TOWARDS UNSUPERVISED OPEN-SET PERSON RE-IDENTIFICATION

Hanxiao Wang, Xiatian Zhu, Tao Xiang, Shaogang Gong

Queen Mary University of London

ABSTRACT

Most existing person re-identification (ReID) methods assume the availability of extensively labelled cross-view person pairs and a closed-set scenario (i.e. all the probe people exist in the gallery set). These two assumptions significantly limit their usefulness and scalability in real-world applications, particularly with large scale camera networks. To overcome the limitations, we introduce a more challenging yet realistic ReID setting termed *OneShot-OpenSet-ReID*, and propose a novel Regularised Kernel Subspace Learning model for ReID under this setting. Our model differs significantly from existing ReID methods due to its ability of effectively learning cross-view identity-specific information from unlabelled data alone, and its flexibility of naturally accommodating pairwise labels if available.

Index Terms— Person Re-identification, Open-set Recognition, Unsupervised Subspace Learning, Kernelisation

1. INTRODUCTION

For making sense of the huge quantity of video data produced by large scale surveillance camera networks in public spaces, automatically identifying and re-identifying specific people across non-overlapping cameras at different physical locations is essential. This task is known as *person re-identification* (ReID). Person ReID by visual matching is inherently challenging [1]. This is due to the great visual similarity among different people, and significant appearance variations of the same people across cameras caused by the cross-view large disparity in viewpoint, illumination, occlusion and background clutter.

To build an automated ReID system, most existing methods are *fully supervised* and require a large amount of exhaustively labelled cross-view matching person pairs for model training [2, 3, 4, 5, 6], which is impractically expensive and not scalable. To reduce the labelling cost, alternative methods attempt to exploit unlabelled data for *semi-supervised learning* [7, 8], or *unsupervised learning* [9, 10, 11, 12, 13]. However, the former still requires labelled cross-view pairs, whereas the latter is unable to exploit cross-view identity discriminative information and thus yields deficient ReID performances. Moreover, most existing ReID methods above assume a *closed-set* setting, i.e. probe population and gallery

population are the same, which makes them unsuitable for real-world open-set applications.

Under an open-set setting, the search scope (probe set) can be much larger than the target people (gallery set). For example, even in a space of moderate size (e.g. an underground station), there could easily be hundreds or even thousands of people passing through within an hour. Moreover, not all target people necessarily appear in the probe camera view, due to, e.g., the complex topology structure of the camera network. To further compound the problem, there may be only *a single shot per person* for target people due to low frame rates of surveillance videos, thus offering insufficient data to learn the appearance variations for each person.

The aim of this work is thus to design an intelligent ReID system for re-identifying the target people on the watch list (i.e. gallery set) from a large volume of video footages without any cross-view manual labelling. We call person ReID under this real-world setting the *OneShot-OpenSet-ReID* (OS²ReID) problem. Such a problem setting raises two requirements for ReID models: (1) to learn from unlabelled data, the only source to obtain cross-view identity-discriminative information; and (2) to perform matching under an open-set assumption, i.e. the probe people are not guaranteed to have a match in the small galley set which contains a single shot for each person. The concept of open-set ReID has started attracting attention recently [14, 15, 16]. Nevertheless, they all assume a fully pairwise labelled closed-world training set and thus none is suitable for the proposed OS²ReID problem.

The **contributions** of this work are: **(I)** We introduce a new and more realistic unsupervised ReID setting called OneShot-OpenSet-ReID (OS²ReID) which differs significantly from the existing closed-set ReID settings and supervised open-set ReID settings. **(II)** We propose a new unsupervised subspace learning model (named as Regularised Kernel Subspace Learning, or RKSL) specifically suitable to the OS²ReID problem. Importantly, it is optimised by a closed-form eigen-problem solution instead of iterative approximations, making it suitable for large scale and real-time applications. **(III)** We further extend our RKSL model for semi-supervised learning, to accommodate any labelled data if available. Extensive experiments on two benchmark datasets VIPeR[17] and CUHK01[18] over five existing models show the advantages of the RKSL model

Fig. 1. Intuition of our cross-view constraint. The unlabelled cross-view data (the left and right pairs) encode information on cross-view appearance variations, e.g. changes in illumination and viewpoint respectively. This subtle information is exploited effectively by the proposed RKSL model for re-identifying the truly matched cross-view people (the middle pair).

over state-of-the-art methods.

2. METHODOLOGY

2.1. Problem Definition

We first define the newly introduced OneShot-OpenSet-ReID (OS²ReID) problem. Suppose one only has unlabelled (in a pairwise inter-camera sense) images of people, including a one-shot watch list of target people G (gallery) from view $\mathcal{X}$ and a larger pool of probe people P from view $\mathcal{Y}$. Given a probe image from P, the objective is to determine: (a) whether it matches anyone in the gallery, and (b) if yes, which one. We treat the above problem as an unsupervised subspace learning task, and formulate a *Regularised Kernel Subspace Learning* (RKSL) model aiming to learn a latent shared cross-view subspace where data pairs of same identities are closer to each other, and those of different identities are further-apart. Specifically, we propose to mine two types of information from unlabelled data described below.

Cross-view appearance variation: Such a variation is caused by unknown changes in viewpoint and illumination across camera networks, prohibiting truly matched target images being correctly found. To model such variations, we have the following observation: If a cross-view image pair share *very* similar appearances, they often capture some cross-view information w.r.t. viewing condition changes, regardless of their identities (Fig. 1). By enforcing such pairs to be closer in the shared subspace, the cross-view appearance variation can be potentially learned.

Within-view appearance similarity: Similar appearances of different people within the same camera view often confuse ReID systems. To address this problem, we add another model learning constraint by reinforcing two visually similar people in the same camera view to be further away in the common subspace, assuming that a person is uniquely represented by a single image, i.e. the one-shot ReID setting where only one image is available for every person in the gallery set without any explicit labelling for the images.

2.2. Model Formulation

In our subspace learning formulation, each data sample represented in a visual feature space $\mathcal{F}$ is projected to a subspace $\mathcal{P}$. The projection is realised by two projection vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ respectively, the model parameters to be learned.

/* *Variables associated with view $\mathcal{X}$ (similar for view $\mathcal{Y}$)* */:
  $\{\breve{\mathbf{x}}_i\}_{i=1}^{n_u}$: Unlabelled data, with feature matrix $\breve{X}$;
  $\{\bar{\mathbf{x}}_i\}_{i=1}^{n_l}$: Labelled data, with feature matrix $\bar{X}$;
  $\{\hat{\mathbf{x}}_i\}_{i=1}^{n_x} = \{\breve{\mathbf{x}}_i\}_{i=1}^{n_u} \cup \{\bar{\mathbf{x}}_i\}_{i=1}^{n_l}$, and $\hat{X} = [\bar{X}; \breve{X}]$;
  $\mathcal{L}_{\hat{x}}$: Graph Laplacian matrix;
  $\mathbf{w}_x$: Projection vectors (model parameters);
  α: Kernelised projection vectors;
/* *Variables across view $\mathcal{X}$ and $\mathcal{Y}$* */:
  $\mathcal{S}_{ij}$: Similarity measure between $\breve{\mathbf{x}}_i$ and $\breve{\mathbf{y}}_j$
/* *Others* */:
  K: Kernel matrix on data, further clarified by subscript.

Fig. 2. Definition of notations.

Formally, with the two constraints described above we formulate our objective function as follows:

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^\top (\sum_{i,j} \mathcal{S}_{ij} \cdot \breve{\mathbf{x}}_i \breve{\mathbf{y}}_j^\top) \mathbf{w}_y}{\sqrt{\mathbf{w}_x^\top (C_{\hat{x}\hat{x}} + R_{\hat{x}}) \mathbf{w}_x \, \mathbf{w}_y^\top C_{\hat{y}\hat{y}} \mathbf{w}_y}} \quad (1)$$

with
$$C_{\hat{x}\hat{x}} = \hat{X}^\top \hat{X} \quad (2) \qquad R_{\hat{x}} = -\frac{\gamma_x}{n_x^2} \hat{X}^\top \mathcal{L}_{\hat{x}} \hat{X} \quad (3)$$
$$C_{\hat{y}\hat{y}} = \hat{Y}^\top \hat{Y}$$

where $C_{\hat{x}\hat{x}}$ and $C_{\hat{y}\hat{y}}$ are within-view covariance matrices. Other notations are given in Fig. 2. Note that in the OS²ReID setting, $\hat{X} = \breve{X}$, and $\hat{Y} = \breve{Y}$ since no cross-view labelled data is available, i.e. $\bar{X} = \bar{Y} = \emptyset$.

In Eq. (1), the nominator $\mathcal{B} = \mathbf{w}_x^\top (\sum_{i,j} \mathcal{S}_{ij} \cdot \breve{\mathbf{x}}_i \breve{\mathbf{y}}_j^\top) \mathbf{w}_y$ enforces a cross-view unlabelled data pair $\{\breve{\mathbf{x}}_i, \breve{\mathbf{y}}_j\}$ in $\mathcal{P}$ to be closer in the projected space if their similarity $\mathcal{S}_{ij}$ in $\mathcal{F}$ is high. The value of $\mathcal{S}_{ij}$ can be set by either learning or non-learning based methods as detailed in Sec. 3. On the other hand, $R_{\hat{x}}$ in the denominator of Eq. (1) represents the within-view similarity constraint, which regularises $\mathbf{w}_x$ so that *within-view* ($\mathcal{X}$) visually similar image data are pulled apart in the subspace. Formally, we denote $A_{\hat{x}}$ as a k-Nearest-Neighbour (kNN) similarity graph on $\hat{X}$. Then we have:

$$\mathbf{w}_x^\top \hat{X}^\top \mathcal{L}_{\hat{x}} \hat{X} \mathbf{w}_x = \frac{1}{2} \sum_{i,j=1}^{n_x} (\mathbf{w}_x^\top \hat{\mathbf{x}}_i - \mathbf{w}_x^\top \hat{\mathbf{x}}_j)^2 A_{\hat{x}}^{ij} \quad (4)$$

where $\mathcal{L}_{\hat{x}}$ is the graph Laplacian [19] of $A_{\hat{x}}$. By appending its additive inverse onto the denominator of Eq. (1), we enforce the adjacent samples in $\mathcal{F}$ to be more separated in $\mathcal{P}$, and in return make the projection $\mathbf{w}_x$ more identity discriminative.

2.3. Accommodating Cross-View Pairwise Labels

To further extend our model flexibility, let us now consider the situation where some labelled cross-view pairs are available. To that end, we introduce a third regularisation term to represent pairwise labelled information by expanding Eq. (1):

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\mathbf{w}_x^\top (\sum_k^{n_l} \bar{\mathbf{x}}_k \bar{\mathbf{y}}_k^\top) \mathbf{w}_y + \eta \cdot \mathbf{w}_x^\top (\sum_{i,j} \mathcal{S}_{ij} \cdot \breve{\mathbf{x}}_i \breve{\mathbf{y}}_j^\top) \mathbf{w}_y}{\sqrt{\mathbf{w}_x^\top (C_{\hat{x}\hat{x}} + R_{\hat{x}}) \mathbf{w}_x \, \mathbf{w}_y^\top (C_{\hat{y}\hat{y}} + R_{\bar{y}}) \mathbf{w}_y}} \quad (5)$$

where $\mathbf{w}_x^\top (\sum_k^{n_l} \bar{\mathbf{x}}_k \bar{\mathbf{y}}_k^\top) \mathbf{w}_y$ is the new regularisation term for encoding the *labelled* cross-view data pairs. The coefficient η is a balancing parameter for controlling the trade-off between the *hard* labelled and *soft* unlabelled cross-view correspondences. Note that we also introduce the regularisation term $R_{\bar{y}} = -\frac{\gamma_y}{n_y^2} \bar{Y}^\top \mathcal{L}_{\bar{y}} \bar{Y}$ for the probe set data whose identities are known to be different by available labels.

2.4. Kernelisation

Given complex variations in viewing condition across cameras, the optimal subspace may not be obtainable by linear projections. We thus further kernelise Eq. (5) by projecting the data into a reproducing kernel Hilbert space $\mathcal{H}$:

$$\rho = \max_{\alpha,\beta} \frac{\alpha^\top K_{\hat{x}\bar{x}}K_{\bar{y}\hat{y}}\beta + \eta \cdot \alpha^\top (\sum_{i,j} \mathcal{S}_{ij} \cdot K_{\hat{x}\breve{x}_i}K_{\breve{y}_j\hat{y}})\beta}{\sqrt{\alpha^\top(K_{\hat{x}\hat{x}}^2 + \mathcal{R}_{\hat{x}})\alpha \, \beta^\top(K_{\hat{y}\hat{y}}^2 + \mathcal{R}_{\bar{y}})\beta}} \quad (6)$$

where α and β are the kernelised projection vectors for the two views respectively, and the kernelised $\mathcal{R}_{\hat{x}}$ and $\mathcal{R}_{\bar{y}}$ are:

$$\mathcal{R}_{\hat{x}} = \epsilon_x K_{\hat{x}\hat{x}} - \frac{\gamma_x}{n_x^2}K_{\hat{x}\hat{x}}\mathcal{L}_{\hat{x}}K_{\hat{x}\hat{x}}, \quad \mathcal{R}_{\bar{y}} = \epsilon_y K_{\hat{y}\hat{y}} - \frac{\gamma_y}{n_y^2}K_{\hat{y}\hat{y}}\mathcal{L}_{\bar{y}}K_{\bar{y}\hat{y}}$$

We use $\epsilon_x K_{\hat{x}\hat{x}}$ and $\epsilon_y K_{\hat{y}\hat{y}}$ to penalise the norms of the respective projection vectors, which is equivalent to Tikhonov regularization. In this work, we set both ϵ_x and ϵ_y to the standard value of 0.5, and utilised the exponential chi-square kernel function. The optimisation of (6) can be performed through a generalised eigen-problem similar to [20, 21] after a simple Lagrangian transformation.

Person ReID: Our RKSL model can be applied to both the proposed OS^{2}ReID and the conventional closed-set ReID scenarios. Specifically, given the watch list (gallery set) G and the probe set P, we first obtain their representations in the projected space $\mathcal{P}$ by applying the proposed RKSL (Eq. (6), with or without labels). Then, we directly use the projected data points in $\mathcal{P}$ to perform ReID with cosine distance [22] as the matching function.

3. EXPERIMENTS

Datasets: Two large benchmark datasets VIPeR [17] and CUHK01 [18], were selected for the evaluation. The VIPeR dataset contains a total of 632 people with one image per person per view, whilst the CUHK01 dataset has 971 people with two images per person per view.

Visual features: We applied the histogram-based image descriptor introduced in [23]. The feature vector (5138 dimensions) is a concatenation of colour, HOG [24], and LBP [25] histograms extracted on horizontally segmented stripes.

3.1. OneShot-OpenSet ReID Evaluation

Under the OneShot-OpenSet (OS^{2}ReID) setting, there is no cross-view labelled data pair available.

Settings: For both datasets, we created gallery set G by randomly selecting 120 people from one camera view, and the probe set P by selecting half of the whole population (316 on VIPeR and 486 on CUHK01) from the other view, with the condition that 100 people exist in both G and P. For gallery set G, only one-shot image per person is included. We evaluated a total of 10 folds and reported their averaged results.

Competitors: We compared the proposed RKSL model against five contemporary and state-of-the-art methods: (1) L_1-norm[1]: a basic distance metric. (2-4) Regularised Dictionary Learning (RDL) [13], SDALF [9], SDC [11]: three most

---
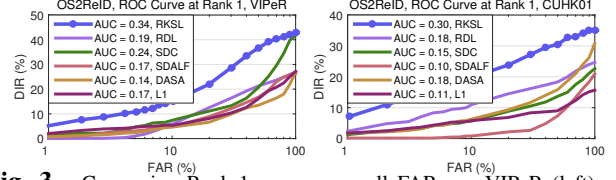[1] We found that L_2-norm distance gave almost identical results.

Fig. 3. Comparing Rank-1 scores over all FARs on VIPeR (left) and CUHK01 (right) under the OneShot-OpenSet ReID setting

notable state-of-the-art unsupervised ReID models with code provided by authors. (5) DASA [26]: a recent unsupervised domain adaptation model. In ReID context, each person is considered as a class, and each camera view as a domain.

Evaluation metric: We utilised the ROC curves of False Accept Rate (FAR) versus Detection and Identification Rate (DIR) for performance evaluation [14]. Note that DIR becomes the Cumulated Matching Characteristics used for the conventional closed-world setting, when FAR = 100%.

Implementation details: We set $\gamma_x/n_x^2 = 0.02$ (Eq. (3)) empirically. For computing $\mathcal{S}_{ij}$, we simply used the additive inverse of the cross-view bipartite kNN graph with L_2 distances, and normalised its values to a range $[0, 1]$. For both cross/within-view kNN graphs, the value of k was set to 15.

Dataset	VIPeR [17]				CUHK01 [18]			
FAR (%)	1	10	50	100	1	10	50	100
L_1-norm	1.9	5.4	16.1	27.2	1.8	5.8	9.0	15.7
DASA [26]	0.6	4.7	13.5	27.0	1.8	6.3	15.8	30.8
SDALF [9]	0.7	4.5	16.6	26.9	0.2	1.2	8.0	21.7
SDC [11]	1.7	7.3	21.5	41.5	1.2	5.8	14.0	23.3
RDL [13]	0.1	6.0	21.3	26.3	2.5	11.0	18.3	24.7
RKSL (Ours)	**4.9**	**15.1**	**36.7**	**42.9**	**7.5**	**20.2**	**32.0**	**36.0**

Table 1. Comparing Rank-1 scores of different methods at varying FARs under the OneShot-OpenSet-ReID setting

Comparative results: It is evident from Fig. 3 and Table 1 that the proposed RKSL model significantly outperforms all the competitors on both datasets, especially with demanding (small) FARs. Particularly, when compared to the second best method (SDC on VIPeR and RDL on CUHK01) at FAR = 10%, the Rank-1 score is doubled (from 7.3 to 15.1) on VIPeR and tripled (from 11.0 to 20.2) on CUHK01 by RKSL. This demonstrates the effectiveness of the proposed kernel subspace learning model in extracting identity-sensitive information from the unlabelled data.

Existing state-of-the-art unsupervised ReID models, RDL, SDC and SDALF, are shown to be less effective for the new OS^{2}ReID setting, compared to RKSL which is tailored specifically for unsupervised open-set ReID. It can also be observed that DASA is much inferior in matching people across views, which suggests that it is difficult for the unsupervised domain adaptation approach to solve this OS^{2}ReID problem where the intrinsic discriminative information can be more subtle and more challenging to extract than in the general object recognition/categorisation problem, especially when their assumption on the two domains containing the same set of classes becomes invalid under the proposed OS^{2}ReID setting.
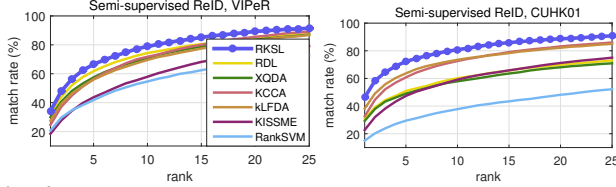
Fig. 4. Comparing the CMC performances on VIPeR (left) and CUHK01 (right) under the semi-supervised ReID setting

3.2. Semi-Supervised ReID Evaluation

We also evaluated the ability the proposed RKSL model to accommodate labelled data, by applying the RKSL model in the conventional semi-supervised settings where some (sparse) cross-view labelled pairs are available.

Settings: We adopted the same setting as in [8, 13]. Specifically, for both datasets, we equally split the whole dataset into two partitions: one for training and the other for testing. One third of the training partition are cross-view pairwise labelled.

Competitors: We compared our RKSL model with two most resent semi-supervised methods, SSCDL [8] and RDL [13], and five most contemporary fully-supervised models including, RankSVM [27], KISSME [2], kLFDA [5], KCCA [23], and XQDA [6]. The same visual feature was adopted for all methods with author published codes, except SSCDL which does not have code available so their reported results were compared. We applied the exponential chi-square kernel in kLFDA [5] and KCCA [23], the same as in RKSL.

Evaluation metric: The conventional Cumulated Matching Characteristics (CMC) curves were utilised for quantitative comparison between different methods.

Implementation details: We used cross-validation to determine the free parameters $(\eta, \gamma_x, \gamma_y)$ for the proposed RKSL model, as well as parameters of all the baseline methods. For computing $\mathcal{S}_{ij}$, since some sparse labels are available, we thus first used the labelled pairs to learn an initial subspace, and then used the distance in this subspace to calculate the similarity over unlabelled data and re-learned a final subspace.

Dataset	VIPeR [17]				CUHK01 [18]			
Ranks (%)	1	5	10	20	1	5	10	20
RankSVM [27]	20.7	41.8	54.6	68.2	15.0	29.4	37.8	48.2
KISSME [2]	18.5	43.7	57.9	74.5	22.7	47.4	59.1	71.2
kLFDA [5]	27.5	56.0	69.6	82.6	38.3	63.7	73.5	82.2
KCCA [23]	24.6	56.2	71.7	85.6	32.6	60.8	72.6	83.2
XQDA [6]	30.0	57.5	70.9	83.5	29.3	49.0	57.9	68.0
RDL [13]	32.5	61.8	74.3	84.1	31.0	50.9	60.2	69.8
SSCDL [8]	25.6	53.7	68.1	83.6	-	-	-	-
RKSL (Ours)	**34.2**	**66.6**	**78.9**	**89.3**	**46.3**	**72.3**	**80.8**	**88.7**

Table 2. Comparing some recognition rates of different methods on VIPeR and CUHK01. The semi-supervised ReID setting.

Comparative results: From Table 2, it is observed that the proposed RKSL model significantly outperforms all baseline methods. Specifically, RKSL provides much better ReID accuracy than the state-of-the-art models RDL and SSCDL, particularly on CUHK01. The plausible reason for the weak performance of RDL [13] on the larger CUHK01 dataset is that, the error propagation from the regularisation term is made

more severe during the iterative learning process due to many visually-similar people. In contrast, we utilise available labels to initialise the subspace learning for better exploiting unlabelled data, i.e. obtaining more accurate cross-view similarity measure, without any chance to mislead model learning. It is also observed that all fully-supervised models yield much worse recognition results than RKSL. For example, on VIPeR, our RKSL improves the Rank-1 score over RankSVM by 13.51%, KISSME by 15.73%, kLFDA by 6.68%, KCCA by 9.59%, and XQDA by 4.20%. Even larger improvements are gained on CUHK01. The main reason for inferior performance by these supervised methods is the limited availability of labelled data and their inability of exploiting the large quantity of unlabelled data. Whilst the proposed RKSL model can effectively utilise both in a unified way, largely relaxing the stringent assumption on labelled data amount and making it flexible in coping with varying amounts of data annotation.

Effects of labelled data sparsity: For evaluating the performance given different amount of data annotation, we further conducted a set of experiments on VIPeR by comparing RKSL with two state-of-art baselines, kLFDA [5] and KCCA [23], when different numbers of labelled pairs are provided. To this end, we changed the labelled data percentage from 10% to 100% and compared their performances on ranks 1, 5, 10. The results in Fig. 5 show that the accuracies achieved by the proposed RKSL model are significantly better at all three ranks, compared to the two baselines. The margins are evidently larger when fewer labelled data are available, which further suggests the effectiveness of our RKSL in exploiting unlabelled data for person-discriminative subspace learning. This further demonstrates the strength and flexibility of our model under a large spectrum of settings.
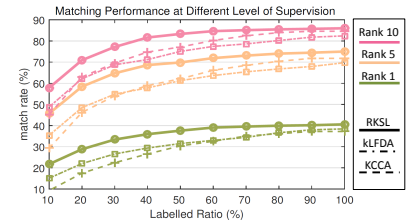

Fig. 5. Semi-supervised matching on VIPeR (matching rate as a function of labelled data percentage)

4. CONCLUSION

We have presented a new and more realistic person ReID setting: *OneShot-OpenSet-ReID* (OS^{2}ReID). To solve the OS^{2}ReID problem, a novel *Regularised Kernel Subspace Learning* (RKSL) model is proposed. The model is unique due to its capability of learning cross-view identity discriminative information from unlabelled data. This makes RKSL readily applicable and scalable to large scale open-set ReID problems. Also, the RKSL model allows to effectively exploit pairwise labels when available. Extensive comparative evaluations were conducted to validate the advantages of the proposed model under both OS^{2}ReID (no pairwise labels) and conventional (with labels and closed-set) settings.

5. REFERENCES

[1] Shaogang Gong, Marco Cristani, Change Loy Chen, and Timothy M. Hospedales, "The re-identification challenge," in *Person Re-Identification*. Springer, 2014.

[2] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, 2012.

[3] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Re-identification by relative distance comparison," *TPAMI*, 2013.

[4] Sateesh Pedagadi, James Orwell, Sergio A. Velastin, and Boghos A. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification.," in *CVPR*, 2013.

[5] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier, "Person re-identification using kernel-based metric learning methods," in *ECCV*. Springer, 2014.

[6] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015.

[7] Chunxiao Liu, Chen Change Loy, Shaogang Gong, and Guijin Wang, "Pop: Person re-identification post-rank optimisation," in *ICCV*, 2013.

[8] Xiao Liu, Mingli Song, Dacheng Tao, Xingchen Zhou, Chun Chen, and Jiajun Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *CVPR*, 2014.

[9] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *CVPR*, 2010.

[10] Bingpeng Ma, Yu Su, and Frédéric Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *BMVC*, 2012.

[11] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Unsupervised salience learning for person re-identification," in *CVPR*, 2013.

[12] Hanxiao Wang, Shaogang Gong, and Tao Xiang, "Unsupervised learning of generative topic saliency for person re-identification," in *BMVC*, 2014.

[13] Elyor Kodirov, Tao Xiang, and Shaogang Gong, "Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification," in *BMVC*, 2015.

[14] Shengcai Liao, Zhipeng Mo, Yang Hu, and Stan Z. Li, "Open-set person re-identification," *CoRR*, vol. abs/1408.0872, 2014.

[15] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Towards open-world person re-identification by one-shot group-based verification," *TPAMI*, vol. PP, no. 99, July 2015.

[16] Brais Cancela, Tim Hospedales, and Shaogang Gong, "Open-world person re-identification by multi-label assignment inference," in *BMVC*, 2014.

[17] Douglas Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007.

[18] Wei Li and Xiaogang Wang, "Locally aligned feature transforms across views.," in *CVPR*, 2013.

[19] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *JMLR*, 2006.

[20] David R. Hardoon, Sandor Szedmak, Or Szedmak, and John Shawe-taylor, "Canonical correlation analysis; an overview with application to learning methods," 2007.

[21] M. B. Blaschko, J. A. Shelton, A. Bartels, C. H. Lampert, and A. Gretton, "Semi-supervised kernel canonical correlation analysis with application to human fMRI," *PR*, 2011.

[22] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[23] Giuseppe Lisanti, Iacopo Masi, and Alberto Del Bimbo, "Matching people across camera views using kernel canonical correlation analysis," in *Proceedings of the International Conference on Distributed Smart Cameras*, 2014.

[24] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan, "An hog-lbp human detector with partial occlusion handling.," in *ICCV*, 2009.

[25] Timo Ahonen, Abdenour Hadid, and Matti Pietikinen, "Face description with local binary patterns: Application to face recognition," *TPAMI*, vol. 28, pp. 2037–2041, 2006.

[26] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *ICCV*, 2013.

[27] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by support vector ranking," in *BMVC*, 2010.