2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Distributed and Unsupervised Cost-Driven Person Re-identification

Niki Martinel
University of Udine
Via Scienze 208, 33100 Udine, Italy
niki.martinel@uniud.it

Gian Luca Foresti
University of Udine
Via Scienze 208, 33100 Udine, Italy
gianluca.foresti@uniud.it

Christian Micheloni
University of Udine
Via Scienze 208, 33100 Udine, Italy
christian.micheloni@uniud.it

***Abstract*—The problem of re-identify persons across single disjoint camera-pairs has received great attention from the community. Despite this, when the re-identification process has to be carried out on a large camera network a different approach has to be considered. In particular, existing approaches have neglected the importance of the network topology (i.e., the structure of the monitored environment) in such a process. To try filling such a gap, we propose a Distributed and Unsupervised Cost-Driven Person Re-Identification framework (DUPRe) which introduces the following contributions: (i) a camera matching cost to measure the re-identification performance between nodes of the network; (ii) a derivation of the distance vector algorithm which allows to learn the network topology hence to prioritize and limit the cameras inquired for the re-identification. Results on two benchmark datasets show that our solution brings to significant network-wise re-identification improvements.**

I. Introduction

In a wide area scene analysis set-up, different issues arise and deny the deployment of camera networks that provide full area coverage. To obtain information across the uncovered areas, new inter-camera data association approaches which allows long term target tracking are needed. When the target is a person, the solution can be achieved by means of person re-identification approaches [1].

The task of re-identify persons in a wide area camera network presents different issues like illumination variations, scale, viewpoint and pose changes. To address such challenges, existing approaches have followed three main lines of research.

Discriminative signature based methods: Clothing patches [2], dense region-based structures [3] and their spatial relationships [4] were explored to extract different color, texture and shape features. Recently, multiple frame analysis techniques [5], [6], coupled dictionaries exploiting labeled and unlabeled data [7] and sparse discriminative classifiers ensuring that the best candidates are ranked at each iteration [8] were proposed.

Feature transformation based methods: A first work in such direction learned a brightness transfer function [9] to transform features between camera dependent spaces. Later, the idea that the feature transformation is not unique was investigated in [10]. Poses and viewpoint changes were also exploited in [11] to select the optimal classifier.

Metric learning based methods: Several methods were proposed to learn a Mahalanobis metric [12] often relying on equivalence constraints [13]. Metric ensembles [14], transfer learning set ups [15] and re-ranking solutions [16] were also investigated. Recently, works focused on local distance comparison problems [17], similarity metric learning and listwise similarities [18].

Almost all such methods consider the problem of re-identify a person between a single camera pair only. When used in a large camera network they require the exchange of a huge amount of data (person signatures) and may also lead to many false matches. In addition, such works are neglecting the information brought in by the network topology and the environment structure which can be exploited to improve the re-identification. In this direction, the paths of the people inside the monitored area generate different correlation degrees between camera pairs (e.g., cameras with similar vs different views). We think that exploiting such correlation degrees for person re-identification can increase the performance.

With respect to all existing methods, the closest ones to our approach have been proposed in [19], [20], [21]. Though sharing the network-wise re-identification idea, [19] differs from our work since it assumes that network topology can be learned using spatio-temporal constraints. In [20], the goal is to provide a feature selection method that minimizes the data needed to represent the appearance of objects. In [21], authors aim to maintain re-identification consistency across the network. Contrarily, our approach focuses on the selection of the optimal subset of cameras to inquire.

Specifically, we propose a framework that exploits the principles of the distance vector routing algorithm to perform the re-identification in a distributed fashion. In particular, it introduces: (i) a camera-pair matching cost as a measure of the correlation degree; (ii) a derivation of the distance vector routing algorithm that exploits the camera-pair matching cost to learn the network topology in an unsupervised fashion. These contributions allow us to prioritize and limit the cameras inquired for person matching. In addition, the update rule of the proposed distance vector routing algorithm leads to network adaptation over time.

II. The DUPRe Framework

As shown in Fig.1, the DUPRe framework is organized in two stages.

During the unsupervised learning phase, gallery signatures form a given camera are computed. Such signatures are matched with the gallery set of a second camera in the network. The resulting dissimilarity scores are used to compute the C2C matching cost that represents the correlation degree

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

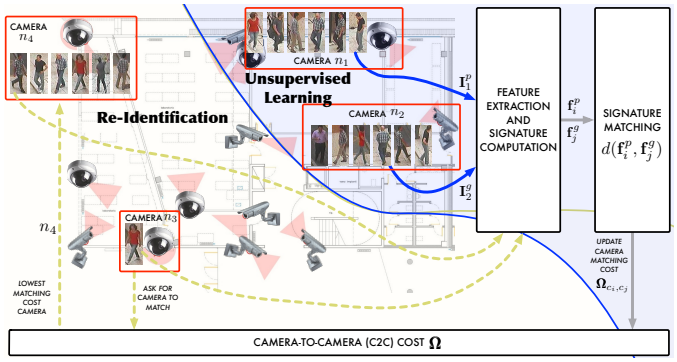


Fig. 1. Proposed DUPRe framework architecture. During the re-identification process, cameras of the network are iteratively inquired in a priority fashion by exploiting the camera-to-camera (C2C) cost which has been previously obtained through the unsupervised learning phase. Cameras with the lowest C2C cost are first inquired. If the returned match is no reliable (high dissimilarity score), cameras with higher C2C costs are iteratively considered until a maximum allowable matching cost have been reached.

between such a camera pair. The process is repeated for every camera pair.

During the re-identification phase, a probe camera computes the person signature. Then, the distance vector routing algorithm exploits the C2C cost to select the cameras that have higher probability to return a correct match, i.e., cameras that have a C2C cost with the probe camera within a predefined value. Such cameras receive the probe signature and compute the dissimilarities with their gallery sets. The resulting dissimilarity scores are returned to the probe cameras. This applies a network-consensus scheme to identify the correct match. Correct matches are then considered to update the C2C costs which are finally propagated to other cameras by exploiting the distance vector algorithm.

A. Preliminaries

Camera Network: $\mathcal{N} = \{n_1, n_2, \cdots, n_m\}$ is the set of m cameras in the network.
Camera Pairs: Given a probe camera n_i, the set of camera pairs is $\mathcal{N}_i = \{(n_i, n_j) | n_j \in \mathcal{N}\}$, where $|\mathcal{N}_i| = m$.
Gallery Set: $\mathcal{G}_i$ denotes the gallery set of persons that have been acquired by the camera n_i.
Person Signature: Let $\mathbf{I}_i^p$ be an image of person p acquired in camera n_i, then $\mathbf{f}_i^p = [\mathbf{x}_1^p, \mathbf{x}_2^p, \cdots, \mathbf{x}_s^p]^T$ is the corresponding signature obtained by concatenating s different features (e.g. color histogram, texture features, etc.). If a sequence of images is available, we assume that each $\mathbf{x}_j^p$ is the accumulation of the j-th feature over the N frames.
Signature Matching: $d\left(\mathbf{f}_i^p, \mathbf{f}_j^g\right) \in [0, 1]$ is the dissimilarity between signatures $\mathbf{f}_i^p$ and $\mathbf{f}_j^g$.
Camera Pair Dependent Threshold: δ_{n_i,n_j} is a threshold used to decide whether two signatures are from a same person.
C2C Cost: $\mathbf{\Omega} \in \mathbb{Z}^{m \times m}$ denotes the matching cost matrix. Each entry $\mathbf{\Omega}_{n_i,n_j} \in \mathbb{Z}$ is the C2C cost between n_i and n_j.

B. Unsupervised Camera-to-Camera Cost Learning

To learn the C2C matching cost for each camera pair (n_i, n_j), the dissimilarity matrix $\mathbf{\Delta}_{n_i,n_j}$ is computed as

$$\Delta_{n_i,n_j}(p, g) = d\left(\mathbf{f}_i^p, \mathbf{f}_j^g\right) \tag{1}$$

where $p = 1, \ldots, |\mathcal{G}_i|$ and $g = 1, \ldots, |\mathcal{G}_j|$.

Then, we exploit the distribution of the values within the dissimilarity matrix to learn δ_{n_i,n_j}. The idea is that such a distribution should present two modes associated with the dissimilarities computed between correct and false matches, respectively. The desired threshold δ_{n_i,n_j} is the one that optimally separates such two modes. This is computed using the histogram entropy-based method in [22].

Armed with the camera pair threshold, we first compute

$$\tilde{\Delta}_{n_i,n_j}(p, g) = \begin{cases} 0, & \text{if } \Delta_{n_i,n_j}(p, g)) > \delta_{n_i,n_j} \\ 1, & \text{else} \end{cases} \tag{2}$$

then, by summing the number of matches for each person p acquired by n_i we obtain the C2C cost as

$$\mathbf{\Omega}_{n_i,n_j} = \sum_{p=1}^{|\mathcal{G}_i|} \begin{cases} -1, & \text{if } \max(\tilde{\Delta}_{n_i,n_j}(p, :)) > 0 \\ +1, & \text{else} \end{cases} \tag{3}$$

where the $\tilde{\Delta}_{c_i,c_j}(p, :)$ denotes the p-th row of matrix $\tilde{\mathbf{\Delta}}_{c_i,c_j}$.

To conclude, it is important to emphasize that, since the approach is unsupervised, wrong matches might also be considered to reduce in the C2C cost. Despite this may limit the performance of our approach, it does not require an *a-priori* manual labeling of a set of persons identities. Since we consider large camera network set-ups, such a feature is very relevant.

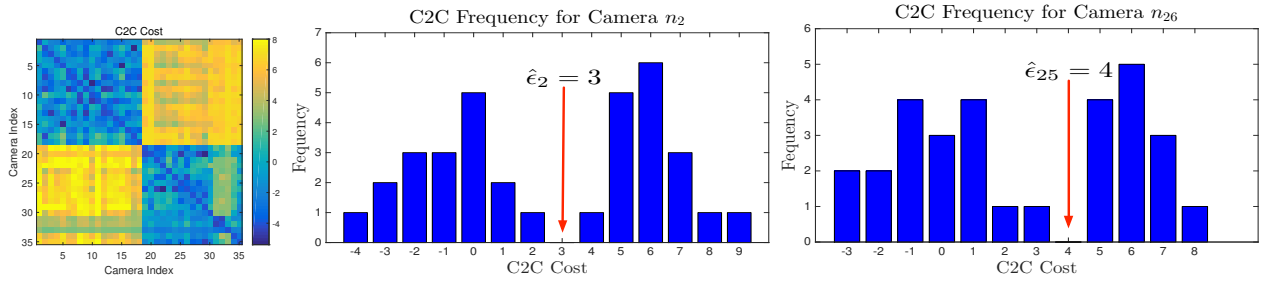
C. Distributed Re-Identification via Network-Wise Consensus

To achieve distributed network-wise re-identification we introduce a variation of the distance vector routing algorithm. Provided that each node of the network can transmit data only via its directly connected neighbors, for every of them, the standard distance vector routing algorithm constructs a routing table specifying the *cost* to reach every *destination*. It also stores a *next-hop* entry indicating the shorter outgoing line to reach a specific *destination*.

Each node knows the *cost* required to reach a neighbor. When a node has to transmit data, it sends to the *next-hop* node a packet containing the data itself and its personal routing table. The node receiving such packet updates its own routing table by fusing its current information with the routing table of the sender. After few updates, each node is able to determine the lowest-cost route to every destination.

We propose to exploit the distance vector routing algorithm to update the C2C cost. Specifically, we substitute the routing table of each camera n_i with the C2C cost table having entries given by $(n_j, \mathbf{\Omega}_{n_i,n_j})$ pairs. Once a match is completed, the C2C cost table is fused with the C2C table received by the inquiring camera.

In the re-identification phase, the C2C cost table of a given camera n_i is used to prioritize and limit the number of the inquired cameras. Priority is given by the increasing C2C cost. The number of queries is limited by the camera-dependent threshold $\hat{\epsilon}_i$ over the C2C costs, e.g. n_j is queried if and only if $\mathbf{\Omega}_{n_i,n_j} < \hat{\epsilon}_i$. The set of all the inquired cameras is denoted as $\mathcal{N}_i^*$. The matching process is completed when, after C2C table



(a) (b) (c)

Fig. 2. Camera network C2C cost example for the DANA36 dataset. In (a) the camera network C2C cost $\mathbf{\Omega}$ is shown. In the color coded plot, yellow values mean high cost, while blue values mean low cost. In (b) and (c) the distributions of the C2C costs of cameras c_3 and c_{23} are shown together with the computed C2C thresholds $\hat{\epsilon}_3$ and $\hat{\epsilon}_{23}$, respectively.

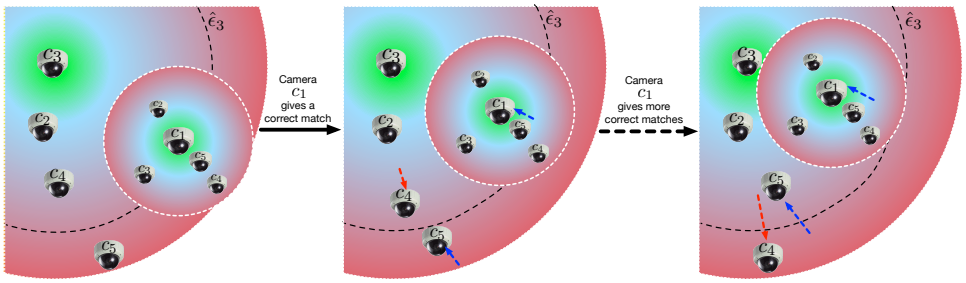


Fig. 3. Consequences of the C2C update. Let consider the C2C cost of camera n_3. Let also assume that due to image noise, misdetections, and false matches, camera n_5, that has a similar view to camera n_3, hence should well match it, is associated with a high $\mathbf{\Omega}_{n_3,n_5}$. The camera-dependent threshold $\hat{\epsilon}_3$ excludes camera n_5 from the set $\mathcal{N}_3^*$ of the inquired cameras. Camera n_1 is in such a set and it has a low $\mathbf{\Omega}_{n_1,n_5}$. After few re-identification iterations, camera n_1 returns many correct matches. As a result of the proposed update rule, camera n_5 correctly moves within the set of inquired cameras $\mathcal{N}_3^*$.

update, there is no already inquired camera in the network satisfying the above condition.

To learn $\hat{\epsilon}_i$, we follow a similar scheme to the one used to obtain δ_{n_i,n_j}. Let consider Fig. 2 showing the C2C costs (i.e., $\mathbf{\Omega}$) and the corresponding distribution computed for the DANA36 dataset. Results show the presence of two distinct modes. We exploit such modes and identify $\hat{\epsilon}_i$ by means of the entropy-based thresholding method [22].

The dissimilarities between a probe p and any gallery g, belonging to an inquired camera n_j, are computed as $d(\mathbf{f}_i^p, \mathbf{f}_j^g)$. Such dissimilarities are then filtered to obtain the set of matching persons $\mathcal{G}_j^* = \{g | d(\mathbf{f}_i^p, \mathbf{f}_j^g) < \delta_{n_i,n_j}\}$.

When the matching process is completed, re-identification consensus is obtained as follows. Let $\hat{\mathcal{G}} = \bigcup_j \mathcal{G}_j^*$ be the set of unique person identities that matched probe p. Then, for each $g \in \hat{\mathcal{G}}$, the consensus dissimilarity is computed as

$$D(p,g) = \prod_{n_j \in \mathcal{N}_i^*} \begin{cases} d(\mathbf{f}_i^p, \mathbf{f}_j^g), & \text{if } g \in \mathcal{G}_j^* \\ 1, & \text{else} \end{cases} . \tag{4}$$

The person g having the lowest dissimilarity score $D(p,g)$ represents the re-identified person.

Once the person has been re-identified via consensus, the C2C cost of the probe camera is updated. Then, as a router learns the costs from its neighbors, the current probe camera n_i learns from a matching camera n_j. In particular, the C2C cost between the probe and each inquired camera $c_j \in \mathcal{C}_i^*$ is updated as

$$\mathbf{\Omega}^*_{c_i,c_j} = \mathbf{\Omega}_{c_i,c_j} \begin{cases} -1 & \text{if } g \in \mathcal{G}_j^* \\ +1, & \text{else} \end{cases} . \tag{5}$$

Then, each of the remaining $c_l \in \{\mathcal{C} \setminus \mathcal{C}_i^*\}$ entries of the C2C cost table $\mathbf{\Omega}_{c_i}$ is updated as

$$\mathbf{\Omega}^*_{c_i,c_l} = \min\left(\mathbf{\Omega}_{c_i,c_l}, \min_{c_j \in \mathcal{C}_i^*} \left(\mathbf{\Omega}^*_{c_i,c_j} + \mathbf{\Omega}_{c_j,c_l}\right)\right). \tag{6}$$

D. Discussion

By exploiting the C2C cost as well as the proposed variation of the distance vector routing algorithm, the DUPRe framework introduces the following features.

- The C2C cost, i.e. the camera pair correlation degree, provide information regarding the structure of the monitored environment. This is considered to inquire cameras of the network in a *lower costs-higher priority* fashion.
- Unsupervised identification of the C2C costs threshold allow to dynamically limit the number of inquired cameras for the network-wise re-identification. This results in a limited network bandwidth usage.
- Fusing dissimilarities via network consensus provides a more robust re-identification.
- Exploiting the updating rule derived from the proposed distance vector routing algorithm variation yields to network adaptation over time. Cameras that return a correct match lower the C2C costs, while non-matching cameras increase it.

(a) (b)

Fig. 4. 10 persons samples from the (a) WARD, and (b) DANA36 datasets. Each row shows the same person acquired by a different camera.

- As shown in Fig. 3, C2C costs adaptation propagates through the network via camera neighbors. As a result, a definition of a more performing set of inquired cameras is obtained and updated over time.

III. Experiments

A. Person Signature

To compute the person signature, we followed a common approach and divided each given image into 6 equally sized horizontal stripes. Color, shape and texture features are extracted from each of such stripes, then concatenated to obtain $\mathbf{f}$. If more images of a single person were available, the features of the same type were averaged over all the frames.

Color: To obtain illumination invariant properties, we have projected each stripe to the RGB, HSV, CIELab and normRGB color spaces. Then, three 1D histograms have been extracted from each color space. Histograms extracted from the HSV and RGB color spaces have been quantized using 8 bins per channel, while histograms extracted from the normRGB and CIELab color spaces have been computed considering 32 bins per channel. All such histograms have been finally concatenated in a 1440-D vector.

Shape: PHOG [23] features have been separately computed for the hue, saturation and value components of each stripe. We have considered 9 orientation bins to compute the single HoG feature vector for each of the 4 pyramid levels. The resulting PHOG vector has 74982 elements.

Texture: Haralick [24] features have been extracted from grayscale stripes. To achieve invariance to rotation we have computed the gray level co-occurrence matrices using the same symmetric offsets as in [24]. The resulting Haralick feature vector lies in a 312-D space.

B. Dissimilarity Computation

We have considered different approaches to compute the dissimilarities between image pairs, i.e., $d(\cdot,\cdot)$. We have conducted the experiments using the standard ℓ^2, χ^2 and cosine distances as well as metric learning approaches. Specifically, we have considered the KISSME [25], LADF [17] and LFDA [26] metric learning approaches.

C. Datasets

Two public benchmark datasets have been considered to evaluate the performance of the proposed approach.

WARD: The Wide Area Re-identification Dataset[1] (WARD) [27] is of particular interest because it has a huge illumination variation apart from resolution and pose changes (see Fig. 4(a) for a few samples). We have conducted the experiments for all the three cameras denote here as camera 1, 2, and 3. As performed by existing methods [10], [21], we have computed the results by considering 35 persons both for the train and for the re-identification phases.

DANA36: The DANA36 dataset consists of 23,641 images, representing 15 persons and 9 vehicles acquired by 36 cameras. 27 outdoor cameras observed the persons and vehicles, while the remaining 9 observed the same persons indoors (see Fig. 4(b) for few sample images).

While such a dataset cannot be considered as fully representative of a real scenario, it has images coming from 36 cameras. Hence, we have considered it to show that, using the proposed framework, we can improve the re-identification performance by inquiring a limited number of cameras in a large network. To highlight these advantages, we have split the dataset as follows: images of 7 out of 15 persons have been taken from camera 1 to camera 18, images for the remaining 8 persons have been taken from each of the remaining cameras. A maximum of 15 randomly selected images per person has been taken. No persons have been acquired by camera 35, thus the resulting dataset contains 3,372 images of 15 person acquired by 35 cameras. Since a training phase is required to learn the C2C cost matrix, we have split the dataset as follows: 4 out of 7 and 4 out of 8 persons have been considered in the training phase; the remaining 3+4 ones have been used in the re-identification phase.

D. Evaluation Protocol

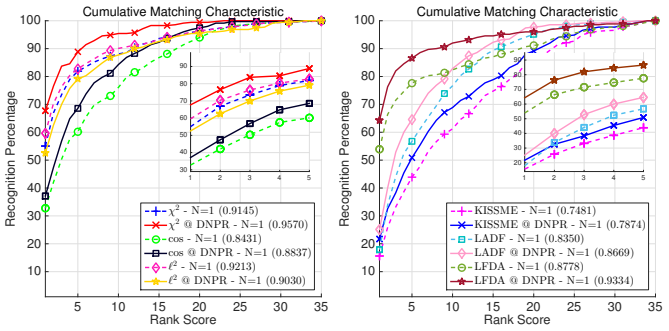
To show the benefits of our framework, we have considered two scenarios. In the former case (@DUPRe), given a camera in which a probe person is viewed, its signature is matched with each gallery set of the inquired cameras. In the latter our distribute framework is not considered and the signature is matched with the gallery set of each camera in the network. Notice that, the considered metrics, i.e., KISSME [25], LADF [17] and LFDA [26], are re-identification methods, hence the following performance analysis should be considered as a comparison with existing works too.

All the results are given using both the single-shot and the multiple-shots strategies [1] where N images have been used to compute the signatures. To fairly evaluate our method against state-of-the-art approaches, we have run the experiments using 10 randomly selected training/test set splits. We report on the averaged results over such 10 trials.

E. Performance Measures

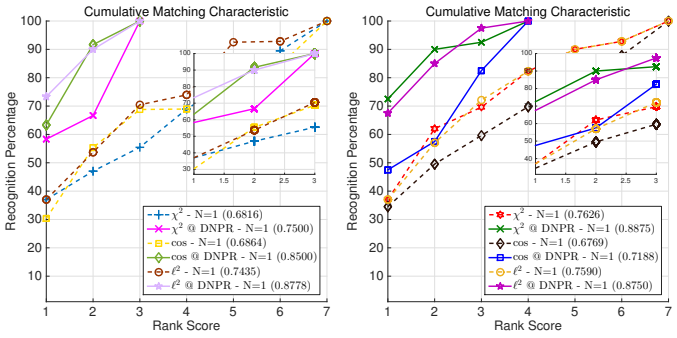
Performances are provided in terms of recognition rate by the Cumulative Matching Characteristic (CMC) curve and

[1]Available at http://users.dimi.uniud.it/~niki.martinel/code.php



(a) (b)

Fig. 5. Single shot performance on the WARD dataset. The two figures show the CMC performance for probe camera 1. The inside picture show the performance on a reduced rank range. The nAUC values for each CMC are given within brackets. In (a), CMC performances have been computed using standard dissimilarity measures. In (b) CMC performances have been computed using metric learning algorithms.



(a) (b)

Fig. 6. Single shot performance on the DANA dataset. Results are shown for probe cameras (a) 1 (outdoor) and (b) 23 (indoor).

the normalized Area Under Curve (nAUC). The CMC curve depicts the recognition performance as a function of the rank score and represents the expectation of finding the correct match in the top-k ones. The nAUC value shows how well a method performs regardless of the dataset size.

F. Re-Identification Performance Analysis

For the following datasets single-shot results are shown for a single probe camera. The performances for the remaining ones are given in the supplementary.

WARD: CMC and nAUC performances achieved using the proposed framework on the WARD dataset are shown in Fig. 5 (single-shot) and Table I (multiple-shots).

In Fig. 5(a) performances obtained by considering standard metrics within the DUPRe framework are shown for probe camera 1. The depicted results show that the CMC as well as the nAUC performance improve when the DUPRe framework is applied. In particular, the rank 1 correct recognition rate achieved by using the χ^2 distance within the DUPRe framework (about 68%) improves the baseline by more than 10%. Notice that, when the Euclidean distance is considered, the performance decreases.

In Fig. 5(b) the depicted performances, obtained for the same probe camera, have been computed using the three considered baseline metric learning methods. The obtained results reflect those achieved by considering the standard metrics. The initial performance are always improved when the learned metrics are exploited by the DUPRe framework. Among all the considered metric learning algorithms, LFDA is the best performing one. It achieves the higher rank 1 correct recognition rate as well as the best overall (nAUC) performances. KISSME is the worst performing one. More interestingly, it performs worse than standard metrics. This is due to the fact that the amount of data provided by the WARD dataset (single shot) is not sufficient to reliably compute the covariance matrices exploited to learn such a metric.

Despite this, as shown in Table I, when multiple images are considered ($N > 1$), the KISSME algorithm yields to the best overall performances. In particular, considering probe camera 1, when $N = 25$ it achieves the best performance possible by recognizing each person at rank 1. A similar result, for probe camera 2, is obtained by LFDA@DUPRe when $N = 5$. Notice that, for the same probe cameras, performances of LFDA@DUPRe degrade when $N > 5$. A more detailed analysis of such result, showed that this behavior is due to the variability of the intra-camera person appearance.

DANA36: Results obtained considering the DANA36 dataset are shown in Fig. 6 (single-shot) and Table II (multiple-shots). Since there are only 8 persons in the training phase, we have not considered metric learning approaches.

Fig. 6 shows the performance of standard metrics and compare those to the ones obtained by considering such methods in the DUPRe framework. Results are depicted for probe cameras 1 (outdoor), Fig. 6(a), and 23 (indoor), Fig. 6(b).

The C2C threshold $\hat{\epsilon}_1$ forces the probe camera 1 to match the probe person with the 3 signatures acquired by inquired cameras 1 to 18. Similarly, signatures from probe camera 23 are matched with the 4 signatures acquired by inquired cameras 19 to 35. In both the cases the true match is included in the 3 and 4 gallery persons. This is the reason why, the baseline methods, considered within the proposed framework, are able to obtain a 100% correct recognition rate at rank 3 and 4, respectively. When such baselines are not exploited by DUPRe, each probe signature is matched with all the 35 cameras, hence with the 7 gallery persons. In such a case, the performance degrades and the 100% correct recognition rate is achieved at rank 7 in all the cases.

When multiple-shots are considered (see Table II), the initial nAUC performance of baseline methods are significantly improved for both the probe cameras. In particular, the best overall performances are achieved by ℓ_2@DUPRe (0.9000) and χ^2@DUPRe (0.9187) for probe camera 1 and 23, respectively. Notice that in both cases the best results are achieved when $N = 5$ is considered. The degradation of the results when $N = 15$ is due to the variability of the intra-camera person appearance. This drawback could be overcome by a better feature pooling strategy.

IV. Conclusions

In this work we have introduced a C2C cost to measure the re-identification performance between pairs of cameras. Using

TABLE I

MULTIPLE-SHOT PERFORMANCE ON THE WARD DATASET. NAUC PERFORMANCES FOR EACH OF THE THREE PROBE CAMERAS ARE SHOWN AS A FUNCTION OF THE NUMBER OF IMAGES PER PERSON (N). FIRST THREE ROWS SHOW THE RESULTS OBTAINED USING STANDARD METRICS. LAST THREE ROWS SHOW THE RESULTS ACHIEVED USING METRIC LEARNING ALGORITHMS. BEST PERFORMANCES FOR EACH (PROBE CAMERA, NUMBER OF IMAGES PER PERSON) ARE SHOWN IN BOLDFACE FONT. THE BEST OVERALL RESULT FOR EACH PROBE CAMERA IS UNDERLINED.

Probe Camera	1				2				3			
$N \rightarrow$	1	5	10	25	1	5	10	25	1	5	10	25
χ^2	0.9145	0.9303	0.9318	0.9317	0.9293	0.9390	0.9400	0.9411	0.9166	0.9383	0.9396	0.9409
χ^2@DUPRe	0.9570	0.9907	0.9885	0.9903	0.9414	0.9909	0.9900	0.9883	0.9305	0.9701	0.9856	0.9822
cos	0.8431	0.8667	0.8679	0.8695	0.8799	0.8942	0.8952	0.8970	0.8836	0.9178	0.9231	0.9240
cos@DUPRe	0.8837	0.9379	0.9324	0.9264	0.9412	0.9565	0.9585	0.94468	0.9065	0.9203	0.9152	0.9142
ℓ_2	0.9213	0.9577	0.9578	0.9578	0.9434	0.9578	0.9578	0.9578	0.8997	0.9569	0.9578	0.9578
ℓ_2@DUPRe	0.9030	0.9529	0.9650	0.9713	0.9708	0.9683	0.9915	0.9842	**0.9367**	0.9798	0.9901	0.9882
KISSME	0.7481	0.9548	0.9557	0.9562	0.6683	0.9553	0.9558	0.9560	0.6000	0.9494	0.9496	0.9511
KISSME@DUPRe	0.7874	**0.9996**	**0.9967**	**1.0000**	0.7020	0.9992	**0.9993**	**0.9994**	0.6301	**0.9971**	**0.9965**	**0.9951**
LADF	0.8350	0.8861	0.8902	0.8840	0.8627	0.9070	0.9120	0.9222	0.8216	0.8988	0.8911	0.8880
LADF@DUPRe	0.8669	0.9247	0.9091	0.9213	0.8886	0.9315	0.9435	0.9438	0.8509	0.9031	0.9035	0.9144
LFDA	0.8788	0.8874	0.8976	0.9114	0.9308	0.9411	0.9478	0.9501	0.8491	0.9440	0.9554	0.9578
LFDA@DUPRe	**0.9334**	0.9443	0.9578	0.9631	**0.9752**	**1.0000**	0.9674	0.9557	0.9058	0.9863	0.9806	0.9673

TABLE II

MULTIPLE-SHOT PERFORMANCE ON THE DANA36 DATASET. NAUC PERFORMANCES FOR THE TWO SELECTED PROBE CAMERAS (I.E., 1 AND 23) ARE SHOWN AS A FUNCTION OF THE NUMBER OF IMAGES PER PERSON (N). THE THREE ROWS SHOW THE RESULTS OBTAINED USING STANDARD METRICS. BEST PERFORMANCE FOR EACH CAMERA/NUMBER OF IMAGES PER PERSON ARE SHOWN IN BOLDFACE FONT. THE BEST OVERALL RESULT FOR EACH PROBE CAMERA IS ALSO UNDERLINED.

Camera	1			23		
$N \rightarrow$	1	5	15	1	5	15
χ^2	0.681	0.772	0.776	0.762	0.787	0.787
χ^2@DUPRe	0.750	0.777	0.750	**0.887**	**0.918**	**0.906**
cos	0.686	0.745	0.731	0.676	0.694	0.716
cos@DUPRe	0.850	0.811	0.811	0.718	0.712	0.725
ℓ_2	0.743	0.867	0.855	0.759	0.766	0.798
ℓ_2@DUPRe	**0.877**	**0.900**	**0.900**	0.875	0.887	**0.906**

the C2C costs as distances between nodes in the distance vector algorithm allows to prioritize and limit the set of inquired cameras. Comparisons with different baseline methods and state-of-the-art approaches show that the proposed approach improves the original solutions. Such promising performances have been achieved through unsupervised consensus.

REFERENCES

[1] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics," *ACM Computing Surveys*, vol. 46, pp. 1–37, Nov. 2013. 1, 4

[2] N. Martinel and C. Micheloni, "Sparse Matching of Random Patches for Person Re-Identification," in *ICDSC*, pp. 1–6, 2014. 1

[3] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Multiple-shot Human Re-Identification by Mean Riemannian Covariance Grid," in *AVSS*, (Klagenfurt, Austria), pp. 179–184, 2011. 1

[4] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and Appearance Context Modeling," in *ICCV*, pp. 1–8, 2007. 1

[5] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person Re-identification by Video Ranking," in *ECCV*, pp. 688–703, 2014. 1

[6] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint Invariant Human Re-Identification in Camera Networks Using Pose Priors and Subject-Discriminative Features," *IEEE TPAMI*, vol. 37, no. 5, pp. 1095–1108, 2015. 1

[7] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-Supervised Coupled Dictionary Learning for Person Re-identification," in *CVPR*, 2014. 1

[8] G. Lisanti, I. Masi, A. D. Bagdanov, and A. D. Bimbo, "Person Re-Identification by Iterative Re-Weighted Sparse Ranking," *IEEE TPAMI*, vol. 37, pp. 1629–1642, Aug. 2015. 1

[9] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views," *CVIU*, vol. 109, pp. 146–162, Feb. 2008. 1

[10] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Re-Identification in the Function Space of Feature Warps," *IEEE TPAMI*, vol. 37, pp. 1656–1669, Aug. 2015. 1, 4

[11] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni, "Modeling feature distances by orientation driven classifiers for person re-identification," *JVCIR*, vol. 38, pp. 115–129, 2016. 1

[12] M. Hirzer, P. M. Roth, K. Martin, and H. Bischof, "Relaxed Pairwise Learned Metric for Person Re-identification," in *ECCV*, vol. 7577, pp. 780–793, 2012. 1

[13] D. Tao, L. Jin, Y. Wang, and X. Li, "Person Reidentification by Minimum Classification Error-Based KISS Metric Learning," *IEEE TCYB*, pp. 1–11, 2014. 1

[14] N. Martinel, C. Micheloni, and G. L. Foresti, "Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning," *IEEE TIP*, vol. 24, pp. 5645–5658, dec 2015. 1

[15] W.-S. Zheng, S. Gong, and T. Xiang, "Towards Open-World Person Re-Identification by One-Shot Group-based Verification," *IEEE TPAMI*, vol. 8828, no. 2, pp. 1–1, 2015. 1

[16] J. Garcia, N. Martinel, C. Micheloni, and A. Gardel, "Person Re-Identification Ranking Optimisation by Discriminant Context Information Analysis," in *ICCV*, 2015. 1

[17] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning Locally-Adaptive Decision Functions for Person Verification," in *CVPR*, pp. 3610–3617, June 2013. 1, 4

[18] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang, "Similarity Learning on an Explicit Polynomial Kernel Feature Map for Person Re-Identification," in *CVPR*, 2015. 1

[19] N. Martinel, C. Micheloni, and C. Piciarelli, "Distributed Signature Fusion for Person Re-Identification," in *ICDSC*, pp. 1–6, 2012. 1

[20] S. F. Tahir and A. Cavallaro, "Cost-effective features for reidentification in camera networks," *IEEE TCSVT*, vol. 24, no. 8, pp. 1362–1374, 2014. 1

[21] A. Das, A. Chakraborty, and A. K. Roy-chowdhury, "Consistent Re-identification in a Camera," in *ECCV*, pp. 330–345, 2014. 1, 4

[22] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985. 2, 3

[23] A. Bosch, A. Zisserman, and X. Munoz, "Image Classification using Random Forests and Ferns," in *ICCV*, pp. 1–8, 2007. 4

[24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE TSMC*, vol. 3, pp. 610–621, Nov. 1973. 4

[25] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, pp. 2288–2295, 2012. 4

[26] S. Pedagadi, J. Orwell, and S. Velastin, "Local Fisher Discriminant Analysis for Pedestrian Re-identification," in *CVPR*, pp. 3318–3325, 2013. 4

[27] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network," in *CVPRW*, (Providence, RI), pp. 31–36, 2012. 4