A Pool of Multiple Person Re-Identification Experts,⋆⋆

Niki Martinel[a,⋆⋆], Christian Micheloni[a], Gian Luca Foresti[a]

[a]*University of Udine, Department of Mathematics and Computer Science, Via Delle Scienze, 206, Udine 33100, Italy*

ABSTRACT

The person re-identification problem, i.e. recognizing a person across non-overlapping cameras at different times and locations, is of fundamental importance for video surveillance applications. Due to pose variations, illumination conditions, background clutter, and occlusions, re-identify a person is an inherently difficult problem which is still far from being solved. In this work, inspired by the recent police lineup innovations, we propose a re-identification approach where Multiple Re-identification Experts (MuRE) are trained to reliably match new probes. The answers from all the experts are then combined to achieve a final decision. The proposed method has been evaluated on three datasets showing significant improvements over state-of-the-art approaches.

1. Introduction

Recognize a person moving across the disjoint fields-of-view (FoVs) of a camera network is a challenging problem known as person re-identification. It is of fundamental importance for wide area video analytics systems, where, due the amount of human supervision, privacy concerns, and maintenance costs involved, a large amount of the environment is not covered by sensors FoVs Martinel et al. (2014b). Many different related applications, like situational awareness Alcaraz and Lopez (2013), scene understanding Nayak et al. (2013), etc. would benefit from it.

In spite of a surge of effort put in by the community in the recent years (see Vezzani et al. (2013)), re-identify a person is still an open issue due to a number of challenges. In particular, in a wide are surveillance scenario cameras are deployed to cover as much area as possible. Thus, the acquired footages have (i) low resolution, (ii) the persons are viewed from different points-of-view, and (iii) their appearance drastically changes due to the different illumination and color conditions as well as their poses. As a result of these variations, the appearance of a person differs significantly in the disjoint views.

To tackle such challenges, current methods mainly follow three different approaches. However, all of them share the idea that, in order to re-identify a person, a feature representation should be computed by considering the visual appearance. Discriminative signature-based methods form the first class of approaches. These focus on novel highly discriminative person signatures that are robust to the aforementioned wide area camera network issues. Feature transformation methods belong to the second class of approaches and aim to model the transformation of the features that is undergoing between disjoint cameras. Finally, metric learning-based methods define the third class of approaches. These aim to learn an optimal signature distance metric such that the intra-class distances are minimized while the inter-class distances are maximized.

In this work, a re-identification framework inspired by the modalities adopted by the organs of justice to conduct crime investigations is proposed. The idea comes from the widely used lineup procedure: an *expert* putative identification of a suspect is confirmed to a level that can count as evidence at trial. As shown in National Research Council (2014), such a practice plays an important role in criminal cases. However, the limits of human vision and memory have, sometimes, lead to failure of identifications. To sidestep such issues, novel modalities have been introduced. Among these, a common practice is to require the intervention of *multiple identification experts*.

The idea is well suited for the re-identification problem. In the proposed work, such a model has been adopted and multiple experts are trained to re-identify persons moving across disjoint

⋆⋆Corresponding author:
e-mail: niki.martinel@uniud.it (Niki Martinel)

cameras. Differently from the existing methods, the single decision taken by a trained expert –that may not be enough to achieve a reliable re-identification– is replaced by an answer obtained by pooling the decision of the multiple trained experts.

2. Related Work

In the recent past Vezzani et al. (2013), many different works have been proposed to tackle the re-identification challenges. In the following, a brief presentation of the recent appearance-based approaches is given.

To obtain *discriminative signature* representations from disjoint camera views, various pursuits have been reported. Multiple local features Martinel and Foresti (2012); Bak et al. (2012), also biologically-inspired Ma et al. (2014a), were used to compute discriminative signatures for each person using multiple images. Other methods investigated dissimilarity-based approaches Satta et al. (2012), adopted collaborative representation that best approximates the query frames Wu et al. (2012) or exploited reference sets to represent the whole body as an assembly of compositional and alternative parts Xu et al. (2013). Recently, coupled dictionaries exploiting labeled and unlabeled data Liu et al. (2014) and sparse discriminative classifiers ensuring that the best candidates are ranked at each iteration were proposed Lisanti et al. (2014).

Due to the significant appearance changes, achieving accurate classification through such method is very difficult. Methods in the second and third classes of approaches aim to overcome such a problem.

In particular, *features transformation-based methods* address the re-identification problem by finding the transformation functions that affect the visual features acquired by disjoint cameras. These methods were initially designed to transform the feature space of one camera to the feature space of another one Javed et al. (2008). Recently, a few methods Li and Wang (2013); Martinel et al. (2015a) had also considered the fact that the transformation is not unique and it depends on several factors (e.g. poses and viewpoint changes Garcia et al. (2014), image resolutions, photometric settings of cameras).

Methods that exploit *optimal feature distances* advantage of a training phase to learn non-Euclidean distances used to compute the match in a different feature space. Several method were proposed by learning a relaxed Mahalanobis metric Hirzer et al. (2012a), by considering multiple metrics Ma et al. (2014b); Martinel et al. (2015b) in a transfer learning set up Li et al. (2012), or by relying on equivalence constraints Kostinger et al. (2012); Tao et al. (2014). Others have focused on local distance comparison problems Li and Wang (2013); Li et al. (2013); Liong et al. (2015).

Finally, it is worth mentioning that the re-identification can be also conducted by exploiting biometric features Micheloni et al. (2009), mainly represented by soft biometrics Nambiar et al. (2015) and gait features Sarkar et al. (2005); Lu and Tan (2010a). Works in such direction introduced methods achieving view invariant properties Liu et al. (2011); Lu and Tan (2010b) also by exploiting multiple view fusion methods Lu and Zhang (2007). The problem of re-identifying a person walking with
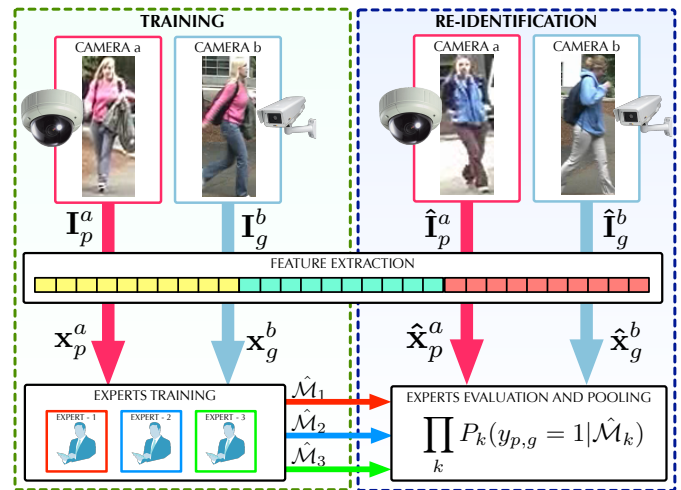
Fig. 1. Proposed expert-based system architecture. A robust feature representation is computed for each image acquired by a camera in the network. In the training phase, such representations computed for image pairs are used to train a set of experts. In the re-identification phase, the trained experts evaluate the new feature representations of an image pair. The answers from all the experts are pooled to obtain the final decision.

arbitrary directions was explored in Lu et al. (2014). Despite the success of such methods, computing such features require a constrained camera deployment and high resolution sensors which are not always available in a wide area camera network. As a result, appearance features are still the dominant choice.

All such methods aim to achieve the optimal re-identification by proposing a single solution. Thus, they believe that the given answer is unique and it is the only one that should be used to decide if two images acquired by disjoint cameras belong to the same person or not. The only work that has a slightly different view, which is partially shared with the proposed method, is Li and Wang (2013). It differs from the proposed approach on the following aspects. In Li and Wang (2013), the feature space is partitioned according to the orientation of a person, then a metric is learned for each partition. During the re-identification, the orientations of the persons in the two images are used to select the metric used to match the corresponding features. Hence, in Li and Wang (2013), it is assumed that the orientation of a person can be computed and a single metric is still enough to provide the final answer. In the proposed work, no assumption is made on the appearance/pose of a person and the answers from all the experts are considered to reach a final decision.

3. The Experts-Based Approach

3.1. Approach Overview

The steps conducted to perform the re-identification using the proposed MuRE approach are illustrated in Fig. 1. As shown, it considers two phases which share two common steps. Given a pair of images acquired by disjoint cameras, these are input to the feature extraction module. This is in charge to compute a discriminative feature representation of each person considering visual clues only. In the training phase, the representations obtained for a training set of image pairs are given to the experts

which individually learn how to optimally discriminate between images of a same or different persons. In the re-identification phase, the trained experts are required to evaluate the representations extracted from a new pair of test images and to provide a pooled answer.

3.2. Experts Training

Let $\mathbf{I}_p^a \in \mathbb{R}^{m \times n}$ and $\mathbf{I}_g^b \in \mathbb{R}^{m \times n}$ be the images of persons p and g which have been acquired by camera a and camera b, respectively. Then, the corresponding feature representations denoted as $\mathbf{x}_p^a \in \mathbb{R}^d$ and $\mathbf{x}_g^b \in \mathbb{R}^d$ can be obtained by computing a suitable representation (e.g., histogram) of the outputs of feature extraction processes $\pi(\mathbf{I}_p^a, i, j)$ and $\pi(\mathbf{I}_g^b, i, j)$ (e.g., gradient orientations) computed for every image pixel at locations $i = 1, \ldots, m$, and $j = 1, \ldots, n$. Since the goal is to re-identify a person moving across disjoint cameras and image pairs are considered in the proposed framework, we can cast the problem as a binary classification one. Thus, to a given image pair $(\mathbf{I}_p^a, \mathbf{I}_g^b)$ corresponds a label $y_{p,g} \in \{0, 1\}$, where $y_{p,g} = 0$ if the images are of a different person (i.e., $p \neq g$), and $y_{p,g} = 1$ otherwise (i.e., $p = g$).

Assuming M persons are viewed by the two cameras, and the data is available for the training phase, then the corresponding feature vectors are collected in the sets $\mathcal{X}^a = \{\mathbf{x}_p^a | \mathbf{x}_p^a, p = 1, \ldots, M\}$ and $\mathcal{X}^b = \{\mathbf{x}_g^b, | \mathbf{x}_g^b, g = 1, \ldots, M\}$. These, together with the set containing all possible values of $y_{p,g}$ denoted here as $\mathcal{Y} = \{y_{p,g} | y_{p,g}, p = 1, \ldots, M, g = 1, \ldots, M\}$, are exploited to separately train K experts. In the current framework each expert can be different from the others, e.g. the first expert can be a Deep Net, the second a Support Vector Machine, the third a non-Euclidean metric, etc. To train each of such experts to discriminate between the set of feature vectors belonging to the same person and the set of feature vectors belonging to different persons suitable expert-dependent learning procedures should be adopted. However, in general, for each expert there exists a cost function which should be minimized to estimate the set of parameters that optimally separates the two sets as

$$\hat{\mathcal{M}}_k = \arg\min_{\mathcal{M}_k} \mathcal{L}_k \left(\mathcal{X}^a, \mathcal{X}^b, \mathcal{Y}, \mathcal{M}_k \right) \tag{1}$$

where $\mathcal{L}_k(\cdot)$ is the k-th expert-dependent cost function to minimize and $\mathcal{M}_k$ characterizes the k-th expert parameters (e.g., the connection weights of a Deep Neural Network, the coefficients of the separating hyperplane and bias of a Support Vector Machine, the entries of the matrix defining a non-Euclidean pseudo-metric, etc.).

3.3. Experts Evaluation and Pooling

The resulting estimated experts parameters $\hat{\mathcal{M}}_k$, for $k = 1, \ldots, K$ are then used in the re-identification phase to match a probe person viewed in camera a and a gallery person detected by camera b. More formally, given a probe image $\hat{\mathbf{I}}_p^a$, a gallery image $\hat{\mathbf{I}}_g^b$, the corresponding feature representations $\hat{\mathbf{x}}_p^a$ and $\hat{\mathbf{x}}_g^b$ are compared by each expert to obtain K separate answers.

In the current framework, it is assumed that each expert is not able to take a strong binary decision on the new sample pair, but it has some uncertainty about it. Hence, the expert answer can

be defined as the probability of a probe person p and a gallery person g being the same, given the observed feature representations and the estimated expert parameters. This translates to

$$P_k \left(y_{p,g} = 1 | \hat{\mathcal{M}}_k \right) = \sigma \left(\mathcal{J}_k(\hat{\mathbf{x}}_p^a, \hat{\mathbf{x}}_g^b, \hat{\mathcal{M}}_k) \right) \tag{2}$$

where $\mathcal{J}_k(\hat{\mathbf{x}}_p^a, \mathbf{x}_g^b, \hat{\mathcal{M}}_k)$ is the k-th expert decision function which output is computed by evaluating the input feature representations $\hat{\mathbf{x}}_p^a$ and $\mathbf{x}_g^b$ with the learned parameters $\hat{\mathcal{M}}_k$.

We assume that the output of an expert decision function $\mathcal{J}_k(\cdot)$ is a similarity score or a distance measure. To translate such an output to a probability value we introduced the $\sigma(\cdot)$ function. More specifically, if the expert output is a similarity score, then to ensure the value is in $[0, 1]$, we have used $\sigma(z) = \frac{1}{1+\exp(-z)}$ (i.e., the logistic function). On the other hand, if the expert output is a distance measure, we have we have used $\sigma(z) = \exp(-z)$.

In order to reach a common decision shared among the experts, the obtained answers must be pooled. Since the K answers are independent from each other and those are defined to be probabilities, the pooled answer can be obtained by computing the conditional probability considering all the K experts. Thus, the final answer is computed as

$$P \left(y_{p,g} = 1 | \hat{\mathcal{M}}_1, \ldots, \hat{\mathcal{M}}_K \right) = \prod_k P_k \left(y_{p,g} = 1 | \hat{\mathcal{M}}_k \right). \tag{3}$$

Such answer is finally used to compute the final ranking for re-identification.

4. Experimental Results

The proposed MuRE approach has been evaluated using three publicly available benchmark datasets: the VIPeR dataset Gray et al. (2007), the 3DPeS dataset Baltieri et al. (2011) and the CHUK02 dataset Li et al. (2012). These datasets have been selected because they provide many challenges faced in real scenarios, i.e., viewpoint, pose and illumination changes, different backgrounds, image resolutions, occlusions, etc. Specific dataset details and related challenges are described below.

4.1. Evaluation Settings

In the current framework, the following settings have been used to compute all the results.

4.1.1. Evaluation Criteria

The re-identification mechanism commonly depends on how the gallery and the probe sets are organized. Given N images per each person in the two sets, two main matching approaches are commonly adopted: i) single-shot ($N = 1$); ii) multiple-shot ($N > 1$). To consider both modalities, in the current framework, the same approach in Martinel et al. (2015a) has been adopted and all the $N \times N$ final answers are average pooled.

As commonly performed by the literature Vezzani et al. (2013), all the following results are shown using the Cumulative Matching Characteristic (CMC) curve and the normalized Area Under Curve (nAUC) values. The CMC curve is a plot of the recognition performance versus the rank score and

Fig. 2. 10 image pairs from the VIPeR dataset. The two rows show the different appearances of the same person viewed by two disjoint cameras.

represents the expectation of finding the correct match within the first k ones. The nAUC is a global indicator which describes how well a re-identification method performs irrespectively of the dataset size. For each dataset, the evaluation procedure is repeated 10 times using independent random splits and the average results are shown. All the results used for comparison with state-of-the-art methods were provided by the authors of the corresponding works.

4.1.2. Person Appearance and Expert Models

To model the person appearance, images are first resized to 64×128 pixels, then the WHOS person descriptor Lisanti et al. (2014) is extracted. As a result, each person is represented by a 5138-dimensional vector which is obtained by concatenating color histograms, LBP texture and HOG shape features.

Due to the recent success of metric learning algorithms, the LFDA Pedagadi et al. (2013), KISSME Kostinger et al. (2012) and LADF Li et al. (2013) approaches have been selected as re-identification experts to evaluate the proposed approach. Results obtained using such methods have been computed using our implementations and the proposed person representation. However, such methods also provide re-identification results on some of the considered datasets. When MuRE is compared to state-of-the-art methods, the results directly provided by the authors of the corresponding works are shown. To indicate which methods have been used in the proposed framework the following notation is used: MuRE (a-b-...), where "a" and "b" are the acronyms denoting the experts methods. The distance output by each of such experts has been translated to a probability value using $\sigma(z) = \exp^{-z}$.

4.2. VIPeR Dataset[1]

The VIPeR dataset Gray et al. (2007) is considered the most challenging one for person re-identification due to the changes in illumination and pose. This dataset contains low spatial resolution images of 632 persons viewed by two different cameras in an outdoor environment (see Fig. 2 for a few samples).

4.2.1. Performance Analysis

Results in Fig. 3a and Table 1 have been computed to evaluate the performance of each single expert. Following a common approach Gray et al. (2007); Lisanti et al. (2014); Martinel et al. (2015a), the results have been computed using 316 persons both

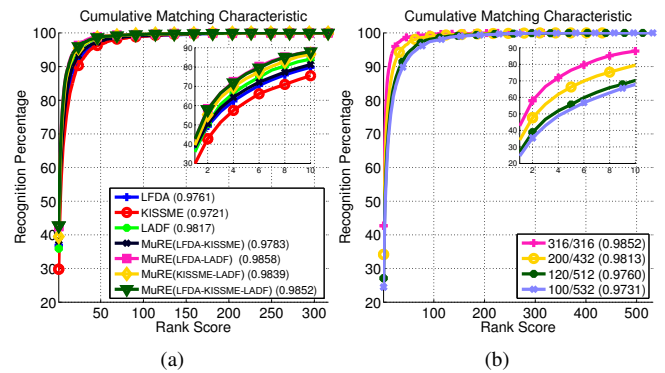---
[1]Available at http://soe.ucsc.edu/~dgray/

(a) (b)

Fig. 3. Results on the VIPeR dataset reported as CMC curves averaged over 10 different trials. In (a) performance of MuRE using different experts. In (b) results of MuRE (LFDA-KISSME-LADF) are shown as a function of the test set size.

Table 1. Comparison of the performance achieved by the selected experts on the VIPeR dataset. Best results for each rank are in boldface font.

Rank →	1	10	20	50	100	nAUC
LFDA Pedagadi et al. (2013)	36.90	79.91	89.62	97.50	99.15	0.9761
KISSME Kostinger et al. (2012)	29.81	75.51	87.75	96.77	99.02	0.9721
LADF Li et al. (2013)	35.95	84.15	93.07	98.32	99.53	0.9817
MuRE (LFDA-KISSME)	38.26	81.33	91.46	97.66	99.21	0.9783
MuRE (LFDA-LADF)	42.50	**88.04**	**95.13**	**99.02**	99.56	**0.9858**
MuRE (KISSME-LADF)	39.49	86.77	94.24	98.77	**99.59**	0.9839
MuRE (LFDA-KISSME-LADF)	**42.72**	**88.04**	94.87	98.73	99.56	0.9852
Max Voting Fusion	39.87	87.04	94.21	98.61	99.19	0.9811

for training and for testing. When more than 1 expert is considered, eq.(3) is used to obtain the pooled answer. Results demonstrate that the optimal overall performance is achieved by combining LFDA and LADF (i.e., by MuRE (LFDA-LADF)). The highest rank 1 score is achieved by pooling all the three experts answers (i.e., MuRE (LFDA-KISSME-LADF)).

In Table 1 a comparison with a max voting fusion scheme is also provided. In such a case, each expert makes a decision, then the max voting rule is used to fuse the decisions of all the experts (i.e., LFDA, KISSME and LADF). Results show that the max voting fusion approach achieves worse performance than the proposed one. In particular, the recognition percentage at rank 1 is 3% lower than MuRE (LFDA-KISSME-LADF).

4.2.2. Comparison with State-of-the-art Methods

In Table 2, the results of the proposed MuRE framework are compared to the ones achieved by current state-of-the-art approaches. The results are reported for the case when 316 persons are considered in both the training and the test set. Results demonstrate that the MuRE (LFDA-KISSME-LADF) approach has rank 1 performance very close to the LMF Zhao et al. (2014)+LADF Li et al. (2013) approach and achieves better results than any other existing method on higher ranks. This is reflected by the reported nAUC value.

As commonly performed in literature An et al. (2013); Ma et al. (2014b), the proposed method has been evaluated considering three additional different train/test sizes. The performance achieved under such scenarios are shown in Fig. 3b and Table 3.

Results demonstrate that our method outperforms all existing approaches and it is robust to significant reductions in the trai-

Table 2. Comparison with state-of-the-art methods on the VIPeR dataset. Best results for each rank are in boldface font.

Rank →	1	10	20	50	100	nAUC
MuRE (LFDA-KISSME-LADF)	42.72	**88.04**	**94.87**	**98.73**	**99.56**	**0.9852**
LMF Zhao et al. (2014)+LADF Li et al. (2013)	**43.29**	85.13	94.12	-	-	-
LOMO+XQDA Liao et al. (2015)	40.00	80.51	91.08	-	-	-
PKFM Chen et al. (2015)	36.8	83.7	91.7	97.8	-	-
SWF Martinel et al. (2014a)	32.97	75.63	86.87	96.17	98.96	0.9701
kBiCoV Ma et al. (2014a)	31.11	70.71	82.44	-	-	-
QALF Zheng et al. (2015)	30.17	62.44	73.81	-	-	-
SalMatch Zhao et al. (2013)	30.16	65.54	79.15	91.49	98.10	0.9542
LAFT Li and Wang (2013)	29.60	69.30	81.34	96.80	-	-
LADF Li et al. (2013)	29.30	78.80	92.20	97.40	-	-
LMF Zhao et al. (2014)	29.10	66.30	81.00	-	-	-
MtMCML Ma et al. (2014b)	28.83	75.82	88.51	-	-	-
ISR Lisanti et al. (2014)	27.43	61.06	72.92	86.69	-	0.9410
PatMatch Zhao et al. (2013)	26.90	62.34	75.63	90.51	97.47	0.9496
WFS Martinel et al. (2015a)	25.81	69.56	83.67	95.12	98.89	-
SSCDL Liu et al. (2014)	25.6	68.1	83.6	-	-	-

Table 3. Comparisons on the VIPeR dataset. Recognition rates per rank score as a function of the test set size. Best results are in boldface font.

Test Set Size	432			512				532		
Rank →	1	10	20	1	5	10	20	1	10	20
MuRE (LFDA-KISSME-LADF)	**34.19**	**79.47**	**89.44**	**27.11**	**55.66**	**70.27**	**83.50**	**24.49**	**67.82**	**81.02**
SWF Martinel et al. (2014a)	24.72	66.29	82.70	14.77	38.06	53.29	68.32	10.67	45.46	65.95
RCCA An et al. (2013)	22	59	75	-	-	-	-	15	47	60
MtMCML Ma et al. (2014b)	20	62	77	-	-	-	-	12	45	61
RPLM Hirzer et al. (2012b)	20	56	71	-	-	-	-	11	38	52
NRDV Zhou et al. (2014)	20	54	67	-	-	-	-	14	44	55
MCE-KISS Tao et al. (2014)	14	49	69	-	-	-	-	-	-	-
RS-KISS Tao et al. (2013)	10	40	61	-	-	-	-	-	-	-
PRDC Zheng et al. (2013)	13	44	60	9.12	24.19	34.40	48.55	9	34	49
MCC Zheng et al. (2013)	-	-	-	5.00	16.32	25.92	39.64	-	-	-
LAFT Li and Wang (2013)	-	-	-	12.90	30.30	42.73	58.02	-	-	-

Fig. 4. 10 image pairs from the CUHK02 dataset. The two rows show the different appearances of the same person viewed by two disjoint cameras.

ning set size. This is a desirable property that avoids the need of large quantities of labeled training data. More in details, when 432 persons are considered as test set, MuRE (LFDA-KISSME-LADF) has a rank 1 correct recognition rate of 34.19%, while the runner up (RCCA An et al. (2013)) has a recognition rate of only 22%. A similar behavior is achieved when the number of test persons increases to 512 and 532.

4.3. CUHK02 Campus Dataset[2]

The CUHK02 Campus dataset Li et al. (2012) has images acquired by 5 disjoint camera pairs (denoted as P1-P5) deployed in a campus environment. Each person has two images in each camera. To evaluate the proposed method and compare it to the state-of-the-art, the same protocol used in Zhao et al. (2013); Li et al. (2012) has been used, hence results for camera pair P1 when $N \in \{1, 2\}$ are provided. In this camera pair, images from the first camera are captured from lateral view,

while images from the second camera are acquired from a frontal view or back view (see Fig. 4).

4.3.1. Performance Analysis

As done for the VIPeR dataset, in Fig. 5a and Fig. 5b the results achieved by different experts are provided in terms of CMC curves. The reported results have been computed using 486 persons for training and 485 persons for testing.

In Fig. 5a, results are for the single-shot approach. In such a case, results show that while MuRE (LFDA-KISSME-LADF) reaches the highest rank 1 correct recognition rate (36.62%), the optimal overall performance is achieved by combining LFDA and LADF (i.e., MuRE (LFDA-lADF)).

Performance shown in Fig. 5b are for the multiple-shot scenario with $N = 2$. Results demonstrate that the MuRE framework yields to better performance than any other baseline method. In particular, when $N = 2$ images are used, MuRE (LFDA-LADF) reaches the highest rank 1 correct recognition rate (57.29%) and the optimal overall performance (with an nAUC of 0.9892). It is worth noticing that in such a case the single LADF expert yields to better overall performance than all MuRE combination (other than MuRE(LFDA-LADF)). This is due to the fact that KISSME performance is very poor compared to other experts. Therefore, including it in the MuRE framework causes a degradation of the performance.

In Fig. 5c, results achieved by the proposed framework using different train/test sizes are shown. Results demonstrate that when the proposed framework is robust to even extreme cases like when only 97 persons are used for training and 874 are used for testing. This is reflected by the fact that, among all the five considered splits, the nAUC values change by less than 3%.
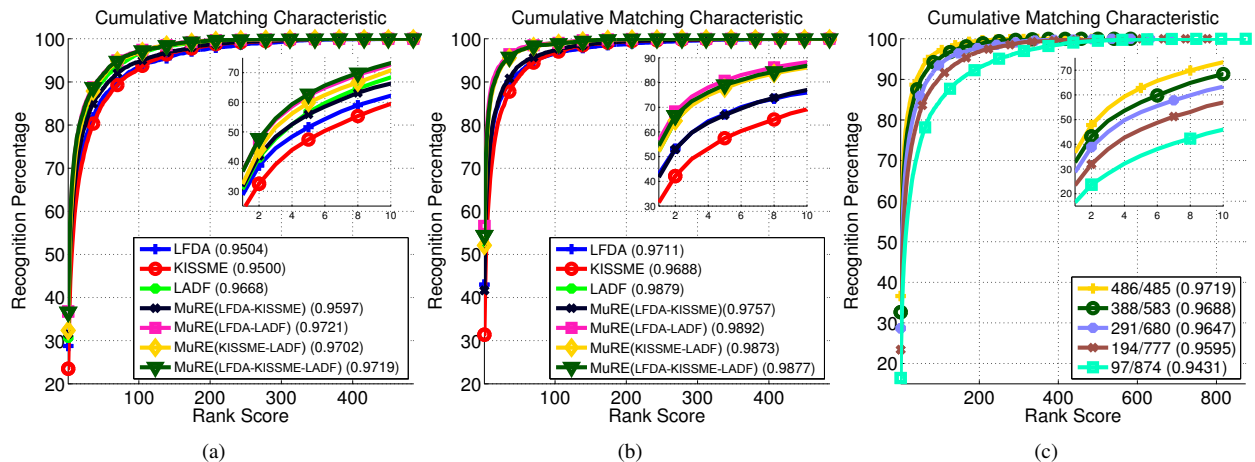
Fig. 5. Results on the CUHK02 dataset reported as CMC curves averaged over 10 different trials. In (a) comparisons with different experts are shown for the single shot-approach. In (b) the results achieved by the MuRE framework and the adopted experts are given for the multiple-shot approach with $N = 2$. In (c) results of the proposed MuRE (LFDA-KISSME-LADF) are shown as a function of the test set size.

Fig. 6. Qualitative performance of MuRE (LFDA-KISSME-LADF) on the CUHK02 dataset. In the first column 8 query persons are shown. The next 20 images per row represent the ranked matching persons. The correct match (also shown in the last column) is highlighted in green.

Finally, since existing algorithms are not achieving a 100% of correct recognition rate at rank 1, human intervention is still required to identify the true match within the given ranking. Thus, providing a suitable ranking for end-user inspection is a desirable feature that a re-identification algorithm should have. To show that MuRE has such a property, qualitative performance are shown in Fig.6: 8 query images and the first 20 ranks produced by MuRE (LFDA-KISSME-LADF) are depicted. In, particular we have included rankings in which the true match is not located in the first position (2nd to 8th rows). Results demonstrate that even in such cases, the MuRE framework is able to correctly capture the inter-camera global appearance changes and produces a suitable ranking (i.e., persons share visual similarities) that can be finally exploited by the end-user.

4.3.2. Comparison with State-of-the-art Methods

In Table 4, the results of the proposed MuRE framework are compared to the ones achieved by current state-of-the-art ap-

Table 4. Comparison with state-of-the-art methods on the CUHK02 (P1) dataset. Best results for each rank are in bold.

Rank →	1	5	10	20	100	nAUC
Max Voting Fusion (N=1)	33.51	60.46	71.83	80.91	96.60	0.9694
MuRE (LFDA-KISSME-LADF) (N=1)	36.62	62.80	73.24	81.98	97.09	0.9719
Max Voting Fusion (N=2)	50.87	76.14	84.59	91.01	98.29	0.9821
MuRE (LFDA-KISSME-LADF) (N=2)	**54.41**	**79.11**	**86.80**	**92.21**	**98.82**	**0.9877**
SalMatch Zhao et al. (2013)	28.45	45.85	55.67	67.95	92.26	0.9374
PatMatch Zhao et al. (2013)	20.39	34.12	41.09	51.56	87.91	0.9065
TML(Our_Generic) Li et al. (2012)	20.53	45.54	56.61	69.62	93.75	-

proaches. Since the CUHK02 dataset has 2 images per person in each camera, multiple-shot performance with $N = 2$ are also shown. The reported results have been computed using 486 persons for training and 485 persons for testing.

Results demonstrate that the MuRE (LFDA-KISSME-LADF) approach has the best performance on every considered rank when 1 or 2 images per person are used. In particular, when $N = 2$ images are considered, MuRE (LFDA-KISSME-LADF) has a rank 1 of 54.41% which almost doubles the previous top rank 1 achieved by SalMatch Zhao et al. (2013).

Comparisons with the max voting fusion scheme are also provided. Results show that under both the single shot and the multiple shot scenarios, the proposed fusion scheme yields to better performance than the max voting one.

4.4. 3DPeS Dataset[3]

The 3DPeS dataset Baltieri et al. (2011) has images from 8 cameras which present different light conditions and viewpoints (see Fig.7). Different sequences of 191 persons have been taken from such a multi-camera system on different days, under strongly changing illumination conditions. Partial occlusions and multiple persons appearing in the same image introduce additional challenges.

Fig. 7. 10 image pairs from the 3DPeS dataset. The two rows show the different appearances of the same person viewed by two disjoint cameras.
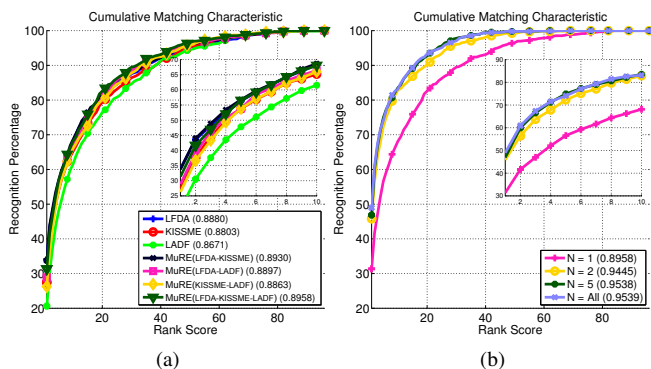
Table 5. Comparison of the proposed method on the 3DPeS dataset. Best results are in bold.

Rank →	1	10	25	50	nAUC
Max Voting Fusion (N=All)	46.19	82.97	94.76	99.51	0.9506
MuRE (LFDA-KISSME-LADF) (N=1)	31.35	68.12	86.25	96.56	0.8958
MuRE (LFDA-KISSME-LADF) (N=2)	45.83	82.92	93.44	99.38	0.9445
MuRE (LFDA-KISSME-LADF) (N=All)	**48.96**	**83.13**	**95.10**	**99.79**	**0.9539**
LFDA Pedagadi et al. (2013)	33.43	69.98	84.80	95.07	0.8870
KISSME Kostinger et al. (2012)	22.94	62.21	80.74	93.21	0.8582
LMNN-R Dikmen et al. (2010)	23.03	55.23	73.44	88.92	0.8191

(a)　　　　(b)

Fig. 8. Results on the 3DPeS dataset reported as CMC curves averaged over 10 trials. In (a) performance of different experts are compared to the MuRE framework under a single-shot modality. In (b) results are shown as a function of the number of images available for each person.

4.4.1. Performance Analysis

In Fig. 8a the results achieved by different experts are provided in terms of CMC curves. The reported results have been computed using 95 persons for training and 96 persons for testing. Differently from the other two datasets, results show that MuRE (LFDA-KISSME-LADF) obtains the optimal overall performance, but the highest rank 1 correct recognition rate (33.46%) is achieved by combining LFDA and LADF.

In Fig. 8b, CMC performance on the 3DPeS dataset obtained by MuRE (LFDA-KISSME-LADF) are provided for $N \in \{1, 2, 5, All\}$. Results show that the performance strongly improves just by considering more than a single image. However, there is a subtle difference in the overall performance for the case when $N = 5$ and $N = All$. Indeed, the obtained nAUC values differ only by 0.0001. Despite this, when all the images are considered the obtained rank 1 improves of about 1.32% with respect to the case when $N = 5$ images are used.

4.4.2. Comparison with State-of-the-art Methods

In Table 5, the performance of the proposed MuRE framework are compared to the ones obtained by LFDA Pedagadi et al. (2013), KISSME Kostinger et al. (2012) and LMNN-R Dikmen et al. (2010). The same experimental protocol of Pedagadi et al. (2013) has been adopted, hence the dataset has been split into a training set and a test set each one composed of 95 randomly selected persons. Since in Pedagadi et al. (2013) no details regarding the number of images used for each person are given, it is assumed that their results have been computed

---

³Available at http://www.openvisor.org/3dpes.asp

using all the available ones. Results show that the proposed method achieves state-of-the-art performance when a single-shot approach is used and outperforms existing methods when $N \geq 2$. In particular, a rank 1 correct recognition rate of 48.96% is achieved when all the available images are used.

5. Discussion

The reported results show that the proposed MuRE framework performs better than any other existing method on all the three considered benchmark datasets. However, as shown in Fig. 9, the approach performance analysis conducted on each dataset has shown that there is not much strong consistency on the performance when two or more experts are considered. Indeed, for two datasets the top rank 1 performance are achieved when only two experts are used, and the optimal global performance are obtained when all experts are considered. For the last dataset, the opposite result is achieved. This brings out of the water a common problem in experts pooling Garg et al. (2004), which is defining (or learning) proper ways of pooling the answers from multiple experts. Since the preliminary results obtained by pooling the experts answers through probability rules are promising, more complex ways of pooling will be investigated in the future.

6. Conclusions

In the proposed work, a re-identification framework inspired by the real police lineup method has been proposed. The recent idea that the intervention of multiple identification experts is better than using a single answer by a single expert has been adapted for person re-identification purposes. In the current framework, different experts have been trained to discriminate between feature representations computed for pairs of images of same or different persons. In the re-identification phase, the answers from all the experts are pooled using probability rules. Results obtained by evaluating the method on 3 benchmark datasets have demonstrated that superior performance than state-of-the-art approaches are achieved.

References

Alcaraz, C., Lopez, J., 2013. Wide-Area Situational Awareness for Critical Infrastructure Protection. IEEE Computer 46, 30–37.

An, L., Kafai, M., Yang, S., Bhanu, B., 2013. Reference-Based Person Re-Identification, in: AVSS.

Bak, S., Corvee, E., Bremond, F., Thonnat, M., 2012. Boosted human re-identification using Riemannian manifolds. Image and Vision Computing 30, 443–452.
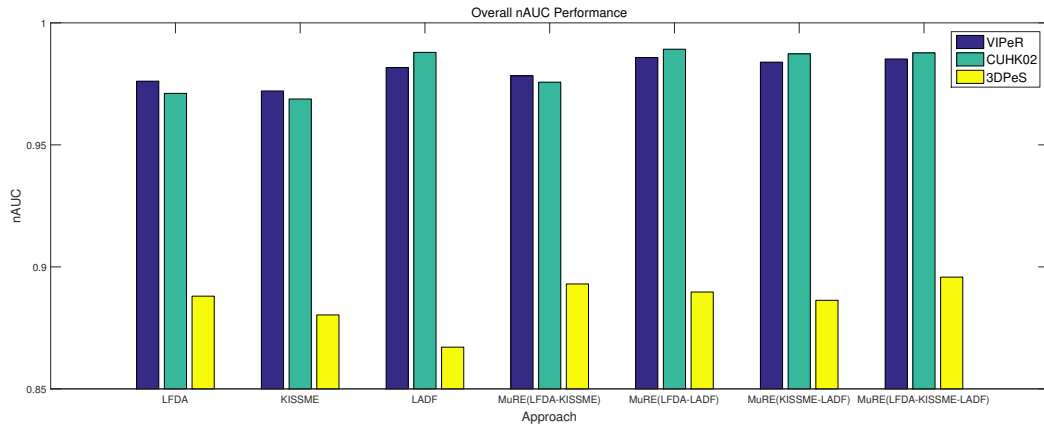
Fig. 9. nAUC performances of single re-identification experts are compared to the ones obtained by using them withing the MuRE framework. Results are shown for all the 3 considered benchmark datasets.

Baltieri, D., Vezzani, R., Cucchiara, R., 2011. 3DPeS: 3D People Dataset for Surveillance and Forensics, in: International ACM Workshop on Multimedia access to 3D Human Objects, pp. 59–64.

Chen, D., Yuan, Z., Hua, G., Zheng, N., Wang, J., 2015. Similarity Learning on an Explicit Polynomial Kernel Feature Map for Person Re-Identification, in: CVPR.

Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N., 2010. Pedestrian Recognition with a Learned Metric, in: ACCV, pp. 501–512.

Garcia, J., Martinel, N., Foresti, G.L., Gardel, A., Micheloni, C., 2014. Person Orientation and Feature Distances Boost Re-Identification, in: ICPR.

Garg, A., Jayram, T.S., Vaithyanathan, S., Zhu, H., 2004. Generalized Opinion Pooling, in: Intl. Symp. on Artificial Intelligence and Mathematics, pp. 79–86.

Gray, D., Brennan, S., Tao, H., 2007. Evaluating appearance models for re-congnition, reacquisition and tracking, in: PETS.

Hirzer, M., Roth, P.M., Bischof, H., 2012a. Person Re-identification by Efficient Impostor-Based Metric Learning, in: AVSS, pp. 203–208.

Hirzer, M., Roth, P.M., Martin, K., Bischof, H., 2012b. Relaxed Pairwise Learned Metric for Person Re-identification, in: ECCV, pp. 780–793.

Javed, O., Shafique, K., Rasheed, Z., Shah, M., 2008. Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views. CVIU 109, 146–162.

Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H., 2012. Large scale metric learning from equivalence constraints, in: CVPR, pp. 2288–2295.

Li, W., Wang, X., 2013. Locally Aligned Feature Transforms across Views, in: CVPR, IEEE. pp. 3594–3601.

Li, W., Zhao, R., Wang, X., 2012. Human Reidentification with Transferred Metric Learning, in: ACCV, pp. 31–44.

Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R., 2013. Learning Locally-Adaptive Decision Functions for Person Verification. CVPR , 3610–3617.

Liao, S., Hu, Y., Zhu, X., Li, S.Z., 2015. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning, in: CVPR.

Liong, V.E., Lu, J., Ge, Y., 2015. Regularized local metric learning for person re-identification. Pattern Recognition Letters , 1–9.

Lisanti, G., Masi, I., Bagdanov, A., Del Bimbo, A., 2014. Person Re-identification by Iterative Re-weighted Sparse Ranking. IEEE TPAMI , 1–1.

Liu, N., Lu, J., Tan, Y.P., 2011. Joint subspace learning for view-invariant gait recognition. IEEE SPL 18, 431–434.

Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J., 2014. Semi-Supervised Coupled Dictionary Learning for Person Re-identification, in: CVPR.

Lu, J., Tan, Y.P., 2010a. Gait-based human age estimation. IEEE TIFS 5, 761–770.

Lu, J., Tan, Y.P., 2010b. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. Pattern Recognition Letters 31, 382–393.

Lu, J., Wang, G., Moulin, P., 2014. Human identity and gender recognition from gait sequences with arbitrary walking directions. IEEE TIFS 9, 51–61.

Lu, J., Zhang, E., 2007. Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. Pattern Recognition Letters 28, 2401–2411.

Ma, B., Su, Y., Jurie, F., 2014a. Covariance Descriptor based on Bio-inspired Features for Person Re-identification and Face Verification. Image and Vision Computing 32, 379–390.

Ma, L., Yang, X., Tao, D., 2014b. Person Re-Identification Over Camera Networks Using Multi-Task Distance Metric Learning. IEEE TIP 23, 3656–3670.

Martinel, N., Das, A., Micheloni, C., Roy-Chowdhury, A., 2015a. Re-Identification in the Function Space of Feature Warps. IEEE TPAMI 37, 1656–1669.

Martinel, N., Foresti, G.L., 2012. Multi-signature based person re-identification. Electronics Letters 48, 764–765.

Martinel, N., Micheloni, C., Foresti, G.L., 2014a. Saliency Weighted Features for Person Re-Identification, in: ECCVW, pp. 291–208.

Martinel, N., Micheloni, C., Foresti, G.L., 2015b. Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning. IEEE TIP 24, 5645–5658.

Martinel, N., Micheloni, C., Piciarelli, C., Foresti, G.L., 2014b. Camera Selection for Adaptive Human-Computer Interface. IEEE TSMC-C 44, 653–664.

Micheloni, C., Canazza, S., Foresti, G.L., 2009. Audio-video biometric recognition for non-collaborative access granting. Journal of Visual Languages and Computing 20, 353–367.

Nambiar, A., Bernardino, A., Nascimento, J., 2015. Shape Context for soft biometrics in person re-identification and database retrieval. Pattern Recognition Letters , 1–9.

National Research Council, 2014. Identifying the Culprit: Assessing Eyewitness Identification.

Nayak, N.M., Zhu, Y., Roy-chowdhury, A.K., 2013. Exploiting Spatio-Temporal Scene Structure for Wide-Area Activity Analysis in Unconstrained Environments. IEEE TIFS , 1–1.

Pedagadi, S., Orwell, J., Velastin, S., 2013. Local Fisher Discriminant Analysis for Pedestrian Re-identification, in: CVPR, pp. 3318–3325.

Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W., 2005. The humanID gait challenge problem: Data sets, performance, and analysis. IEEE TPAMI 27, 162–177.

Satta, R., Fumera, G., Roli, F., 2012. Fast person re-identification based on dissimilarity representations. Pattern Recognition Letters 33, 1838–1848.

Tao, D., Jin, L., Wang, Y., Li, X., 2014. Person Reidentification by Minimum Classification Error-Based KISS Metric Learning. IEEE TCYB , 1–1.

Tao, D., Jin, L., Wang, Y., Yuan, Y., Li, X., 2013. Person Re-Identification by Regularized Smoothing KISS Metric Learning. IEEE TCSVT 23, 1675–1685.

Vezzani, R., Baltieri, D., Cucchiara, R., 2013. People reidentification in surveillance and forensics. ACM Computing Surveys 46, 1–37.

Wu, Y., Minoh, M., Mukunoki, M., Li, W., Lao, S., 2012. Collaborative Sparse Approximation for Multiple-Shot Across-Camera Person Re-identification, in: AVSS, pp. 209–214.

Xu, Y., Lin, L., Zheng, W.S., Liu, X., 2013. Human Re-identification by Matching Compositional Template with Cluster Sampling, in: ICCV, pp. 3152–3159.

Zhao, R., Ouyang, W., Wang, X., 2013. Person Re-identification by Salience Matching, in: ICCV, pp. 2528–2535.

Zhao, R., Ouyang, W., Wang, X., 2014. Learning Mid-level Filters for Person

537   Re-identification. CVPR , 144–151.

538 Zheng, L., Wang, S., Tian, L., He, F., Liu, Z., Tian, Q., 2015. Query-Adaptive
539   Late Fusion for Image Search and Person Re-identification, in: CVPR.

540 Zheng, W.S., Gong, S., Xiang, T., 2013. Re-identification by Relative Distance
541   Comparison. IEEE TPAMI 35, 653–668.

542 Zhou, T., Qi, M., Jiang, J., Wang, X., Hao, S., Jin, Y., 2014. Person Re-
543   identification based on nonlinear ranking with difference vectors. Informa-
544   tion Sciences .