45.12	68.31	72.15	77.12	95.1
CBTF with SCA [3]	24.12	43.12	67.2	72.3	77	95.2
CWBTF with SCA [50]	24.82	43.62	67.9	72.5	77	95.4
WFS [37]	25.81	-	**69.86**	-	**83.67**	-
Improved SCA	31.42	45.3	68.21	71.2	75.8	95.35
CBTF with improved SCA [3]	31.8	**47.2**	67.21	72.5	78.3	95.7
CWBTF with improved SCA	**32.14**	46.25	69.21	**73.28**	79.6	**96.18**

Table 4. Experiment on the CAVIAR4REID dataset reported by top-ranked matching rate (%) and normalized area under curve (nAUC) (%) for different methods compared with our CWBTF approaches.

Methods	r = 1	r = 5	r = 10	r = 15	r = 20	nAUC
ICT [35]	8	32.75	52.25	65.25	77	71.19
ECT [35]	9.5	40	60	73.25	83.25	75.19
MBTF [1]	14.25	44	63.5	76.75	83.75	77.56
WBTF [4]	16.25	45.5	65.5	76.75	83.75	77.79
SDALF [10]	11.5	38.5	62	74.5	83	73.80
CPS [11]	20.25	53	71	83. 25	89.25	82.01
SCA [12]	22.75	**59.25**	71.5	81.75	89.25	82.63
CBTF [3]	17.82	44	62	75.33	84.88	77.79
CWBTF with CPS [50]	21.75	55.10	74	85.13	91.63	83.28
CBTF with SCA [3]	23.10	**59.5**	74.33	83.5	90.25	82.90
CWBTF with SCA [50]	23.15	59.05	75.13	84.27	90.52	83.82
WFS [37]	**32.4**	**89**	**92**	**96**	**98**	83.61
Improved SCA	23.95	54.75	73.75	86	91.5	83.10
CBTF with improved SCA [3]	**24.75**	56.2	74.75	85	90.5	83.61
CWBTF with improved SCA	24.25	58	**76.05**	**86**	**91.97**	**84.11**

5.4. Effect of Feature Representation: Analysis with WHOS Feature

Goal. The objective of this experiment is to verify the effectiveness of our CWBTF approach by changing the feature representation. Specifically, our goal is to show the performance of the proposed method by replacing used features with weighted histograms of overlapping stripes (WHOS) feature representation [53]. Ideally, we would expect similar performance improvement by our method, irrespective of the feature used to represent each person in a network of cameras.

Implementation Details. We use the publicly available code of WHOS to test the performance and set the parameters the same as recommended in the published work. Except the change in features, we followed the same experimental settings while comparing with other methods.

Results. Figure 8 shows results for the PRID 2011 and SAIVT-SoftBio datasets. From Figure 8, the following observations can be made: (i) our approach outperforms all compared methods, which suggests that the proposed CWBTF technique works significantly well irrespective of the feature used to represent persons in a camera network; (ii) among the alternatives, CBTF is the most competitive. However, the gap is still significant compared to CWBTF; (iii) the improvement over CBTF shows that the proposed brightness transfer function is very effective in transferring brightness across the camera, irrespective of the feature representation.
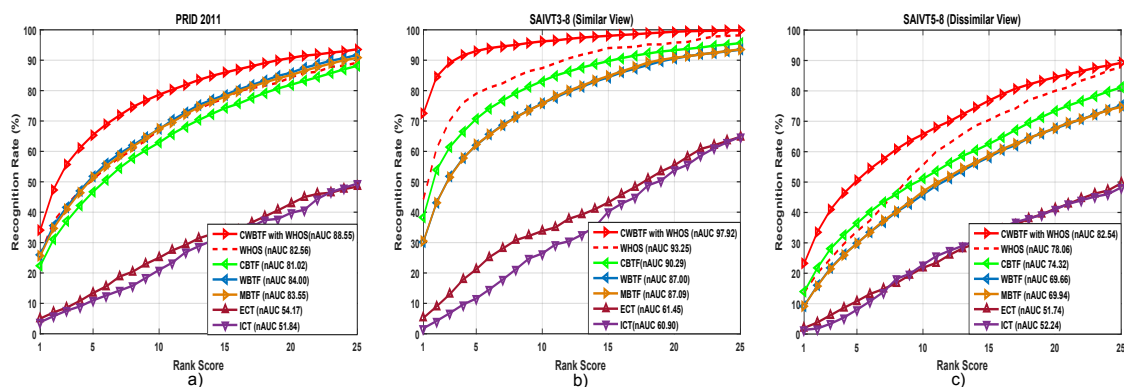
Figure 8. Re-identification performance with weighted histograms of overlapping stripes (WHOS) feature representation. Plots (**a**–**c**) show CMC curves and nAUC for PRID 2011, and similar and dissimilar views of the SAIVT-SoftBio dataset, respectively.
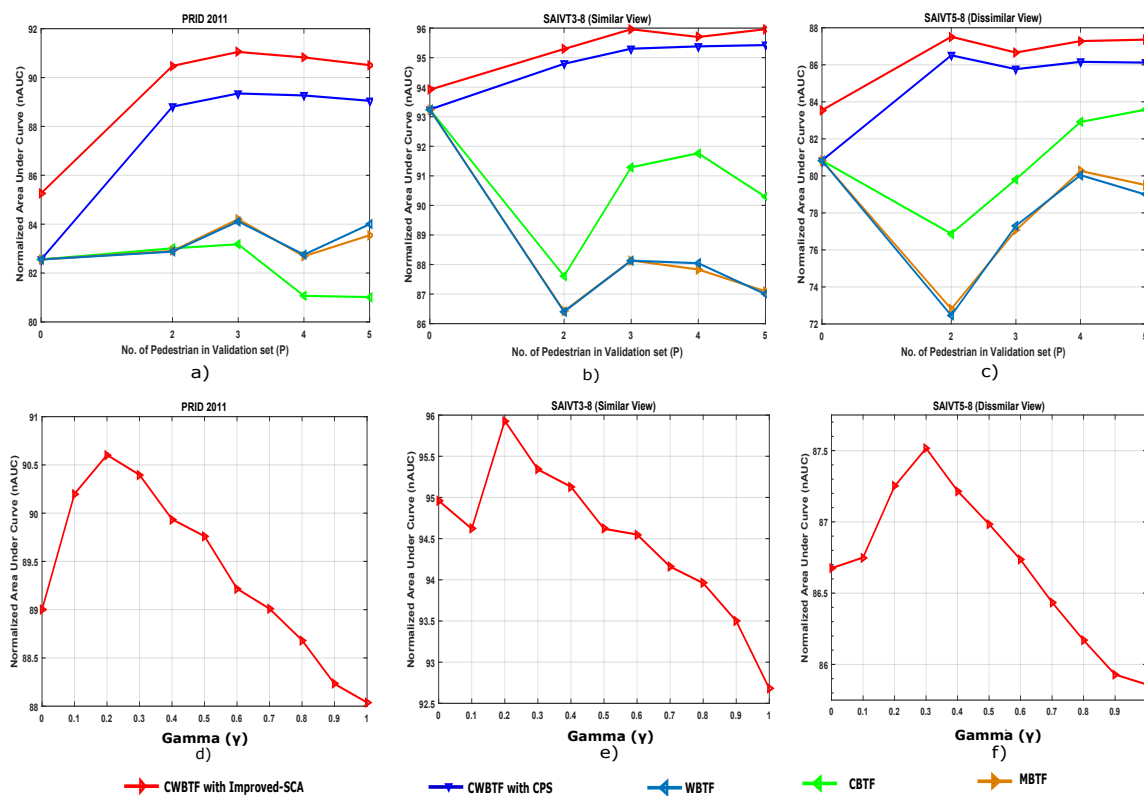
5.5. Effect of Parameter Variations

Goal. The main objective of this experiment is to analyze the performance of CWBTF by varying the parameters γ and the number of pedestrians in the validation set, P (i.e., γ and P in Equation (6)) . In the ideal case, we would like to see a minor change in performance with the change of parameters.

Implementation Details. We tested our approach for $P = [0, 2, 3, 4, 5]$ pedestrians 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 the framework becomes the standard ReID framework. We also tested our approach by varying the γ between 0 and 1.

Results. Figure 9 shows results for PRID 2011, and similar and dissimilar views of SAIVT-SoftBio. From these figures, the following observations can be made: (i) our approach outperforms all compared methods, which suggests that the proposed CWBTF technique works consistently well irrespective of the number of pedestrians P in the validation set; (ii) the variations of nAUC against γ are consistent for all datasets as shown in Figure 9d–f; (iii) for all the datasets, the proposed approach achieves best results when γ varies around 0.2 to 0.3.

Figure 9. Re-identification performance with parameter variations. Plots (**a**–**c**) show variations of nAUC with P on PRID 2011, and similar and dissimilar views of SAIVT-SoftBio, respectively. Plots (**d**–**f**) show variations of nAUC with γ on PRID 2011, and similar and dissimilar views of SAIVT-SoftBio, respectively.

6. Conclusions

This paper proposes the use of cumulative histograms from multiple detections to learn better and more robust transfer functions. Firstly, we proposed a robust brightness transfer on multiple detections to map the appearance from one camera to another multiple detection, which makes our approach quite effective and efficient. As a second contribution, we exploited the segmentation provided by stel component analysis using a large number of parts for a robust re-identification framework. 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 a few frames. Our results clearly demonstrate a significant improvement over previous ways to model appearance variations. Moreover, our proposed approach is general and can be applied to model appearance variation problems beyond person re-identification.

Author Contributions: All the authors made significant contributions to this work. Alessandro Perina and Amran Bhuiyan conceived and designed the experiment as well as wrote the paper. Amran Bhuiyan performed the described experiments. Alessandro Perina substantively revised the work and contributed the materials. Vittorio Murino is responsible for conception, revision and supervision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Javed, O.; Shafique, K.; Rasheed, Z.; Shah, M. Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views. *Comput. Vis. Image Underst.* **2008**, *109*, 146–162.
2. Porikli, F. Inter-camera color calibration by correlation model function. In Proceedings of the International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 14–17 September 2003.
3. Prosser, B.J.; Gong, S.; Xiang, T. Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions. In Proceedings of the British Machine Conference, New York, NY, USA , 1–4 September 2008.
4. Datta, A.; Brown, L.M.; Feris, R.; Pankanti, S. Appearance modeling for person re-identification using weighted brightness transfer functions. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012.
5. Chen, X.; Bhanu, B. Soft Biometrics Integrated Multi-target Tracking. In Proceedings of the 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014.
6. Bird, N.D.; Masoud, O.; Papanikolopoulos, N.P.; Isaacs, A. Detection of loitering individuals in public transportation areas. In *IEEE Trans. Intell. Trans. Syst.* **2005**, *6*, 167–177.
7. Gheissari, N.; Sebastian, T.B.; Hartley, R. Person reidentification using spatiotemporal appearance. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006.
8. Wang, X.; Doretto, G.; Sebastian, T.; Rittscher, J.; Tu, P. Shape and appearance context modeling. In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.
9. Hamdoun, O.; Moutarde, F.; Stanciulescu, B.; Steux, B. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In Proceedings of the Second ACM/IEEE International Conference on Distributed Smart Cameras, Stanford, CA, USA, 7–11 September 2008.
10. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
11. Cheng, D.S.; Cristani, M.; Stoppa, M.; Bazzani, L.; Murino, V. Custom pictorial structures for re-identification. In Proceedings of the 22nd British Machine Vision Conference, Dundee, Scotland, 29 August–2 September 2011.
12. Bhuiyan, A.; Perina, A.; Murino, V. Person re-identification by discriminatively selecting parts and features. In Proceedings of the Computer Vision—ECCV 2014 Workshops, Zurich, Switzerland, 6–7 September 2014.
13. Corvee, E.; Bremond, F.; Thonnat, M. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1713–1727.
14. Corvee, E.; Bremond, F.; Thonnat, M. Person re-identification using spatial covariance regions of human body parts. In Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, MA, USA, 29 August–1 September 2010.
15. Bak, S.; Corvee, E.; Bremond, F.; Thonnat, M. Multiple-shot human re-identification by mean riemannian covariance grid. In Proceedings of the 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Klagenfurt, Austria, 30 August–2 September 2011.
16. Wu, Y.; Minoh, M.; Mukunoki, M.; Li, W.; Lao, S. Collaborative sparse approximation for multiple-shot across-camera person re-identification. In Proceedings of the Collaborative Sparse Approximation for Multiple-Shot Across-Camera Person Re-Identification, Beijing, China, 18–21 September 2012.

17. Kviatkovsky, I.; Adam, A.; Rivlin, E. Color invariants for person reidentification.*IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1622–1634.

18. Gray, D.; Tao, H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008.

19. Schwartz, W.R.; Davis, L.S. Learning discriminative appearance-based models using partial least squares. In Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image, Rio de Janiero, Brazil, 11–15 October 2009.

20. Mignon, A.; Jurie, F. Pcca: A new approach for distance learning from sparse pairwise constraints. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.

21. Hirzer, M.; Roth, P.M.; Köstinger, M.; Bischof, H. Relaxed pairwise learned metric for person re-identification. In Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2014.

22. Li, Z.; Chang, S.; Liang, F.; Huang, T.S.; Cao, L.; Smith, J.R. Learning locally-adaptive decision functions for person verification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.

23. Zheng, W.-S.; Gong, S.; Xiang, T. Reidentification by relative distance comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 635–668.

24. Köstinger, M.; Hirzer, M.; Wohlhart, P.; Roth, Peter M.; Bischof, H. Large scale metric learning from equivalence constraints. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.

25. Pedagadi, S.; Orwell, J.; Velastin, S.; Boghossian, B. Local fisher discriminant analysis for pedestrian re-identification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Portland, OR, USA, 23–28 June 2013.

26. Zheng, W.-S.; Gong, S.; Xiang, T. Person re-identification by probabilistic relative distance comparison. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011.

27. Xiong, F.; Gou, M.; Camps, O.; Sznaier, M. Person re-identification using kernel-based metric learning methods. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.

28. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Deep Metric Learning for Person Re-identification. In Proceedings of the 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014.

29. Cheng, D.; Gong, Y.; Zhou, S.; Wang, J.; Zheng, N. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

30. Varior, R.R.; Haloi, M.; Wang, G. Gated siamese convolutional neural network architecture for human re-identification. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.

31. Ahmed, E.; Jones, M.; Marks, T.K. An improved deep learning architecture for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

32. Xiao, T.; Li, H.; Ouyang, W.; Wang, X. Learning deep feature representations with domain guided dropout for person re-identification. *arXiv* **2016**, arXiv:1604.07528.

33. Varior, R.R.; Shuai, B.; Lu, J.; Xu, D.; Wang, G. A siamese long short-term memory architecture for human re-identification. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.

34. Yan, Y.; Ni, B.; Song, Z.; Ma, C.; Yan, Y.; Yang, X. Person Re-identification via Recurrent Feature Aggregation. In Proceedings of the 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.

35. Avraham, T.; Gurvich, I.; Lindenbaum, M.; Markovitch, S. Learning implicit transfer for person re-identification. In Proceedings of the Computer Vision, Florence, Italy, 7–13 October 2012.

36. Yulia, B.; Tamar, A.; Michael, L. Transitive Re-Identification. In Proceedings of the British Machine Vision Conference, Bristol, UK, 9–13 September 2013.

37. Martinel, N.; Das, A.; Micheloni, C.; Roy-Chowdhury, A.K. Re-identification in the function space of feature warps. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1656–1669.

38. Jojic, N.; Perina, A.; Cristani, M.; Murino, V.; Frey, B. Stel component analysis: Modeling spatial correlations in image class structur. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.

39. Bazzani, L.; Cristani, M.; Murino, V. Symmetry-driven accumulation of local features for human characterization and re-identification. *Comput. Vis. Image Underst.* **2013**, *117*, 130–144.

40. Forssén, P.E. Maximally stable colour regions for recognition and matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007.

41. Cheng, D.S.; Cristani, M. Person re-identification by articulated appearance matching. In *Person Re-Identification*; Springer: Berlin/Heidelberg, Germany, 2014.

42. Hirzer, M.; Beleznai, C.; Roth, P.M.; Bischof, H. Person re-identification by descriptive and discriminative classification. In Proceedings of the 17th Scandinavian Conference, Ystad, Sweden, May 2011.

43. Bialkowski, A.; Denman, S.; Sridharan, S.; Fookes, C.; Lucey, P. A database for person re-identification in multi-camera surveillance networks. In Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, Australia, 3–5 December 2012.

44. Su, C.; Yang, F.; Zhang, S.; Tian, Q.; Davis, L.S.; Gao, W. Multi-task learning with low rank attribute embedding for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.

45. Cancela, B.; Hospedales, T.M.; Gong, S. Open-world person re-identification by multi-label assignment inference. In Proceedings of the British Machine Vision Conference (BMVC), Nottingham, UK, 1–5 September 2014.

46. Karanam, S.; Li, Y.; Radke, R.J. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.

47. Das, A.; Chakraborty, A.; Roy-Chowdhury, A.K. Consistent re-identification in a camera network. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.

48. Kodirov, E.; Xiang, T.; Fu, Z.; Gong, S. Person Re-Identification by Unsupervised ℓ_1 Graph Learning. In Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016.

49. Bhuiyan, A.; Mirmahboub, B.; Perina, A.; Murino, V. Person Re-identification Using Robust Brightness Transfer Functions Based on Multiple Detections. In Proceedings of the 18th International Conference, Genoa, Italy, 7–11 September 2015.

50. Bhuiyan, A.; Perina, A.; Murino, V. Exploiting multiple detections to learn robust brightness transfer functions in re-identification systems. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec, QC, Canada, 27–30 September 2015.

51. Mirmahboub, B.; Kiani, H.; Bhuiyan, A.; Perina, A.; Zhang, B.; Del Bue, A.; Murino, V. Person re-identification using sparse representation with manifold constraints. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2015.

52. Panda, R.; Bhuiyan, A.; Murino, V.; Roy-Chowdhury, A.K. Unsupervised Adaptive Re-identification in Open World Dynamic Camera Networks. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2015.

53. Lisanti, G.; Masi, I.; Bagdanov, A.D.; Del Bimbo, A. Person re-identification by iterative re-weighted sparse ranking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1629–1642.