Bidirectional Sparse Representations for Multi-shot Person Re-identification

Solène CHAN-LANG[1], Quoc Cuong PHAM[1], Catherine ACHARD[2]
[1]CEA LIST, Vision and Content Engineering Laboratory,
Point Courrier 173, F-91191 Gif-sur-Yvette, France
[2]Sorbonne University, UPMC Univ Paris 06, CNRS, UMR 7222, ISIR,
F-75005, Paris, France
Solene.CHAN-LANG@cea.fr, Quoc-Cuong.PHAM@cea.fr,catherine.achard@upmc.fr

Abstract

With the development of surveillance cameras, person re-identification has gained much interest, however re-identifying people across cameras remains a challenging problem which not only requires a good feature description but also a reliable matching scheme. Our method can be applied with any feature and focuses on the second requirement. We propose a robust bidirectional sparse coding method that improves simple sparse coding performances. Some recent work have already explored sparse representation for the re-identification task but none has considered the problem from both the probe and the gallery perspectives. We propose a bidirectional sparse representations method which searches for the most likely match for the test element in the gallery set and makes sure that the selected gallery match is indeed closely related to the probe. Extensive experiments on two datasets, CUHK03 and iLIDS-VID, show the effectiveness of our approach.

1. Introduction

Person re-identification consists in matching people captured with different cameras. The images available for that task not only depend on the people themselves, but also on the environment, the cameras and the person detector. All those factors make the re-identification very challenging. Indeed, the color rendering of someone's clothes can be drastically different depending on the camera and on whether the image has been captured in an indoor or outdoor place, by a sunny or cloudy day. Moreover the quality of the detector, the camera viewpoint and the person's pose can lead to huge misalignment issues. Consequently, the variability between two images of the same person can be greater than that of two different people.

Most of the studies on person re-identification focus on appearance modelling [6, 8, 9, 24, 27] and on learning a metric [7, 10, 11, 28, 29, 37] so that the intraclass variability becomes smaller than the interclass variability. Some recent work tackle more specific difficulties such as occlusion [33, 39] and low image resolution [12]. Open-world person re-identification is discussed in [5, 38].

Our work focuses on the matching part of person re-identification. It can work with any hand-crafted feature, or after a Mahalanobis metric learning step. It is best used in the multi-shot setting where multiple images of gallery and test cameras are available. This should be the case in real applications since those images usually come from tracking algorithms which extract tracklets of people.

Sparse coding has been shown to yield good results for the person re-identification task [13, 23]. It is based on the sparse representation of probe elements using gallery elements where therefore probe and gallery elements do not play a symmetric role. The main idea of our method is to exploit not only the effectiveness of sparse coding used in its traditionnal way but also to make it even more robust by symmetrizing the re-identification problem.

The bidirectional sparse coding method we present here is a three steps approach. The first step is the direct collaborative sparse representation where probe elements are approximated by a sparse linear combination of gallery elements. The second step is the reverse sparse representation where a collaborative sparse representation of gallery elements is computed using the probe elements and a generic dictionary. The third step is the ranking step: residual errors from the two previous steps are combined and ranked by increasing values. This method is more robust than the simple sparse representation. It significantly improves results in the multi-shot case, especially in the first ranks.

2. Related work

Numerous approaches have been developped to tackle the person re-identification problem. Early person re-identification methods essentially attempted to define good

978-1-5090-3811-4/16/$31.00 ©2016 IEEE

hand-crafted descriptors that would discriminate people well. Based on body symmetry, Farenzena *et al.* [8] define the parts of the body from which color and texture features are extracted while Cheng *et al.* [6] use a probabilistic model to find 6 body parts. Bak *et al.* [4] use covariance matrices on patches as descriptors. In those methods, different similarity measures are computed according to the features extracted or the body part where it comes from and a heuristic combination of those are used for the final ranking. Later, methods involving a learning step on training data emerged. Gray and Tao [9] used Adaboost to learn a set of discriminative features. Other early works aimed at handling the difference in camera color rendering by learning a camera color transfer function to recover the transformations from one camera to another [3, 30].

Recent methods are now mainly based on metric learning, neural networks and transfer learning. Metric learning approaches aim at learning a metric that better separates images belonging to the same person from those belonging to different people. While methods presented in [20, 21] learn one single metric to distinguish people captured from two cameras, Liong *et al.* [22] propose a combination of global and local metrics. The main drawback of current metric learning methods is that they do not perform well when the metric is learnt on a database and tested on another one.

Employing large and representative training data helps avoiding overfitting. Neural networks are a tool that can handle large amount of data. In person re-identification, several works [1, 17] propose to learn a deep neural network that takes a pair of images as input and outputs whether they describe the same person. Another work [35] trains a siamese network on three body parts. However training a neural network requires large representative and labeled training datasets that are hard to obtain.

The labeling problem is even worse for methods which make use of visual attributes. In [16], in addition to usual color and texture features, an SVM is trained to detect attributes such as the presence of hat, short, skirts, etc. Besides people's identity, labeled clothing is also needed for training the SVM. One way to use attributes without labeled data, is presented in [26]. Attributes are latent, they are learnt through a probabilistic model. Another way to go beyond the lack of labels is to use transfer learning which aims at transfering the knowledge acquired on a task to a similar but different task. Shi *et al.* [31] transfer knowledge of clothing attributes learnt on labeled fashion dataset to person re-identification. Zheng *et al.* [36] propose to learn for each query the usefulness of a feature based on reference information collected from any kind of images datasets.

Our work is mostly related to sparse representation and dictionary learning. Sparse representations have been widely explored in face recognition for their robustness against occlusion and corrupted data, but it is only recently that person re-identification approaches have taken interest in this kind of methods. In dictionary learning approaches, sparse representation is used as a feature. The idea is to learn a single dictionary [14, 15] or a coupled dictionary [2, 25] so that gallery and probe elements of the same person have similar sparse representations. In sparse representation methods [13, 23], the ranking is based on the residual errors computed thanks to the sparse coding. However, since the representations are sparse, no information is available to actually rank people from a certain rank onwards. In [23], Lisanti *et al.* present an iterative method to deal with this ranking issue. The work in [13] aims at better exploiting gallery multi-shot by introducing a group penalisation term which favors sparse reconstructions where the few participating dictionary elements belong to even fewer identities. Yet another way to exploit sparse representations in person re-identification is proposed by [18]. Distances and sparse representations are computed on features extracted both from the whole image and from body parts. The matching is based on the distances which are weighted by confidence values computed with the residual errors.

Though several methods using sparse representation have already been proposed for person re-identification, as far as we know, the proposed method is the first to exploit the asymmetric role of probe and gallery elements in sparse reconstructions. Our method improves performances compared to single direct Lasso sparse representation. It is a matching method that can be used with any feature. Therefore one can choose input features that result from a supervised learning step on training data.

3. Bidirectional sparse representations

3.1. Problem notations

We consider problems with two cameras. Known identities are captured by the gallery camera. Test images come from the probe camera. All test identities are present in the gallery.

Let K be the number of distinct people in the gallery. For each person $k \in [1, K]$, N_k images are available in the gallery camera so that there is a total of $N_g = \sum_{k=1}^{K} N_k$ images in the gallery. Each image is described by a column vector of length d. Let g_k^i be the feature associated with the i^{th} image of person k, and $G_k = [g_k^1, \ldots, g_k^{N_k}]$ be the matrix containing all the features of gallery person k. The concatenation of all the gallery features form the gallery dictionary $G = [G_1, \ldots, G_K]$ of size $d \times N_g$.

When given n images of a probe person l, the aim of person re-identification is to find the most likely match in the gallery. Let p_l^j stand for the feature vector of the j^{th} image of probe person l and $P_l = [p_l^1, \ldots, p_l^n]$ be the matrix containing all the features of probe person l. $P = [P_1, \ldots, P_K]$ is the concatenation of the probe dictionaries.

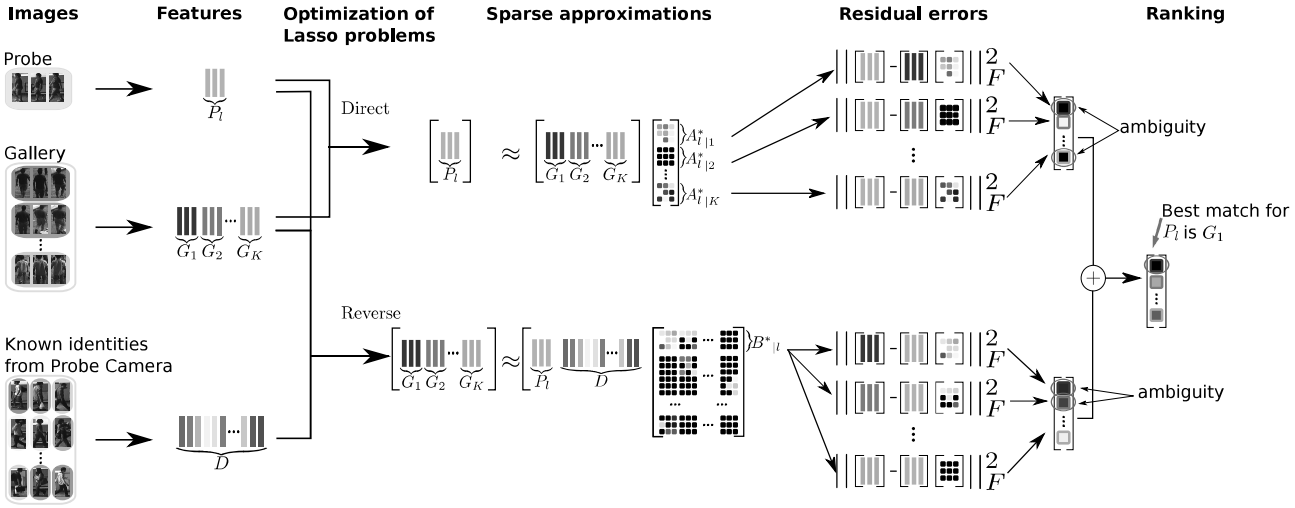

Figure 1. Overview of ABSR for multi-shot person re-identification with the direct and reverse sparse representations and the combination of the associated residual errors. Sparse matrices and residual errors are in gray scale, the darker the square, the smaller the value.

3.2. Sparse coding of probe elements

The idea of collaborative sparse representation is to approximate an element of an unknown class by a sparse linear combination of a few reference elements that belong to different known classes. Reconstructing the test element only once with a sparse combination of elements from several classes forces reference elements to compete against each other and only the most similars are selected. The unknown class is associated to the class that participates the most.

In our person re-identification problem, we compute the sparse representation A_l^* of probe elements P_l using reference gallery dictionary G by solving the following optimization problem:

$$A_l^* = \arg\min_{A_l} ||P_l - GA_l||_F^2 + \lambda||A_l||_1 \qquad (1)$$

where $||.||_F$ is the Frobenius norm and $||.||_1$ is the $\mathcal{L}_1$-norm. $||P_l - GA_l||_F^2$ favors a small reconstruction error, $||A_l||_1$ enforces the sparsity of A_l, and λ balances the trade-off between the overall reconstruction error and the sparsity of the representation. This $\mathcal{L}_1$ penalized optimization problem is also known as the Lasso problem.

Let $err_{P_l,G_k,G} = ||P_l - G_k A_{l|k}^*||_F^2$ be the residual error of the reconstruction of probe elements P_l using only the gallery identity k where $A_{l|k}^*$ is the submatrix of A_l^* containing only the rows describing the participation of gallery dictionary G_k in the reconstruction of P_l. One way to rank gallery identities is by using their increasing residual error. The best match is given by:

$$k^* = \arg\min_k err_{P_l,G_k,G} \qquad (2)$$

3.3. Sparse coding of gallery elements

A probe person l being better reconstructed by a gallery person k than k' only means that k is more similar to the test person than k', it does not imply that the gallery person k is actually similar to the probe person. In order to build a robust correspondence between probe and gallery person, a reciprocal relation is needed. P_l must be well reconstructed by G_k when competing against other $G_{k'}$ and G_k must be well represented by P_l as well when competing with other $P_{l'}$. The additional optimization problem to be solved, is the optimization problem of Eq. (1) where the role of G and P are switched:

$$B_k^* = \arg\min_{B_k} ||G_k - PB_k||_F^2 + \gamma||B_k||_1 \qquad (3)$$

where γ is a trade-off parameter between the reconstruction error and the sparsity.

However, since in real world applications, not all identities probe images are available at a given testing time, it is not possible to use this perfectly symmetric approach to compute the sparse representation of gallery features with the complete probe dictionary and infer the residual errors. Instead, we propose to substitute in Eq. (3) the unavailable probe elements of $P = [P_1, \ldots, P_l, \ldots, P_K]$ with a generic dictionary D defined as described in section 3.5, *i.e.* P is replaced by $T(P_l, D) = [P_l, D]$. For every gallery identity k, we now solve the following optimization problem:

$$B_k^* = \arg\min_{B_k} ||G_k - [P_l, D]B_k||_F^2 + \gamma||B_k||_1 \qquad (4)$$

Let $err_{G_k,P_l,T(P_l,D)} = ||G_k - P_l B_{k|l}^*||_F^2$ be the residual error of the reconstruction of the gallery elements G_k

with the probe elements P_l where $B^*_{k|l}$ is the submatrix of B^*_k containing only the rows describing the participation of probe elements P_l in the reconstruction of G_k. These residual errors represent how well gallery elements G_k are reconstructed by the probe elements P_l when put in competition with the generic dictionary D. These residual errors alone can be used for ranking gallery identities. Probe person can be associated to the identity verifying:

$$k^* = \arg\min_k err_{G_k,P_l,T(P_l,D)} \quad (5)$$

3.4. Combination of both representations

In order to have a robust person re-identification system, we combine both sparse representations' residual errors: $err^{g\leftrightarrow p}_{k\leftrightarrow l} = err_{P_l,G_k,G} + err_{G_k,P_l,T(P_l,D)}$ so that the best match for probe id l is given by:

$$k^* = \arg\min_k err^{g\leftrightarrow p}_{k\leftrightarrow l} \quad (6)$$

We name SBSR, for Symmetric Bidirectional Sparse Representation, the method combining equations Eq. (1) and Eq. (3) that can be used in the ideal case when all the images of all the probe identities are available at testing time, in an offline setting for example. We call Asymmetric Bidirectional Sparse Representation (ABSR) the method combining equations Eq. (1) and Eq. (4) because the collaborative representations are not computed the same way for probe and gallery elements (the reverse representation requires an additional dictionary). ABSR can be used in an online context where test identities arrive one by one.

The strength of our method lies in its bidirectional aspect. It is particularly useful for elements that are subject to ambiguity. If probe elements P_l have about the same small residual errors for several gallery identities, the right match might be in the first ranks but not necessarily be the first choice. The extent of P_l's contribution to the reconstruction of the gallery dictionary elements gives additional information about the similarity between the probe and each gallery identity. Though the reverse sparse representation might also present ambiguities, it usually involves a slightly different small group of people in which case combining both sparse representation residual errors will favor commun elements. Combining both sparse representations helps leveraging ambiguities for similar people. This is illustrated in the overview of our method in Figure 1.

3.5. Choice of D

Given a test identity l captured with the probe camera, the idea of the reverse sparse representation is to find the gallery identities k for whose reconstruction the test identity l participates the most when competing against some other identities. We denote D_l the features associated to the identities that are competing against P_l.

In SBSR, all probe images are available at testing time, so P_l competes against $\{P_{l'}\}_{l'\in[|1,\dots,K|]\setminus l}$ in the reconstruction of gallery elements G, *i.e.*

$$D_l = [|P_1, \dots, P_{l-1}, P_{l+1}, \dots, P_K|] \quad (7)$$

When test elements arrive one by one, $P_{l'}$ is unavailable when testing identity l for $l' \neq l$, so another choice for D_l is needed. In our ABSR method, the choice we make is to form a generic dictionary D by selecting one image from the probe camera for each person in a known training set, *i.e.*

$$\forall l, D_l = D = [|P_{tr_1}, ...P_{tr_M}|] \quad (8)$$

where tr_i refers to the i^{th} identity of the training set which contains images from M people. With this choice, D should be varied enough to represent well every gallery identity. Features in D and P_l are extracted from the same camera, allowing a fair competition between identities. On the contrary, if D contains features extracted from gallery camera images, while P_l comes from the probe camera, the reconstruction of G will be biased, favoring images captured with the gallery camera which have similar color rendering, pose and viewpoint, rather than focusing on the difference between people.

4. Experimental results

4.1. Implementations details and feature extraction

Optimization. The optimizations problems presented in Eq. (1), Eq. (3) and Eq. (4) are $\mathcal{L}_1$-norm minimization problems. They can be solved using proximal algorithms. We used the SPAMS library [1].

Parameters. The parameters λ and γ have been set to 0.2 in all our experiments.

Features. Images are resized to 128×64 pixels and we extract state-of-the art LOMO features [20]. Distance metric learning method XQDA [20] is applied to training set features giving a symmetric definite positive matrix M that can be decomposed into $M = L^T L$. Our final base features are LOMO features projected into a lower dimensional space by the projection matrix L and normalized to unit $\mathcal{L}_2$-norm.

4.2. Datasets and evaluation protocol

CUHK03. The CUHK03 dataset is composed of two sets of 13164 images of 1360 people captured by 6 cameras in a university campus. One set of images forms the CUHK03 detected dataset, and the other set forms the CUHK03 labeled dataset. In the first case, the bounding boxes have

[1] spams-devel.gforge.inria.fr/downloads.html

been automatically found by a detector, while in the second case, the bounding boxes have been manually labeled. Each person is observed by one of the three pairs of cameras and has an average of 4.8 images in each camera view. We adopt the same 20 splitings as in [17] with 100 identities for testing and 1260 identities for training. However, instead of the single-shot test settings usually used, we adopt a multi-shot protocol using all the available images in both probe and gallery cameras as described in subsection 4.3.

iLIDS-VID. iLIDS-VID contains sequences of images of 300 people captured by 2 cameras in a busy airport, so the main difficulty comes from the occlusions. The sequences have variable length (22 to 192 images). We use 10 random splits of our data into training and testing sets, each containing images of 150 identities. We adopt a multi-shot protocol using all the images of the sequences.

4.3. Results.

We use CMC, Cumulative Match Curve, to evaluate the performance of our method. A CMC curve shows for every rank r the percentage of people well recognised at any rank smaller or equal to r.

Multi-shot protocol on CUHK03. Our bidirectional sparse representations approach is a multi-shot method, so instead of using the usual single-shot protocol on the CUHK03 dataset, we will compare our method with the multi-shot results that we computed using the XQDA approach [20] whose code is available online. The Figure 2 shows the CMC on the CUHK03 detected dataset for four testing protocols (SvsS, MvsS, SvsM and MvsM) using only the XQDA metric learning approach. The letters S and M stand for Single-shot and Multiple-shots. The first letter corresponds to images in the probe camera and the second one to those in the gallery camera. Thus MvsS (Multiple versus Single) means that for each person, several images from the probe camera are available but only one from the gallery camera is. In [20], the multi-shot protocol is not clearly specified, here we defined the distance between a probe identity and a gallery identity as the average distance between every pair of features from those two identities in the corresponding camera. The Figure 2 clearly shows how useful it is to have several images of the same person. With the SvsS protocol, the XQDA method yields 47.3% at first rank and up to 65.7% for the MvsM protocol.

Robustness of the ABSR approach. For CUHK03 detected, CUHK03 labeled and iLIDS-VID, we can see respectively in Tables 1, 2 and 3 that compared to simple metric learning XQDA, using Lasso sparse representation after the XQDA metric step, greatly improves first rank re-identification rate with a 4% or 5% raise for CUHK03 datasets and 9% for iLIDS-VID compared to XQDA only. Using our ABSR approach we further obtain a 3% increase

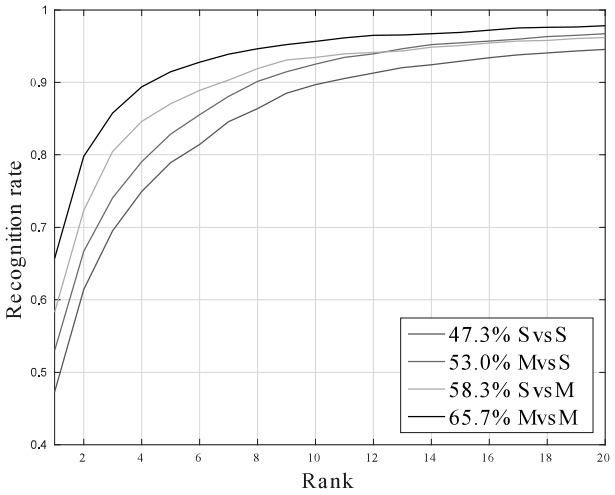

Figure 2. CMCs on CUHK03 detected. Results for different testing protocols: SvsS (Single vs Single), MvsS (Multi-shot in probe, Single-shot for gallery), SvsM (Single-shot in probe, Multi-shot in gallery), MvsM (Multi-shot in both probe and gallery).

Method	Rank 1	Rank 5	Rank 10	Rank 20
XQDA [20]	65.7	91.5	95.6	97.8
Lasso + XQDA	70.6	92.0	96.2	97.6
ABSR + XQDA	**74.1**	**92.5**	**96.6**	**98.1**
SBSR + XQDA	**77.8**	**94.5**	**97.6**	**98.9**

Table 1. CMC scores at ranks 1, 5, 10 and 20 on CUHK03 detected. Best and second best results are in bold. SBSR is in gray since it requires all testing images at testing time.

on all three datasets. Overall with a multi-shot protocol, our ABSR + XQDA method outperforms XQDA method alone by 8.4% on CUHK03 detected, by 6.9% on CUHK03 labeled and by 12.1% on iLIDS-VID. On iLIDS-VID, we also compared our method with recent state-of-the art papers. The results are reported from the papers [14, 19, 32, 34], we did not recompute their code on the same 10 random partitions that we used for our experiments, but since results are averaged over 10 random partitions, the results should be representative enough of the performances of each approach. Combining our bidirectional sparse representation matching method with metric learning XQDA, we clearly surpass AFDA [19] (37.5%) and MTL-LORAE [32] (43.0%) on iLIDS-VID dataset.

Influence of dictionary D. With our simple choice for D, Tables 1, 2 and 3 already show a 3% improvement for ABSR over Lasso for rank-1 recognition rate on all tested datasets. If all the probe images corresponding to the gallery identities are available at testing time, we can gain up to 7% at first rank with SBSR compared to Lasso. Table 4 presents experiments conducted on CUHK03 detected

Method	Rank 1	Rank 5	Rank 10	Rank 20
XQDA [20]	70.4	93.9	97.2	**98.8**
Lasso + XQDA	74.5	94.5	97.0	97.9
ABSR + XQDA	**77.3**	**95.2**	**97.4**	98.6
SBSR + XQDA	**81.4**	**96.5**	**98.1**	**98.7**

Table 2. CMC scores at ranks 1, 5, 10 and 20 on CUHK03 labeled. Best and second best results are in bold. SBSR can be applied only if all testing images are available at testing time, so it is in gray.

Method	Rank 1	Rank 5	Rank 10	Rank 20
DVDL* [14]	25.9	48.2	57.3	68.9
AFDA* [19]	37.5	62.7	73.3	81.8
DVR* [34]	39.5	61.1	71.7	81.0
MTL-LORAE*[32]	43.0	60.0	70.2	85.3
XQDA [20]	53.0	78.5	86.9	93.4
Lasso + XQDA	62.6	84.8	90.6	94.9
ABSR + XQDA	**65.1**	**86.8**	**92.9**	**96.6**
SBSR + XQDA	**68.5**	**87.9**	**93.0**	**96.3**

Table 3. CMC scores at ranks 1, 5, 10 and 20 on iLIDS-VID. Best and second best results are in bold. SBSR is in gray since it can be applied only if all testing images are available at testing time. * refers to results reproduced from the quoted papers, they might not have been computed with the same partitions.

nIds	50	75	100	200	300	400	500
rank 1	72.3	72.7	73.1	73.9	74.0	74.0	74.2
nIds	600	700	800	900	1000	1100	1200
rank 1	**74.3**	74.2	74.0	74.2	74.0	74.1	74.0

Table 4. Comparison of first rank recognition rate on CUHK03 detected for different number of training identities represented in the generic dictionary D.

where the number of identities in the dictionary D varies. For each partition, the tests are computed for 10 random selections of s identities out of 1260 training identities, with s varying from 50 to 1260. Randomly selecting only 200 identities already seems to make D a generic dictionary.

Influence of features. The results discussed so far involved LOMO features on which a supervised transformation had already been performed. To assess the effectiveness of our approach, we also tested it with raw LOMO features, another feature and its XQDA transformed version. We chose a simple feature: HS (8×8 bins) and RGB ($4 \times 4 \times 4$ bins) histograms are extracted from stripes of 16 pixels height with an overlap of 8 pixels and concatenated, resulting in a 1960 dimensional feature. The results for CUHK03 detected are summarized in Figure 3 and Table 1. The same trend can be observed for every tested feature: ABSR performs better than Lasso. For raw LOMO features, the first rank recognition for Euclidean distance is

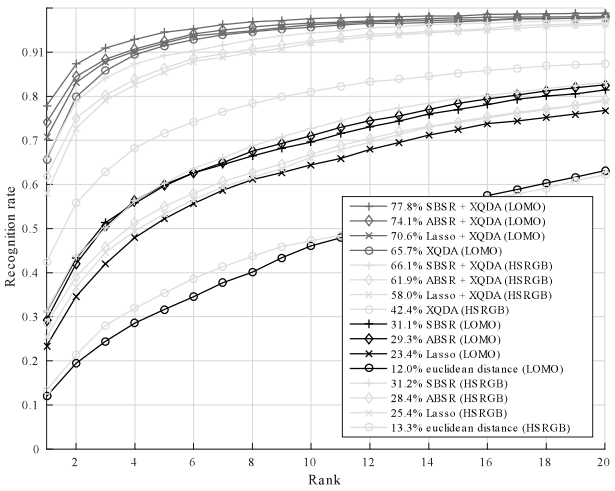

Figure 3. CMCs on CUHK03 detected for the MvsM protocol. Comparison of Euclidean distance, Lasso, ABSR, SBSR for LOMO and HSRGB features and their XQDA projected version.

only 12.0%, while using Lasso, it is raised up to 23.4%, and with our ABSR method, it becomes 29.3%. Even if using only XQDA metric learning with LOMO features already achieves 65.7% recognition rate at first rank, combining it with ABSR further increases performances to 74.1%.

5. Conclusions

In this paper, we presented two variants of our bidirectional sparse representation method for person re-identification: a symmetric one (SBSR) and an asymmetric one (ABSR). These approaches aim at improving the matching step in the person re-identification task, and can be applied with any feature. The idea is to exploit the robustness of sparse coding using a bidirectional scheme. We solve the Lasso problem in the traditional way by reconstructing probe elements with gallery dictionary, but also in the other way round by reconstructing gallery elements with probe elements completed by a generic dictionary. We showed with two different raw features that our BSR methods indeed improve performances by raising the first rank recognition by more than 15% compared to Euclidean distance. When applied to state-of the art-feature LOMO combined with the XQDA metric learning step, we also manage to get a 8.4% increase in recognition rate compared to XQDA alone. Overall, our method outperforms state-of-the art methods for multi-shot protocols on both CUHK03 and iLIDS-VID datasets with a first rank recognition rate of 74.1% for CUHK03 detected, 77.3% for CUHK labeled and 65.1% for iLIDS-VID. Future work will focus on learning a better generic dictionary and on tackling the much harder problem of open-world person re-identification.

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[2] L. An, X. Chen, S. Yang, and B. Bhanu. Sparse representation matching for person re-identification. *Information Sciences*, 355356, 2016.

[3] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch. Learning implicit transfer for person re-identification. In *European Conference on Computer Vision Workshops and Demonstrations*, 2012.

[4] S. Bak, E. Corvee, F. Brémond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. In *The IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010.

[5] B. Cancela, T. M. Hospedales, and S. Gong. Open-world person re-identification by multi-label assignment inference. In *The British Machine Vision Conference*, 2014.

[6] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *The British Machine Vision Conference*, 2011.

[7] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *Asian conference on Computer vision*, 2010.

[8] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. *The IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008.

[10] M. Hirzer, P. M. Roth, and H. Bischof. Person re-identification by efficient impostor-based metric learning. In *The IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2012.

[11] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *European Conference on Computer Vision*, 2012.

[12] X.-Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[13] S. Karanam, Y. Li, and R. Radke. Sparse re-id: Block sparsity for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.

[14] S. Karanam, Y. Li, and R. J. Radke. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In *The IEEE International Conference on Computer Vision*, December 2015.

[15] E. Kodirov, T. Xiang, and S. Gong. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In M. W. J. Xianghua Xie and G. K. L. Tam, editors, *The British Machine Vision Conference*, 2015.

[16] R. Layne, T. M. Hospedales, S. Gong, et al. Person re-identification by attributes. In *The British Machine Vision Conference*, 2012.

[17] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[18] X. Li, A. Wu, M. Cao, J. You, and W.-S. Zheng. Towards more reliable matching for person re-identification. In *The IEEE International Conference on Identity, Security and Behavior Analysis*, 2015.

[19] Y. Li, Z. Wu, S. Karanam, and R. J. Radke. Multi-shot human re-identification using adaptive fisher discriminant analysis. In M. W. J. Xianghua Xie and G. K. L. Tam, editors, *The British Machine Vision Conference*, 2015.

[20] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[21] S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *The IEEE International Conference on Computer Vision*, 2015.

[22] V. E. Liong, J. Lu, and Y. Ge. Regularized local metric learning for person re-identification. *Pattern Recognition Letters*, 68, 2015.

[23] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *IEEE transactions on pattern analysis and machine intelligence*, 2015.

[24] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: what features are important? In *European Conference on Computer Vision Workshops and Demonstrations*, 2012.

[25] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-supervised coupled dictionary learning for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[26] X. Liu, M. Song, Q. Zhao, D. Tao, C. Chen, and J. Bu. Attribute-restricted latent topic model for person re-identification. *Pattern recognition*, 2012.

[27] B. Ma, Y. Su, F. Jurie, et al. Bicov: a novel image representation for person re-identification and face verification. In *British Machive Vision Conference*, 2012.

[28] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[29] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[30] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *BMVC*, 2008.

[31] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[32] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao. Multi-task learning with low rank attribute embedding for person re-identification. In *The IEEE International Conference on Computer Vision*, December 2015.

[33] S. Wang, M. Lewandowski, J. Annesley, and J. Orwell. Re-identification of pedestrians with variable occlusion and scale. In *The IEEE International Conference on Computer Vision Workshops*, 2011.

[34] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by discriminative selection in video ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[35] D. Yi, Z. Lei, and S. Z. Li. Deep metric learning for practical person re-identification. *arXiv preprint arXiv:1407.4979*, 2014.

[36] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *Computer Vision and Pattern Recognition*, 2015.

[37] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[38] W.-S. Zheng, S. Gong, and T. Xiang. Towards open-world person re-identification by one-shot group-based verification. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2016.

[39] W.-S. Zheng, X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong. Partial person re-identification. In *The IEEE International Conference on Computer Vision*, 2015.