Noname manuscript No.
(will be inserted by the editor)

Person Re-Identification in Identity Regression Space

Hanxiao Wang · Xiatian Zhu · Shaogang Gong · Tao Xiang

Received: date / Accepted: date

Abstract Most existing person re-identification (re-id) methods are unsuitable for real-world deployment due to two reasons: *Unscalability to large population size*, and *Inadaptability over time*. In this work, we present a unified solution to address both problems. Specifically, we propose to construct an Identity Regression Space (IRS) based on embedding different training person identities (classes) and formulate re-id as a regression problem solved by identity regression in the IRS. The IRS approach is characterised by a closed-form solution with high learning efficiency and an inherent incremental learning capability with human-in-the-loop. Extensive experiments on four benchmarking datasets (VIPeR, CUHK01, CUHK03 and Market-1501) show that the IRS model not only outperforms state-of-the-art re-id methods, but also is more scalable to large re-id population size by rapidly updating model and actively selecting informative samples with reduced human labelling effort.

Keywords Person Re-Identification · Feature Embedding Space · Regression · Incremental Learning · Active Learning

1 Introduction

Person re-identification (re-id) aims to match identity classes of person images captured under non-overlapping camera views (Gong et al, 2014). It is inherently challenging due to significant cross-view appearance changes (Fig. 1(a)) and high visual similarity among different people (Fig. 1(b)).

Hanxiao Wang is with Electrical and Computer Engineering Department, Boston University, Boston MA 02215, USA. E-mail: hxw@bu.edu.
Xiatian Zhu is with Vision Semantics Limited, London E1 4NS, UK. E-mail: eddy@visionsemantics.com.
Shaogang Gong and Tao Xiang are with School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK. E-mail: {s.gong, t.xiang}@qmul.ac.uk.

(a) Significant person appearance change across camera views.

(b) High visual similarity among different people.

Fig. 1 Illustration of person re-identification challenges.

Most existing re-id methods focus on designing identity discriminative features and matching models for reducing intra-person appearance disparity whilst increasing inter-person appearance individuality. This is often formulated as a supervised learning problem through classification (Koestinger et al, 2012; Liao et al, 2015), pairwise verification (Li et al, 2014; Shi et al, 2016), triplet ranking (Zheng et al, 2013; Wang et al, 2016d), or a combination thereof (Wang et al, 2016a). While achieving ever-increasing re-id performance on benchmarking datasets (Zheng et al, 2016; Karanam et al, 2016), these methods are restricted in scaling up to real-world deployments due to two fundamental limitations:

(I) *Small Sample Size*: The labelled training population is often small (e.g. hundreds of persons each with a few images) and much smaller (e.g. $<\frac{1}{10}$) than typical feature dimensions. This is because collecting cross-view matched image pairs from different locations is not only tedious but also difficult. The lack of training samples is known as the Small Sample Size (SSS) problem (Chen et al, 2000), which may cause singular intra-class and poor inter-class scatter

matrices. Given that metric learning re-id methods aim to minimise the within-class (intra-person) variance whilst maximising the inter-class (inter-person) variance, the SSS problem is therefore likely to make the solutions suboptimal.

(II) *Inadaptability*: Existing re-id methods often adopt off-line batch-wise model learning with the need for sufficiently large sized training data collected via a time consuming manual labelling process. This *first-labelling-then-training* scheme is not scalable to real-world applications that require deployments at many previously unseen surveillance locations with little or no labelled data in advance. Also, real-world label collection is more incremental, i.e. additional label data are sequentially available for model update over time. It is hence desirable for a re-id model to grow and adapt continuously to progressively available up-to-date labelled data. Existing re-id methods can only afford model re-training from scratch, causing both high computational cost and response latency to a user. They are thus unsuitable for human-in-the-loop model adaptation.

In this work, we solve the two issues by formulating person re-id as a *regression* problem (Hoerl and Kennard, 1970). Unlike existing methods designed to learn *collectively* from all the training identities a *generic* feature embedding space optimised for classification, verification or ranking, we propose to construct an *individually* semantic feature embedding space for identity regression optimised on *each training identity*, referred to as an *Identity Regression Space* (IRS) defined by all training identity classes. Each dimension of IRS corresponds to a specific training person class, i.e. all training images of the same identity class are represented by a single unit vector lying in one unique dimension (axis). Our modelling objective is therefore to train a regression model that maps (embeds) the original image feature space to this identity regression space.

We formulate a re-id incremental learning framework with three fundamental advantages: *First*, it allows quicker re-id system deployment after learning from only a small amount of labelled data. *Second*, the learned re-id model facilitates the subsequent labelling tasks by providing human a ranking order of unlabelled samples with the labelling targets (i.e. true matches) in top ranks at high likelihoods. This reduces manual search time and effort as compared to the conventional exhaustive eye-balling of unstructured person images. *Third*, the re-id model progressively improves from new labelled data to further facilitate future labelling. This interactive effect is cumulative in a loop: More frequently the model updates, more benefit we obtain in both reducing labelling effort and increasing model deployment readiness.

Our **contributions** are three-folds: **(1)** We propose the concept of an *Identity Regression Space* (IRS) by formulating re-id as a regression problem for tackling the inherent Small Sample Size (SSS) challenge. This is in contrast to existing methods relying on classification, verification, or ranking learning spaces which are subject to the SSS problem. The IRS model is featured by an efficient closed-form feature embedding solution without the need for solving an expensive eigen-system and alternative optimisation. **(2)** We introduce an incremental learning algorithm for efficient online IRS model update. This facilitates rapidly updating a IRS re-id model from piecewise new data *only*, for progressively accommodating update-to-date labelled data and viewing condition dynamics, hence avoiding less efficient model re-training from scratch. **(3)** We develop an active learning algorithm for more cost-effective IRS model update with human-in-the-loop, an under-studied aspect in existing re-id methods. Extensive experiments on four popular datasets VIPeR (Gray et al, 2007), CUHK01 (Li et al, 2012), CUHK03 (Li et al, 2014) and Market-1501 (Zheng et al, 2015) show the superiority and advantages of the proposed IRS model over a wide range of state-of-the-art person re-id models.

2 Related Work

Person Re-ID. Existing person re-id studies focus on two main areas: feature representation and matching model. In the literature, a number of hand-crafted image descriptors have been designed for achieving general non-learning based view-invariant re-id features (Farenzena et al, 2010; Zhao et al, 2013; Wang et al, 2014a; Ma et al, 2012; Yang et al, 2014; Matsukawa et al, 2016). However, these representations alone are often insufficient to accurately capture complex appearance variations across cameras. A common solution is supervised learning of a discriminative feature embedding, subject to classification, pairwise or triplet learning constraints (Liao and Li, 2015; Wang et al, 2014b, 2016a).

Our work belongs to the supervised learning based approach but with a few unique advantages. *First*, our IRS is designed with each dimension having discriminative semantics, rather than learning to optimise. We uniquely train a regression mapping from the raw feature space to the interpretable IRS with a close-formed optimisation solution (Hoerl and Kennard, 1970; Hastie et al, 2005) more efficient than solving eigen-problems (Liao et al, 2015; Zhang et al, 2016a) and iterative optimisation (Zheng et al, 2013; Liao and Li, 2015). The IRS addresses the SSS problem in a similar spirit of the NFST re-id model (Chen et al, 2000; Yu and Yang, 2001; Zhang et al, 2016a) by projecting same-identity images into a single point. Importantly, our model uniquely confirms to a well-designed embedding space rather than relying on intra-person scatter matrix which may render the solution less discriminative. *Second*, we further propose an incremental learning algorithm for sequential model update at new scene and/or dynamic deployments without model re-training from scratch. *Finally*, we investigate active sampling for more cost-effective re-id model update.

Subspace Learning. The IRS is a discriminative subspace learning method, similar to distance metric learning (Yang and Jin, 2006), Fisher Discriminant Analysis (FDA) (Fisher, 1936; Fukunaga, 2013), and cross-modal feature matching (Hardoon et al, 2007; Sharma et al, 2012; Kang et al, 2015). Representative metric learning re-id methods include PRDC (Zheng et al, 2013), KISSME (Koestinger et al, 2012), XQDA (Liao et al, 2015), MLAPG (Liao and Li, 2015), LADF (Li et al, 2013), and so forth. PRDC maximises the likelihood of matched pairs with smaller distances than unmatched ones. KISSME measures the probability similarity of intra-class and inter-class feature differences under the Gaussian distribution assumption. sharing the spirit of Bayesian Face model (Moghaddam et al, 2000). KISSME and Bayesian Face are inefficient given high-dimensional features. XQDA overcomes this limitation by uniting dimension reduction and metric learning. MLAPG tackles the efficiency weakness in learning Mahalanobis function. While achieving significant performance gains, these methods focus *only* on one-time batch-wise model learning while ignore incremental learning capability. Our model is designed to fill this gap.

Incremental Learning. Incremental learning (IL) concerns model training from data streams (Poggio and Cauwenberghs, 2001). Often, IL requires extra immediate on-line model update for making the model ready to accept new data at any time. IL has been explored in many different vision tasks, e.g. image classification (Lin et al, 2011; Ristin et al, 2014). The closest works w.r.t. our model are three re-id methods (Liu et al, 2013; Wang et al, 2016c; Martinel et al, 2016).

Specifically, Liu et al (2013) consider to optimise an error-prone post-rank search for refining quickly the ranking lists. this method is inherently restricted and unscalable due to the need for human feedback on all probe images independently. Wang et al (2016c) solves this limitation by learning incrementally a unified generalisable re-id model from all available human feedback. Martinel et al (2016) similarly consider incremental model update in deployment for maintaining re-id performance over time. Compared to these IL re-id methods, the IRS is uniquely characterised with more efficient optimisation (i.e. a closed-form solution) with the capability of low response latency. This is made possible by casting re-id model learning as a regression problem in the concept of well-design identity embedding space, in contrast to classification (Liu et al, 2013), verification (Martinel et al, 2016), or ranking (Prosser et al, 2010; Wang et al, 2016c) learning problem. Given that all these methods adopt their respective human verification designs and incremental learning strategies under distinct evaluation settings, it is impossible to conduct quantitative evaluation among them.

Active Learning. Active learning (AL) is a strategy for reducing human labelling effort by selecting most informative samples for annotation (Settles, 2012; Kang et al, 2004). Despite extensive AL studies on generic object classification (Osugi et al, 2005; Cebron and Berthold, 2009; Hospedales et al, 2012; Ebert et al, 2012; Loy et al, 2012; Käding et al, 2015; Wang et al, 2016e), there exist little re-id attempts with only two works to our knowledge: active person identification (Das et al, 2015) and temporal re-id adaptation (Martinel et al, 2016).

Specifically, Das et al (2015) learn a multi-class classifier on known identity classes for recognising training classes, therefore not a re-id model. Moreover, this model cannot support efficient incremental learning as (Martinel et al, 2016) and IRS, due to expensive re-training from scratch and hence less suitable for AL with human in the loop. Martinel et al (2016) explore also AL for incremental re-id model update. In comparison, our AL algorithm is more extensive and comprehensive (exploitation & exploration vs. exploitation alone) with better learning efficiency (no need for iterative optimisation and graph based data clustering). IRS is thus more suitable for human-in-the-loop driven incremental learning.

Ridge Regression. Ridge regression (Hoerl and Kennard, 1970; Hastie et al, 2005) is one of the most-studied learning algorithms. It has an efficient closed-form solution. with exiting optimised algorithms (Paige and Saunders, 1982) readily applicable to large sized data. We ground the IRS re-id model on ridge regression for inheriting the learning efficiency and scalability advantages, Existing attempts for identity verification problems by class-label regression include (Liao et al, 2014; Sharma et al, 2012; Kang et al, 2015). Liao et al (2014) adopted a linear regression based discriminant analysis method for re-id. Sharma et al (2012) and Kang et al (2015) proposed locality regularised class-label regression methods for recognition and retrieval.

Beyond these existing works, we systematically explore different label coding methods, non-linear regression kernelisation, model efficiency enhancement and labelling effort minimisation in an under-studied incremental re-id learning setting. Moreover, we bridge ridge regression and FDA (Fisher, 1936; Fukunaga, 2013) in feature embedding space design for more discriminatively encoding identity sensitive information. While the relationship between FDA and linear regression has been studied for binary-class (Duda et al, 2012) and multi-class (Hastie et al, 2005; Park and Park, 2005) classification, this is the first study that formulates the two jointly in a single framework for person re-id.

Data Scarcity. There are other generic approaches to solving the SSS challenge. Two common schemes are domain transfer (Layne et al, 2013; Ma et al, 2013; Peng et al, 2016; Geng et al, 2016; Li et al, 2017) and data augmentation (synthesis) (McLaughlin et al, 2015; Zheng et al, 2017). The former relies on auxiliary data (e.g. ImageNet or other re-id datasets) while the latter generates additional training data both for enriching the discriminative information accessible to model training. Conceptually, they are complementary to the proposed IRS with the focus on learning a more discrim-

inative embedding space on the given training data from either scratch or pre-trained models. As shown in our evaluation, these approaches can be jointly deployed for further improving model generalisation (Table 9).

3 Identity Regression

3.1 Problem Definition

We consider the image-based person re-identification problem (Gong et al, 2014). The key is to overcome the unconstrained person appearance variations caused by significant discrepancy in camera viewing condition and human pose (Fig. 1). To this end, we aim to formulate a feature embedding model for effectively and efficiently discovering identity discriminative information of cross-view person images.

Formally, we assume a labelled training dataset $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_i, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{d \times n}$ where $\boldsymbol{x}_i \in \mathbb{R}^{d \times 1}$ denotes the d-dimensional feature vector of image $\boldsymbol{x}_i$, with the corresponding identity label vector $\boldsymbol{l} = [l_1, \cdots, l_i, \cdots, l_n] \in \mathbb{Z}^{1 \times n}$, where $l_i \in \{1, \cdots, c\}$ represents the identity label of image $\boldsymbol{x}_i$ among a total of c identities. So, these n training images describe c different persons captured under multiple camera views. We omit the camera label here for brevity. The model learning objective is to obtain a discriminative feature embedding $\boldsymbol{P} \in \mathbb{R}^{d \times m}$, i.e. in the embedding space, the distance between intra-person images is small whilst that of inter-person images is large regardless of their source camera views. In most existing works, the above criterion of compressing intra-person distributions and expanding inter-person distributions is encoded as classification / verification / ranking losses and then a feature embedding is learned by optimising the corresponding objective formulation. However, due to the SSS problem, the learned embedding space is often suboptimal and less discriminative. Also, there is often no clear interpretation on the learned embedding space.

Our method is significantly different: Prior to the model training, we first explicitly define an *ideal feature embedding space*, and then train a regression from the raw feature space to the defined embedding space. The learned regression function is our discriminative feature embedding. Specifically, we define a set of "*ideal*" target vectors in the embedding space, denoted by $\boldsymbol{Y} = [\boldsymbol{y}_1^\top, \cdots, \boldsymbol{y}_n^\top]^\top \in \mathbb{R}^{n \times m}$, and explicitly assign them to each of the training sample $\boldsymbol{x}_i$, with $\boldsymbol{y}_i \in \mathbb{R}^{1 \times m}$ referring to $\boldsymbol{x}_i$'s target point in the feature embedding space, $i \in \{1, 2, \cdots, n\}$ and m referring to the feature embedding space dimension. In model training, we aim to obtain an optimal feature embedding $\boldsymbol{P}$ that transforms the image feature $\boldsymbol{x}$ into its mapping $\boldsymbol{y}$ with labelled training data $\boldsymbol{X}$. During model deployment, given a test probe image $\tilde{\boldsymbol{x}}^p$ and a set of test gallery images $\{\tilde{\boldsymbol{x}}_i^g\}$, we first transform them into the embedding space with the learned feature embedding $\boldsymbol{P}$, denoted as $\tilde{\boldsymbol{y}}^p$ and $\{\tilde{\boldsymbol{y}}_i^g\}$ respectively. Then, we compute the pairwise matching distances between $\tilde{\boldsymbol{y}}^p$ and $\{\tilde{\boldsymbol{y}}_i^g\}$ by the Euclidean distance metric. Based on matching distances, we rank all gallery images in ascendant order. Ideally, the true match of the probe person is supposed to appear among top ranks.

3.2 Identity Regression Space

To learn an optimal regression function as feature embedding, one key question in our framework is how to design the target "*ideal*" embedding space, in other words, how to set $\boldsymbol{Y}$. We consider two principles in designing distribution patterns of training samples in the embedding space:

1. *Compactness:* This principle concerns image samples belonging to the *same person class*. Even though each person's intra-class distributions may be different in the raw feature space, we argue that in an optimal embedding space for re-id, the variance of all intra-class distributions should be suppressed. Specifically, for every training person, regardless of the corresponding sample size, all samples should be collapsed to a single point so that the embedding space becomes maximally discriminative with respect to person identity.
2. *Separateness:* This principle concerns image samples belonging to the *different person classes*. Intuitively, the points of different person identities should be maximally separated in the embedding space. With a more intuitive geometry explanation, these points should be located on the vertices of a regular simplex with equal-length edges, so that the embedding space treats equally any training person with a well-separated symmetric structure.

Formally, we assign a unit-length vector on each dimension axis in the feature embedding space to every training person identity, i.e. we set $\boldsymbol{y}_i = [y_{i,1}, \cdots, y_{i,m}]$ for the i-th training person (Fig. 2(a)) as:

$$y_{i,j} = \begin{cases} 1, & \text{if} \quad l_i = j; \\ 0, & \text{if} \quad l_i \neq j. \end{cases} \quad \text{with} \quad j \in [1, 2, \cdots, m], \qquad (1)$$

where l_i is the identity label of image $\boldsymbol{x}_i$. We name this way of setting $\boldsymbol{Y}$ as *OneHot Feature Coding*. The embedding space defined by Eq. (1) has a few interesting properties:

1. Each dimension in the embedding space corresponds to one specific training person's identity;
2. Training persons are evenly distributed in the embedding space and the distances between any two training persons are identical;
3. Geometrically, the points of all training person identities together form a standard simplex.

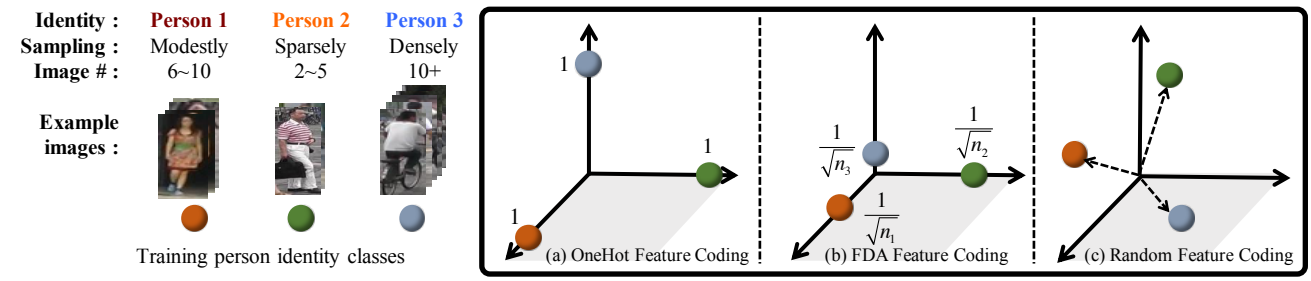


Fig. 2 Illustration of feature embedding spaces obtained by three training class coding methods. Note, n_i in (b) refers to the training image number of person i extracted from any cameras.

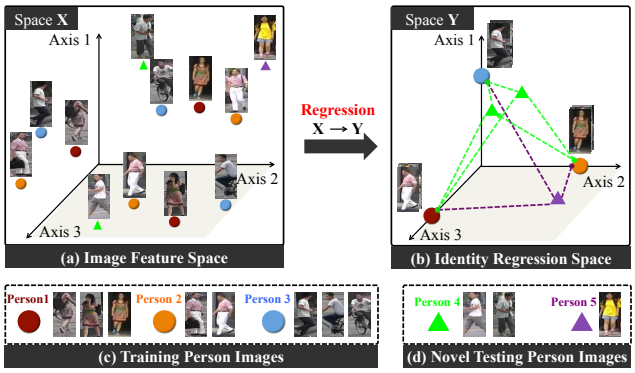


Fig. 3 Illustration of our Identity Regression Space (IRS) person re-id model. During model training, by regression we learn an identity discriminative feature embedding from (a) the image feature space to (b) the proposed identity regression space defined by (c) all training person classes (indicated by circles). During deployment, we can exploit the learned feature embedding to re-identify (d) novel testing person identities (indicated by triangles) in IRS.

Because each dimension of this embedding space can be now interpreted by one specific training identity, we call such an embedding space an *identity regression space*. Having the identity regression space defined by Eq. (1), we propose to exploit the multivariate ridge regression algorithm (Hoerl and Kennard, 1970; Zhang et al, 2010).

In particular, by treating $\boldsymbol{Y}$ as the regression output and $\boldsymbol{P}$ as the to-be-learned parameter, we search for a discriminative projection by minimising the mean squared error as:

$$\boldsymbol{P}^* = \arg\min_{\boldsymbol{P}} \frac{1}{2}\|\boldsymbol{X}^\top \boldsymbol{P} - \boldsymbol{Y}\|_F^2 + \lambda\|\boldsymbol{P}\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm, λ controls the regularisation strength. Critically, this formulation has an efficient closed-form solution:

$$\boldsymbol{P}^* = \left(\boldsymbol{X}\boldsymbol{X}^\top + \lambda \boldsymbol{I}\right)^\dagger \boldsymbol{X}\boldsymbol{Y}, \tag{3}$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose inverse, and $\boldsymbol{I}$ the identity matrix. Since our model learning is by regression towards a training identity space, we call this method the "Identity Regression Space" (**IRS**) model (Fig. 3). The IRS re-id feature learning requirement leads naturally to exploiting the ridge regression method for learning the mapping between image features and this semantic identity space. The novelty of this approach is not in Eq. (2) itself, but the IRS learning concept in the re-id context. Note that, we do not select deep models (Xiao et al, 2016) in our IRS implementation due to their intrinsic weakness for model incremental learning. Nevertheless, in our experiments we also evaluated IRS with a deep learning model (Section 5.1, IV and V). Technically, OneHot based IRS *feature coding* and *embedding* differs fundamentally from deep learning classification models due to two modelling differences: (1) Whilst the latter adopts one-hot *class label vectors*, the underlying optimised deep features (e.g. the feature layer outputs) are not of one-hot style, i.e. not an IRS embedding. (2) A single softmax prediction may correspond to multiple different logit (i.e. feature) inputs. Specifically, even if two logit inputs are different, as long as the corresponding element is *relatively* larger than others, both their softmax outputs will be close to the same one-hot vector. In other words, for deep classification models the underlying feature representations of each class are not unique. Therefore, deep classification model are trained under a weaker learning constraint than the IRS whose feature embedding is trained strictly with only one ground-truth feature vector per class. The regression algorithm selection is independent of the generic IRS concept.

Remark. Unlike Fisher Discriminant Analysis (Fisher, 1936), the proposed IRS has no need for the intra-class and between-class scatter matrices. This renders our model more suitable for addressing the Small Sample Size (SSS) problem since the intra-class scatter matrix of sparse training data will become singular, which results in computational difficulty (Fukunaga, 2013). To solve this SSS problem, one straightforward approach is performing dimensionality reduction (e.g. principal component analysis) before model learning (Pedagadi et al, 2013). This however may cause the loss of discriminative power. An alternative method is directly rectifying the intra-class scatter by adding a non-singular regularisation matrix (Mika et al, 1999; Xiong et al, 2014; Liao et al, 2015). Nonetheless, both approaches as above suffer from the degenerate eigenvalue problem (i.e. several eigenvectors share the same eigenvalue), which makes the solution sub-optimal with degraded discrimination (Zheng et al, 2005). As a more principled solution, the Null Fo-

leySammon Transform (NFST) modifies the Fisher discriminative criterion – Finding null projecting directions on which the intra-class distance is zero whilst the between-class distance is positive – so that more discriminant projections corresponding to the infinitely large Fisher criterion can be obtained (Chen et al, 2000; Guo et al, 2006). The NFST has also been recently employed to solve the SSS problem in re-id (Zhang et al, 2016a). While reaching the largest Fisher objective score via exploiting the null space of intra-class scatter matrix by NFST, the between-class scatter is not maximised and therefore still an incomplete Fisher discriminative analysis. It is easy to see that the proposed IRS model shares the spirit of NFST in terms of projecting same-class images into a single point in order to achieve the extreme class *compactness* and most discriminative feature embedding. However, unlike the NFST's positive between-class scatter constraint – a weaker optimisation constraint likely resulting in lower discriminative power, the model proposed here optimises instead the between-class *separateness* by enforcing the orthogonality between any two different person classes in the target feature space to maximise the class discrimination and separation in a stronger manner. In terms of model optimisation, we resort to the more efficient ridge regression paradigm rather than the Fisher criterion. Overall, we consider that our IRS conceptually extends the NFST by inheriting its local compact classes merit whilst addressing its global class distribution modelling weakness in a more efficient optimisation framework. In our evaluations, we compare our IRS model with the NFST and show the advantages from this new formulation in terms of both model efficiency and discriminative power.

Alternative Feature Coding. Apart from the OneHot feature coding (Eq. (1)), other designs of the embedding space can also be readily incorporated into our IRS model. We consider two alternative feature coding methods. The first approach respects the Fisher Discriminant Analysis (FDA) (Fisher, 1936; Fukunaga, 2013) criterion, named *FDA Feature Coding*, which is adopted in the preliminary version of this work (Wang et al, 2016b). Formally, the FDA criterion can be encoded into our IRS model by setting target identity regression space as (Fig. 2(b)):

$$y_{ij} = \begin{cases} \frac{1}{\sqrt{n_i}}, & \text{if} \quad l_i = j; \\ 0, & \text{if} \quad l_i \neq j. \end{cases} \quad \text{with} \quad j \in [1, 2, \cdots, m]. \quad (4)$$

where n_i and l_i refers to the total image number and identity label of training person i. A detailed derivation is provided in Appendix A. As opposite to Eq. (1) which treats each person identity equally (e.g. assigning them with unit-length vectors in the embedding space), this FDA coding scheme assigns variable-length vectors with the length determined by n_i. As shown in (Fig. 2(b)), with the FDA criterion, the resulting training identity simplex in the embedding space is no longer regular. This may bring benefits for typical classification problems by making size-sensitive use of available training data for modelling individual classes as well as possible, but not necessarily for re-id. Particularly, modelling training classes in such a biased way may instead hurt the overall performance since the re-id model is differently required to generalise the knowledge from training person classes to previously unseen testing ones other than within the training ones as in conventional classification.

The second alternative is *Random Feature Coding*. That is, we allocate for each training identity a m-dimensional random vector with every element following a uniform distribution over the range of $[0, 1]$ (Fig. 2(c)). Random coding has shown encouraging effect in shape retrieval (Zhu et al, 2016) and face recognition (Zhang et al, 2013). In this way, individual dimensions are no longer identity-specific and training identity regression space are shared largely irregularly. We will evaluate the effectiveness of these three feature coding methods in Sec. 5.1.

3.3 Kernelisation

Given complex variations in viewing condition across cameras, the optimal subspace may not be obtainable by linear projections. Therefore, we further kernelise the IRS model (Eq. (3)) by projecting the data from the original visual feature space into a reproducing kernel Hilbert space $\mathcal{H}$ with an implicit feature mapping function $\phi(\cdot)$. The inner-product of two data points in $\mathcal{H}$ can be computed by a kernel function: $h_{\mathrm{k}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$. By h_{k} (we utilised the typical RBF or Gaussian kernel in our implementation), we obtain a kernel representation $\boldsymbol{K} \in \mathbb{R}^{n \times n}$, based on which a corresponding non-linear projection solution can be induced as:

$$\boldsymbol{Q}^* = \left(\boldsymbol{K}\boldsymbol{K}^\top + \lambda \boldsymbol{K}\right)^\dagger \boldsymbol{K}\boldsymbol{Y}. \quad (5)$$

Once test samples are transformed into the kernel space with h_k, we can similarly apply the learned projection $\boldsymbol{Q}^*$ as the linear case. We use the kernel version throughout all experiments due to its capability of modelling the non-linearity which is critical for open space re-id in images with complex person appearance variations across camera views.

4 Incremental Identity Regression

In Sec. 3, we presented the proposed IRS person re-id model. Similar to the majority of conventional re-id methods, we assume a batch-wise model learning setting: First collecting all labelled training data and then learning the feature embedding model (Fig. 4 (a)). In real-world scenario, however, data annotation is likely to arrive in sequence rather

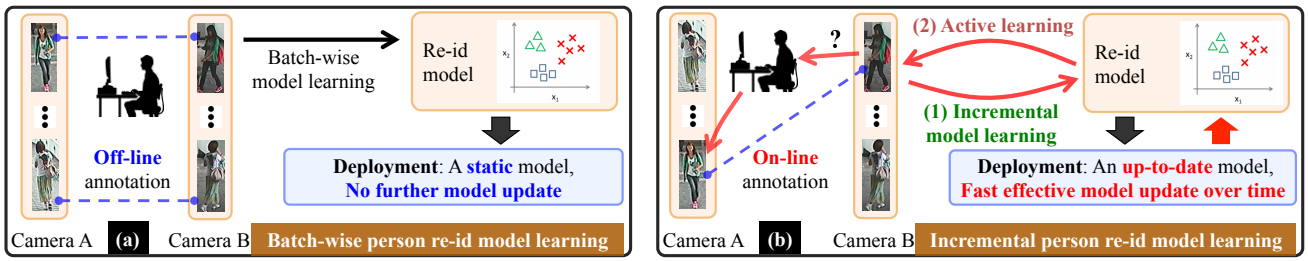


Fig. 4 Illustration of different person re-id model learning settings. **(a)** Batch-wise person re-id model learning: A re-id model is first learned on an exhaustively labelled training set, and then fixed for deployment without model update; **(b)** Incremental person re-id model learning: Training samples are collected sequentially on-the-fly with either random or active unlabelled data selection, and the re-id model keeps up-to-date by efficient incremental learning from the newly labelled data over time.

than at one time particularly when deployed to new arbitrary scenes. In such case, a practical system requires the incremental learning capability for cumulatively learning and updating the re-id model over deployment process (Fig. 4 (b)-(1)). On the other hand, incremental learning is essential for temporal model adaptation, e.g. handling the dynamics in the deployment context (Martinel et al, 2016). A simple and straightforward scheme is to re-train the model from scratch using the entire training dataset whenever any newly labelled samples become available. Obviously, this is neither computational friendly nor scalable particularly for resource/budget restricted deployment.

To overcome this limitation, we introduce an incremental learning algorithm, named IRS$^{\text{inc}}$, for enabling fast model update without the need for re-training from scratch. Suppose at time t, we have the feature matrix $\boldsymbol{X}_t \in \mathbb{R}^{d\times n_t}$ of n_t previously labelled images of c_t person identities, along with $\boldsymbol{Y}_t \in \mathbb{R}^{n_t\times m}$ their indicator matrix defined by Eq. (1). We also have the feature matrix $\boldsymbol{X}' \in \mathbb{R}^{d\times n'}$ of n' newly labelled images of c' new person classes, with $\boldsymbol{Y}' \in \mathbb{R}^{n'\times(c_t+c')}$ the corresponding indicator matrix similarly defined by Eq. (1). After merging the new data, the updated feature and identity embedding matrix can be represented as:

$$\boldsymbol{X}_{t+1} = [\boldsymbol{X}_t,\ \boldsymbol{X}'], \quad \boldsymbol{Y}_{t+1} = \begin{bmatrix} \boldsymbol{Y}_t \oplus \boldsymbol{0} \\ \boldsymbol{Y}' \end{bmatrix}, \tag{6}$$

where $(\cdot)\oplus\boldsymbol{0}$ denotes the matrix augmentation operation, i.e. padding an appropriate number of zero columns on the right. By defining

$$\boldsymbol{T}_t = \boldsymbol{X}_t\boldsymbol{X}_t^\top, \tag{7}$$

and applying Eq. (6), we have

$$\boldsymbol{T}_{t+1} = \boldsymbol{T}_t + \boldsymbol{X}'\boldsymbol{X}'^\top. \tag{8}$$

For initialisation, i.e. when $t=0$, we set $\boldsymbol{T}_0 = \boldsymbol{X}_0\boldsymbol{X}_0^\top + \lambda\boldsymbol{I}$. Also, we can express the projection $\boldsymbol{P}_t \in \mathbb{R}^{d\times m}$ (Eq. (3)) of our IRS model at time t as

$$\boldsymbol{P}_t = \boldsymbol{T}_t^\dagger \boldsymbol{X}_t\boldsymbol{Y}_t. \tag{9}$$

Our aim is to obtain the feature embedding $\boldsymbol{P}_{t+1}$, which requires to compute $\boldsymbol{T}_{t+1}^\dagger$. This can be achieved by applying the Sherman-Morrison-Woodbury formula (Woodbury, 1950) to Eq. (8) as:

$$\boldsymbol{T}_{t+1}^\dagger = \boldsymbol{T}_t^\dagger - \boldsymbol{T}_t^\dagger\boldsymbol{X}'\big(\boldsymbol{I}+\boldsymbol{X}'^\top\boldsymbol{T}_t^\dagger\boldsymbol{X}'\big)^\dagger\boldsymbol{X}'^\top\boldsymbol{T}_t^\dagger. \tag{10}$$

Eq. (3) and Eq. (6) together give us:

$$\begin{aligned}\boldsymbol{P}_{t+1} &= \boldsymbol{T}_{t+1}^\dagger\boldsymbol{X}_{t+1}\boldsymbol{Y}_{t+1} \\ &= (\boldsymbol{T}_{t+1}^\dagger\boldsymbol{X}_t\boldsymbol{Y}_t)\oplus\boldsymbol{0} + \boldsymbol{T}_{t+1}^\dagger\boldsymbol{X}'\boldsymbol{Y}'.\end{aligned} \tag{11}$$

Further with Eq. (10) and Eq. (9), we can update $\boldsymbol{P}$ as:

$$\begin{aligned}\boldsymbol{P}_{t+1} = \Big(\boldsymbol{P}_t - \boldsymbol{T}_t^\dagger\boldsymbol{X}'\big(\boldsymbol{I}+\boldsymbol{X}'^\top\boldsymbol{T}_t^\dagger\boldsymbol{X}'\big)^\dagger\boldsymbol{X}'^\top\boldsymbol{P}_t\Big)\oplus\boldsymbol{0} \\ +\boldsymbol{T}_{t+1}^\dagger\boldsymbol{X}'\boldsymbol{Y}'.\end{aligned} \tag{12}$$

Note, the model update (Eq. (10) and Eq. (12)) only involves newly coming data samples. Hence, our method does not require to store the training data once used for model update. As only cheap computational cost is involved in such linear operations, the proposed algorithm well suits for on-line responsive re-id model learning and updating in deployment at large scales in reality.

Implementation Consideration. The IRS$^{\text{inc}}$ model supports incremental learning given either a single new sample ($n'=1$) or a small chunk of new samples ($n'\geqslant 2$). If the data chunk size $n' \ll d$ (where d is the feature dimension), it is faster to perform n' separate updates on each new sample instead of by a whole chunk. The reason is that, in such a way the Moore-Penrose matrix inverse in Eq. (10) and Eq. (12) can be reduced to n' separate scaler inverse operations, which is much cheaper in numerical computation.

4.1 Active Learning for Cost-Effective Incremental Update

The incremental learning process described above is *passive*, i.e. a human annotator is supposed to label randomly chosen data without considering the potential value of each

selected sample in improving the re-id model. Therefore, data annotation by this random way is likely to contain redundant information with partial labelling effort wasted. To resolve this problem, we explore the active learning idea (Settles, 2012) for obtaining more cost-effective incremental re-id model update (Fig. 4 (b)-(2)).

Active IRS$^{\text{inc}}$ Overview. In practice, we often have access to a large number of *unlabelled* images $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{G}}$ captured by disjoint cameras. Assume at time step $t \in \{1, \cdots, \tau\}$ with τ defining the pre-determined human labelling budget, we have the up-to-date IRS$^{\text{inc}}$ model m_t (corresponding to the feature embedding $\boldsymbol{P}_t$), along with $\widetilde{\mathcal{P}}_t$ and $\widetilde{\mathcal{G}}_t$ denoting the remaining unlabelled data. To maximise labelling profit, we propose an *active labelling* algorithm for IRS$^{\text{inc}}$ with the main steps as follows:

1. An image $\boldsymbol{x}_t^p \in \widetilde{\mathcal{P}}_t$ of a new training identity l_t is *actively* selected by model m_t, according to its potential usefulness and importance measured by certain active sampling criteria (see details below);
2. A ranking list of unlabelled images $\widetilde{\mathcal{G}}_t$ against the selected $\boldsymbol{x}_t^p$ is then generated by m_t based matching distances;
3. For the selected $\boldsymbol{x}_t^p$, a human annotator is then asked to manually identify the cross-view true matching image $\boldsymbol{x}_t^g \in \widetilde{\mathcal{G}}_t$ in the ranking list, and then generate a new annotation $(\boldsymbol{x}_t^p, \boldsymbol{x}_t^g)$;
4. The IRS$^{\text{inc}}$ re-id model is updated to m_{t+1} (i.e. $\boldsymbol{P}_{t+1}$) from the new data annotation $(\boldsymbol{x}_t^p, \boldsymbol{x}_t^g)$ by our incremental learning algorithm (Eq. (10) and Eq. (12)).

Among these steps above, the key lies in how to select a good image $\boldsymbol{x}_t^p$. To this end, we derive a "Joint Exploration-Exploitation" (**JointE**^{2}) active sampling algorithm composed of three criteria as follows (Figure 5).

(I) Appearance Diversity Exploration. Intuitively, the appearance diversity of training people is a critical factor for the generalisation capability of a re-id model. Thus, the preferred next image to annotate should lie in the most unexplored region of the population $\widetilde{\mathcal{P}}_t$. Specifically, at time t, the distance between any two samples $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ by the current re-id model is computed as:

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2|m_t) = (\boldsymbol{x}_1 - \boldsymbol{x}_2)^\top \boldsymbol{P}_t \boldsymbol{P}_t^\top (\boldsymbol{x}_1 - \boldsymbol{x}_2). \tag{13}$$

Given the unlabelled $\widetilde{\mathcal{P}}_t$ and labelled $\mathcal{P}_t$ part of the set $\widetilde{\mathcal{P}}$ ($\widetilde{\mathcal{P}}_t \bigcup \mathcal{P}_t = \widetilde{\mathcal{P}}$), we can measure the diversity degree of an unlabelled sample $\boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t$ by its distance against the *within-view nearest neighbour* in $\mathcal{P}_t$ (Figure 5 (a)):

$$\begin{aligned} \varepsilon_1(\boldsymbol{x}_i^p) = \min\ & d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^p|m_t), \\ \text{s.t. }\ & \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t,\ \ \boldsymbol{x}_j^p \in \mathcal{P}_t. \end{aligned} \tag{14}$$

Eq. (14) defines the distance of an *unlabelled* sample $\boldsymbol{x}_i^p$ from the labelled set, i.e. the distance between $\boldsymbol{x}_i^p$ and its nearest labelled sample. This is not an optimisation operation. It is a nearest sample search by "min" operation. By maximising the nearest distances, more diverse person appearance can be covered and learned for more rapidly increasing the knowledge of the IRS$^{\text{inc}}$ model, avoiding repeatedly learning visually similar training samples.

(II) Matching Discrepancy Exploration. A well learned re-id model is supposed to find the true match of a given image with a small cross-view matching distance. In this perspective, our second criterion particularly prefers the samples with large matching distances in the embedding space, i.e. the re-id model m_t remains largely unclear on what are the likely corresponding cross-view appearances of these "unfamiliar" people. Numerically, we compute the matching distance between an unlabelled sample $\boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t$ and the cross-view true match (assumed as *cross-view nearest neighbour*) in $\widetilde{\mathcal{G}}$ (Figure 5 (b)):

$$\begin{aligned} \varepsilon_2(\boldsymbol{x}_i^p) = \min\ & d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^g|m_t), \\ \text{s.t. }\ & \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t,\ \ \boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}. \end{aligned} \tag{15}$$

That is, the unlabelled images with greater $\varepsilon_2(\boldsymbol{x}_i^p)$ are preferred to be selected.

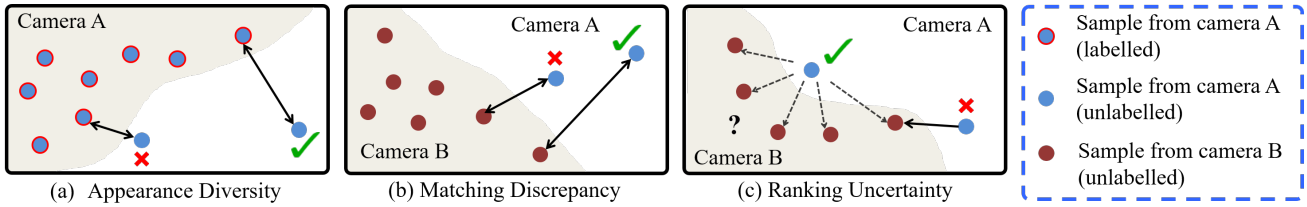
(III) Ranking Uncertainty Exploitation. Uncertainty-based exploitative sampling schemes have been widely investigated for classification problems (Joshi et al, 2009; Settles and Craven, 2008; Ebert et al, 2012). The essential idea is to query the least certain sample for human to annotate. Tailored for re-id tasks with this idea, given the similar appearance among different identities, a weak re-id model may probably generate similar ranking scores for those visually ambiguous gallery identities with respect to a given probe. Naturally, it should be useful and informative to manually label such "challenging" samples for enhancing a person re-id model's discrimination power particularly with regard to such person appearance (Figure 5 (c)). To obtain such person images, we define a matching distance based probability distribution over all samples $\boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}$ for a given cross-view image $\boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}$:

$$p_{m_t}(\boldsymbol{x}_j^g|\boldsymbol{x}_i^p) = \frac{1}{Z_i^t} e^{-d(\boldsymbol{x}_i^p, \boldsymbol{x}_j^g|m_t)}, \tag{16}$$

where

$$Z_i^t = \sum_k e^{-d(\boldsymbol{x}_i^p, \boldsymbol{x}_k^g|m_t)},\ \ \boldsymbol{x}_k^g \in \widetilde{\mathcal{G}}.$$

The quantity $p_{m_t}(\boldsymbol{x}_j^g|\boldsymbol{x}_i^p)$ gives a high entropy when most ranking scores are adjacent to each other, indicating great information to mine from the perspective of information theory (Akaike, 1998). In other words, the model has only a low confidence on its generated ranking list considering that only a very few number of cross-camera samples are likely



(a) Appearance Diversity (b) Matching Discrepancy (c) Ranking Uncertainty

Fig. 5 Illustration of the proposed active exploration and exploitation selection criteria for more cost-effective incremental re-id model learning.

Algorithm 1: Active IRS$^{\text{inc}}$

Data:
(1) Unlabelled image set $\widetilde{\mathcal{P}}$ and $\widetilde{\mathcal{G}}$ from disjoint cameras;
(2) Regularisation strength λ;
(3) Labelling budget τ.

Result:
(1) Discriminative feature embedding matrix $\boldsymbol{P}$;

Initialisation:
(1) Randomly label a small seed set $\boldsymbol{X}_0$, $\boldsymbol{Y}_0$;
(2) Set $\boldsymbol{T}_0^{\dagger} = (\boldsymbol{X}_0\boldsymbol{X}_0^{\top} + \lambda\boldsymbol{I})^{\dagger}$;
(3) Set $\boldsymbol{P}_0 = \boldsymbol{T}_0^{\dagger}\boldsymbol{X}_0\boldsymbol{Y}_0$ (Eq. (3)).

Active Labelling:
for $t = 0 : \tau - 1$ **do**
(1) Select an unlabelled sample $\boldsymbol{x}_t^p \in \widetilde{\mathcal{P}}_t$ (Eq. (18));
(2) Rank the images in $\widetilde{\mathcal{G}}_t$ against the selection $\boldsymbol{x}_t^p$;
(3) Human annotator verifies the true match in $\widetilde{\mathcal{G}}_t$;
(4) Generate a new annotation $(\mathcal{I}_t^p, \mathcal{I}_t^g)$;
(5) Update $\boldsymbol{T}_{t+1}^{\dagger}$ (Eq. (10));
(6) Update $\boldsymbol{P}_{t+1}$ (Eq. (12)).
end
return $\boldsymbol{P} = \boldsymbol{P}_{\tau}$;

to be true matches rather than many of them. Consequently, our third criterion is designed as:

$$\varepsilon_3(\boldsymbol{x}_i^p) = -\sum_j p_{m_t}(\boldsymbol{x}_j^g|\boldsymbol{x}_i^p)\log p_{m_t}(\boldsymbol{x}_j^g|\boldsymbol{x}_i^p), \quad \text{s.t. } \boldsymbol{x}_i^p \in \widetilde{\mathcal{P}}_t, \ \ \boldsymbol{x}_j^g \in \widetilde{\mathcal{G}}. \tag{17}$$

which aims to select out those associated with high model ranking ambiguity.

Joint Exploration-Exploitation. Similar to the model in (Cebron and Berthold, 2009; Ebert et al, 2012), we combine both exploitation and exploration based criteria into our final active selection standard, formally as:

$$\varepsilon(\boldsymbol{x}_i^p) = \varepsilon_1(\boldsymbol{x}_i^p) + \varepsilon_2(\boldsymbol{x}_i^p) + \varepsilon_3(\boldsymbol{x}_i^p). \tag{18}$$

To eliminate scale discrepancy, we normalise ε_1, ε_2, ε_3 to the unit range $[0, 1]$ respectively before fusing them. Specifically, given ε_1 scores of all unlabelled samples, we normalise them by dividing the maximal value so that the highest ε_1 is 1. The same operation is performed on ε_2 and ε_3.

In summary, with Eq. (18), all the unlabelled samples in $\widetilde{\mathcal{P}}$ can be sorted accordingly, and the one with highest $\varepsilon(\boldsymbol{x}_i^p)$ is then selected for human annotation. An overview of our proposed active learning based incremental model learning and updating is presented in Algorithm 1. We will show the effect of our proposed active labelling method in our evaluations (Sec. 5.2).

Table 1 Statistics of person re-id datasets. BBox: Bounding Box.

Dataset	Cameras	Persons	Labelled BBox	Detected BBox
VIPeR	2	632	1,264	0
CUHK01	2	971	1,942	0
CUHK03	6	1,467	14,097	14,097
Market-1501	6	1,501	0	32,668

4.2 Kernelisation

We kernelise similarly the incremental IRS algorithm as in Sec. 3.3. Specifically, we first obtain the kernel representation of new training data and then conduct model incremental learning in the Hilbert space. We utilise the kernelised model with its non-linear modelling power in all incremental re-id model learning experiments including active sampling with human-in-the-loop.

5 Experiments

Datasets. For model evaluation, four person re-id benchmarks were used: VIPeR (Gray et al, 2007), CUHK01 (Li et al, 2012), CUHK03 (Li et al, 2014), and Market-1501 (Zheng et al, 2015), as summarised in Table 1. We show in Fig. 6 some examples of person images from these datasets. Note that the datasets were collected with different data sampling protocols: (a) VIPeR has one image per person per view; (b) CUHK01 contains two images per person per view; (c) CUHK03 consists of a maximum of five images per person per view, and also provides both manually labelled and auto-detected image bounding boxes with the latter posing more challenging re-id test due to unknown misalignment of the detected bounding boxes; (d) Market-1501 has variable numbers of images per person per view. These four datasets present a good selection of re-id test scenarios with different population sizes under realistic viewing conditions exposed to large variations in human pose and strong similarities among different people.

Features. To capture the detailed information of person appearance, we adopted three state-of-the-art feature representations with variable dimensionalities from 10^4 to 10^2:

(a) VIPeR(Gray et al, 2007) (b) CUHK01(Li et al, 2012) (c) CUHK03(Li et al, 2014) (d) Market-1501(Zheng et al, 2015)

Fig. 6 Example person images from four person re-id datasets. Two images of each individual columns present the same person.

(1) *Local Maximal Occurrence* (LOMO) feature (Liao et al, 2015): The LOMO feature is based on a HSV colour histogram and Scale Invariant Local Ternary Pattern (Liao et al, 2010). For alleviating the negative effects caused by camera view discrepancy, the Retinex algorithm (Land and McCann, 1971) is applied to pre-process person images. The feature dimension of LOMO is rather high at $26,960$, therefore expensive to compute.
(2) *Weighted Histograms of Overlapping Stripes* (WHOS) feature (Lisanti et al, 2014, 2015): The WHOS feature contains HS/RGB histograms and HOG (Wang et al, 2009) of image grids, with a centre support kernel as weighting to approximately segmented person foreground from background clutters. We implemented this feature model as described by Lisanti et al (2014). The feature dimension of WHOS is moderate at $5,138$.
(3) *Convolutional Neural Network* (CNN) feature (Xiao et al, 2016): Unlike hand-crafted LOMO and WHOS features, deep CNN person features are learned from image data. Specifically, we adopted the DGD CNN (Xiao et al, 2016) and used the FC_7 output as re-id features. The DGD feature has a rather low dimension of 256, thus efficient to extract. Following Xiao et al (2016), we trained the DGD by combining labelled and detected person bounding box images (a total $26,246$ images) with the original authors released codes. We then deployed the trained DGD to extract deep features of the test image data for CUHK03 (the same domain). On Market-1501, the CUHK03 trained DGD was further fine-tuned on the $12,936$ Market-1501 training images for domain adaptation. On VIPeR and CUHK01, the CUHK03 trained DGD was *directly* deployed *without* any fine-tuning as there are insufficient training images to make effective model adaptation, with only 632 and $1,940$ training images for VIPeR and CUHK01 respectively.

Model Training Settings. In evaluations, we considered extensively comparative experiments under two person re-id model training settings: **(I)** *Batch-wise model training*: In this setting, we followed the conventional supervised re-id scheme commonly utilised in most existing methods, that is, first collecting all training data and then learning a re-id model *before* deployment. **(II)** *Incremental model training*: In contrast to the batch-wise learning, we further evaluated a more realistic data labelling scenario where more training labels are further collected over time *after* model deployment. The proposed IRS^{inc} model was deployed for this incremental learning setting.

5.1 Batch-Wise Person Re-Id Evaluation

Batch-Wise Re-Id Evaluation Protocol. To facilitate quantitative comparisons with existing re-id methods, we adopted the standard supervised re-id setting to evaluate the proposed IRS model. Specifically, on *VIPeR*, we split randomly the whole population of the dataset (632 people) into two halves: One for training (316) and another for testing (316). We repeated 10 trials of random people splits and utilised the averaged results. On *CUHK01*, we considered two benchmarking training/test people split settings: (1) 485/486 split: randomly selecting 485 identities for training and the other 486 for testing (Liao et al, 2015; Zhang et al, 2016a); (2) 871/100 split: randomly selecting 871 identities for training and the other 100 for testing (Ahmed et al, 2015; Shi et al, 2016). As CUHK01 is a multi-shot (e.g. multiple images per person per camera view) dataset, we computed the final matching distance between two people by averaging corresponding cross-view image pairs. Again, we reported the results averaged over 10 random trials for either people split. On *CUHK03*, following Li et al (2014) we repeated 20 times of random 1260/100 people splits for model training/test and reported the averaged accuracies under the single-shot evaluation setting(Zhang et al, 2016a). On *Market-1501*, we used the standard training/test (750/751) people split provided by Zheng et al (2015). On all datasets, we exploited the cumulative matching characteristic (CMC) to measure the re-id accuracy performance. On Market-1501, we also considered the recall measure of multiple truth matches by mean Average Precision (mAP), i.e. first computing the area

Table 2 Re-Id performance comparison on the VIPeR benchmark. (*): Multiple features fusion.

Dataset	VIPeR			
Rank (%)	R1	R5	R10	R20
LADF (Li et al, 2013)	29.3	61.0	76.0	88.1
MFA (Yan et al, 2007)	32.2	66.0	79.7	90.6
kLFDA (Xiong et al, 2014)	38.6	69.2	80.4	89.2
XQDA (Liao et al, 2015)	40.0	68.1	80.5	91.1
MLAPG (Liao and Li, 2015)	40.7	69.9	82.3	92.4
NFST (Zhang et al, 2016a)	42.3	71.5	82.9	92.1
LSSCDL (Zhang et al, 2016b)	42.7	-	84.3	91.9
TMA (Martinel et al, 2016)	43.8	-	83.8	91.5
HER (Wang et al, 2016b)	45.1	74.6	85.1	93.3
DML (Yi et al, 2014)	28.2	59.3	73.5	86.4
DCNN+ (Ahmed et al, 2015)	34.8	63.6	75.6	84.5
SICI (Wang et al, 2016a)	35.8	-	-	-
DGD (Xiao et al, 2016)	38.6	-	-	
Gated S-CNN (Varior et al, 2016a)	37.8	66.9	77.4	-
MCP (Cheng et al, 2016)	**47.8**	74.7	84.8	91.1
IRS (WHOS)	44.5	**75.0**	**86.3**	**93.6**
IRS (LOMO)	45.1	74.6	85.1	93.3
IRS (CNN)	33.1	59.9	71.5	82.2
MLF* (Zhao et al, 2014)	43.4	73.0	84.9	93.7
ME* (Paisitkriangkrai et al, 2015)	45.9	77.5	88.9	95.8
CVDCA* (Chen et al, 2016c)	47.8	76.3	86.3	94.0
FFN-Net* (Wu et al, 2016)	51.1	81.0	91.4	**96.9**
NFST* (Zhang et al, 2016a)	51.2	82.1	90.5	95.9
HER* (Wang et al, 2016b)	53.0	79.8	89.6	95.5
GOG* (Matsukawa et al, 2016)	49.7	-	88.7	94.5
SCSP* (Chen et al, 2016a)	53.5	**82.6**	**91.5**	96.7
IRS (WHOS+LOMO+CNN)*	**54.6**	81.5	90.3	95.7

under the Precision-Recall curve for each probe, then calculating the mean of Average Precision over all probes (Zheng et al, 2015).

In the followings, we evaluated: (i) Comparisons to state-of-the-arts, (ii) Effects of embedding space design, (iii) Effects of features, (iv) Deep learning regression, (v) Complementary of transfer learning and IRS, (vi) Comparisons to subspace/metric learning models, (vii) Regularisation sensitivity, and (viii) Model complexity.

(I) Comparisons to the State-of-The-Arts. We first evaluated the proposed IRS model by extensive comparisons to the existing state-of-the-art re-id models under the standard supervised person re-id setting. We considered a wide range of existing re-id methods, including both hand-crafted and deep learning models. In the following experiments, we deployed the *OneHot Feature Coding* (Eq. (1) in Sec. 3.2) for the identity regression space embedding of our IRS model unless stated otherwise. We considered both single- and multi-feature based person re-id performance, and also compared re-id performances of different models on auto-detected person boxes when available in CUHK03 and Market-1501.

Evaluation on VIPeR. Table 2 shows a comprehensive comparison on re-id performance between our IRS model (and its variations) and existing models using the VIPeR benchmark (Gray et al, 2007). It is evident that our IRS model with a non-deep feature LOMO, IRS(LOMO), is better than all existing methods[1] except the deep model MCP (Cheng et al, 2016), with Rank-1 45.1% vs. 47.8% respectively. Interestingly, using our CUHK03 trained CNN deep feature *without* fine-tuning on VIPeR, i.e. IRS(CNN), does not offer extra advantage (Rank-1 33.1%), due to the significant domain drift between VIPeR and CUHK03. This becomes more clear when compared with the CUHK01 tests below. Moreover, given a score-level fusion on the matching of three different features, IRS(WHOS+LOMO+CNN), the IRS can benefit from further boosting on its re-id performance, obtaining the best Rank-1 rate at 54.6%. These results demonstrate the effectiveness of the proposed IRS model in learning identity discriminative feature embedding because of our *unique* approach on identity regression to learning a re-id feature embedding space, in contrast to existing established ideas on classification, verification or ranking based supervised learning of a re-id model.

Table 3 Re-id performance comparison on the CUHK01 benchmark. (*): Multiple features fusion.

Dataset	CUHK01 (486/485 split)			
Rank (%)	R1	R5	R10	R20
kLFDA (Xiong et al, 2014)	54.6	80.5	86.9	92.0
XQDA (Liao et al, 2015)	63.2	83.9	90.0	94.2
MLAPG (Liao and Li, 2015)	64.2	85.4	90.8	94.9
NFST (Zhang et al, 2016a)	65.0	85.0	89.9	94.4
HER (Wang et al, 2016b)	68.3	86.7	92.6	96.2
DCNN+ (Ahmed et al, 2015)	47.5	71.6	80.3	87.5
MCP (Cheng et al, 2016)	53.7	84.3	91.0	93.3
DGD (Xiao et al, 2016)	66.6	-	-	-
IRS (WHOS)	48.8	73.4	81.1	88.3
IRS (LOMO)	68.3	86.7	92.6	96.2
IRS (CNN)	**68.6**	**89.3**	**93.9**	**97.2**
ME* (Paisitkriangkrai et al, 2015)	53.4	76.4	84.4	90.5
FFN-Net* (Wu et al, 2016)	55.5	78.4	83.7	92.6
GOG* (Matsukawa et al, 2016)	67.3	86.9	91.8	95.9
NFST* (Zhang et al, 2016a)	69.1	86.9	91.8	95.4
HER* (Wang et al, 2016b)	71.2	90.0	94.4	97.3
IRS (WHOS+LOMO+CNN)*	**80.8**	**94.6**	**96.9**	**98.7**
Dataset	CUHK01 (871/100 split)			
FPNN (Li et al, 2014)	27.9	59.6	73.5	87.3
DCNN+ (Ahmed et al, 2015)	65.0	-	-	-
JRL (Chen et al, 2016b)	70.9	92.3	96.9	98.7
EDM (Shi et al, 2016)	69.4	-	-	-
SICI (Wang et al, 2016a)	71.8	-	-	-
IRS (WHOS)	77.0	92.8	96.5	99.2
IRS (LOMO)	80.3	94.2	96.9	99.5
IRS (CNN)	84.4	98.2	**99.8**	**100**
IRS (WHOS+LOMO+CNN)*	**88.4**	**98.8**	99.6	**100**

Evaluation on CUHK01. Table 3 shows a comprehensive comparison of the IRS model with existing competitive re-id models on the CUHK01 benchmark (Li et al, 2012). It is clear that the proposed IRS model achieves the best re-id accuracy under both training/test split protocols. Note that, HER (Wang et al, 2016b) is IRS-FDA(LOMO). Specifically, for the 486/485 split, our IRS(CNN) method surpassed the deep learning DGD model (Xiao et al, 2016), the second best

[1] The HER model presented in our preliminary work (Wang et al, 2016b) is the same as IRS(LOMO) with FDA coding (Eq. (4)), i.e. HER = IRS-FDA(LOMO). On the other hand, IRS(LOMO) in Tables 2, 3, 4 and 5 is IRS-OneHot(LOMO). The effects of choosing different coding is evaluated later (Table 6).

Table 4 Re-id performance comparison on the CUHK03 benchmark. (*): Multiple features fusion.

Dataset	CUHK03 (Manually)			
Rank (%)	R1	R5	R10	R20
kLFDA (Xiong et al, 2014)	45.8	77.1	86.8	93.1
XQDA (Liao et al, 2015)	52.2	82.2	92.1	96.3
MLAPG (Liao and Li, 2015)	58.0	87.1	94.7	98.0
NFST (Zhang et al, 2016a)	58.9	85.6	92.5	96.3
HER (Wang et al, 2016b)	60.8	87.0	95.2	97.7
DCNN+ (Ahmed et al, 2015)	54.7	86.5	93.9	**98.1**
EDM (Shi et al, 2016)	61.3	-	-	-
DGD (Xiao et al, 2016)	75.3	-	-	
IRS (WHOS)	59.6	87.2	92.8	96.9
IRS (LOMO)	61.6	87.0	94.6	98.0
IRS (CNN)	**81.5**	**95.7**	**97.1**	98.0
ME* (Paisitkriangkrai et al, 2015)	62.1	89.1	94.3	97.8
NFST* (Zhang et al, 2016a)	62.6	90.1	94.8	98.1
HER* (Wang et al, 2016b)	65.2	92.2	96.8	**99.1**
GOG* (Matsukawa et al, 2016)	67.3	91.0	96.0	-
IRS (WHOS+LOMO+CNN)*	**81.9**	**96.5**	**98.2**	98.9
Dataset	CUHK03 (Detected)			
KISSME (Koestinger et al, 2012)	11.7	33.3	48.0	-
XQDA (Liao et al, 2015)	46.3	78.9	83.5	93.2
MLAPG (Liao and Li, 2015)	51.2	83.6	92.1	96.9
L_1-Lap (Kodirov et al, 2016)	30.4	-	-	-
NFST (Zhang et al, 2016a)	53.7	83.1	93.0	94.8
DCNN+ (Ahmed et al, 2015)	44.9	76.0	83.5	93.2
EDM (Shi et al, 2016)	52.0	-	-	-
SICI (Wang et al, 2016a)	52.1	84.9	92.4	-
S-LSTM (Varior et al, 2016b)	57.3	80.1	88.3	-
Gated S-CNN (Varior et al, 2016a)	68.1	88.1	94.6	-
IRS (WHOS)	50.6	82.1	90.4	96.1
IRS (LOMO)	53.4	83.1	91.2	96.4
IRS (CNN)	**80.3**	**96.3**	**98.6**	**99.0**
NFST* (Zhang et al, 2016a)	54.7	84.8	94.8	95.2
GOG* (Matsukawa et al, 2016)	65.5	88.4	93.7	-
IRS (WHOS+LOMO+CNN)*	**83.3**	**96.2**	**97.9**	**98.6**

Table 5 Re-id performance comparison on the Market-1501 benchmark. (*): Multiple features fusion.

Dataset	Market-1501			
Query Per Person	Single-Query		Multi-Query	
Metric (%)	R1	mAP	R1	mAP
KISSME (Koestinger et al, 2012)	40.5	19.0	-	-
MFA (Yan et al, 2007)	45.7	18.2	-	-
kLFDA (Xiong et al, 2014)	51.4	24.4	52.7	27.4
XQDA (Liao et al, 2015)	43.8	22.2	54.1	28.4
SCSP (Chen et al, 2016a)	51.9	26.3	-	-
NFST (Zhang et al, 2016a)	55.4	29.9	68.0	41.9
TMA (Martinel et al, 2016)	47.9	22.3	-	-
SSDAL (Su et al, 2016)	39.4	19.6	49.0	25.8
S-LSTM (Varior et al, 2016b)	-	-	61.6	35.3
Gated S-CNN (Varior et al, 2016a)	65.8	39.5	76.0	48.4
IRS (WHOS)	55.2	27.5	60.3	33.5
IRS (LOMO)	57.7	29.0	68.0	37.8
IRS (CNN)	**72.7**	**48.1**	**80.2**	**58.5**
SCSP* (Chen et al, 2016a)	51.9	26.4	-	-
NFST* (Zhang et al, 2016a)	61.0	35.7	71.6	46.0
IRS (WHOS+LOMO+CNN)*	**73.9**	**49.4**	**81.4**	**59.9**

in this comparison, by Rank-1 $2.0\%(68.6-66.6)$. For the 871/100 split, IRS(CNN) yields a greater performance boost over DGD with improvement on Rank-1 at $12.6\%(84.4-71.8)$. It is also worth pointing out that the DGD model was trained using data from other 6 more datasets and further carefully fine-tuned on CUHK01. In contrast, our IRS(CNN) model was only trained on CUHK03 without fine-tuning on CUHK01, and the CNN architecture we adopted closely resembles to that of DGD. By fusing multiple features, the performance margin of IRS(WHOS+LOMO+CNN) over the existing models is further enlarged under both splits, achieving Rank-1 $11.7\%(80.8-69.1)$ boost over NFST (Zhang et al, 2016a) and Rank-1 $16.6\%(88.4-71.8)$ boost over SICI (Wang et al, 2016a), respectively. Compared to VIPeR, the overall re-id performance advantage of the IRS model on CUHK01 is greater over existing models. This is due to not only identity prototype regression based feature embedding, but also less domain drift from CUHK03 to CUHK01, given that the CNN feature used by IRS was trained on CUHK03.

Evaluation on CUHK03. The person re-id performance of different methods as compared to the IRS model on CUHK03 (Li et al, 2014) is reported in Table 4. We tested on both the manually labelled and automatically detected bounding boxes. Similar to VIPeR and CUHK01, our IRS model surpassed clearly all compared methods in either single- or multi-feature setting given manually labelled bounding boxes. Importantly, this advantage remains when more challenging detected bounding boxes were used, whilst other strong models such as NFST and GOG suffered more significant performance degradation. This shows both the robustness of our IRS model against misalignment and its greater scalability to real-world deployments.

Evaluation on Market-1501. We evaluated the re-id performance of existing models against the proposed IRS model on the Market-1501 benchmark (Zheng et al, 2015). The bounding boxes of all person images of this dataset were generated by an automatic pedestrian detector. Hence, this dataset presents a more realistic challenge to re-id models than conventional re-id datasets with manually labelled bounding boxes. Table 5 shows the clear superiority of our IRS model over all competitors. In particular, our IRS model achieved Rank-1 73.9% for single-query and Rank-1 81.4% for multi-query, significantly better than the strongest alternative method, the deep Gated S-CNN model (Varior et al, 2016a), by $8.1\%(73.9–65.8)$ (single-query) and $5.4\%(81.4–76.0)$ (multi-query). Similar advantages hold when compared using the mAP metric.

In summary, these comparative evaluations on the performance of batch-wise re-id model learning show that the IRS model outperforms comprehensively a wide range of existing re-id methods including both hand-crafted and deep learning based models. This validates the effectiveness and advantages of learning a re-id discriminative feature embedding using the proposed approach on identity regression.

(II) Effects of Embedding Space Design. To give more insight on why and how the IRS model works, we evaluated the effects of embedding space design in our IRS model. To this end, we compared the three coding methods as described in Sec. 3.2: *OneHot Feature Coding* in the proposed *Identity Regression Space*, *FDA Feature Coding* by Wang

Table 6 Effects of embedding space on re-id performance in our proposed IRS model. The LOMO visual feature were used on all datasets. We adopted the 485/486 people split on CUHK01 and the manually labelled person images on CUHK03. SQ: Single-Query; MQ: Multi-Query.

Dataset	VIPeR				CUHK01				CUHK03				Market-1501			
Rank (%)	R1	R5	R10	R20	R1	R5	R10	R20	R1	R5	R10	R20	R1(SQ)	mAP(SQ)	R1(MQ)	mAP(MQ)
OneHot Feature Coding	**45.1**	**74.6**	**85.1**	**93.3**	**68.3**	**86.7**	**92.6**	**96.2**	**61.6**	**87.0**	94.6	**98.0**	**57.7**	**29.0**	**68.0**	**37.8**
FDA Feature Coding	**45.1**	**74.6**	**85.1**	**93.3**	**68.3**	**86.7**	**92.6**	**96.2**	60.8	**87.0**	**95.2**	97.7	55.6	27.5	67.5	36.8
Random Feature Coding	44.8	73.4	84.8	92.7	61.3	83.4	89.5	94.2	51.7	79.4	87.4	93.0	47.4	21.1	48.5	23.2

et al (2016b), and *Random Feature Coding* by Zhu et al (2016). In this experiment, we used the LOMO feature on all four datasets, the 485/486 people split on CUHK01, and the manually labelled bounding boxes on CUHK03. For Random Coding, we performed 10 times and used the averaged results to compare with the OneHot Feature Coding and the FDA Feature Coding. The results are presented in Table 6. We have the following observations:

(i) The embedding space choice plays a clear role in IRS re-id model learning and a more "semantic" aligned (both OneHot and FDA) coding has the advantage for learning a more discriminative IRS re-id model. One plausible reason is that the Random Coding may increase the model learning difficulty resulting in an inferior feature embedding, especially given the small sample size nature of re-id model learning. Instead, by explicitly assigning identity class "semantics" (prototypes) to individual dimensions of the embedding space, the feature embedding learning is made more selective and easier to optimise.

(ii) Both the OneHot and FDA Feature Coding methods yield the same re-id accuracy on both VIPeR and CUHK01. This is because on either dataset each training identity has the same number of images (2 for VIPeR and 4 for CUHK01), under which the FDA Coding (Eq. (4)) is equivalent to the OneHot Feature Coding (Eq. (1)).

(iii) Given the different image samples available per training person identity on CUHK03 and Market-1501, FDA Coding is slightly inferior to OneHot Feature Coding. This is interesting given the robust performance of FDA on conventional classification problems. Our explanation is rather straightforward if one considers the unique characteristics of the re-id problem where the training and test classes are *completely* non-overlapping. That is, the test classes have no training image samples. In essence, the re-id problem is conceptually similar to the problem of Zero-Shot Learning (ZSL), in contrast to the conventional classification problems where test classes are sufficiently represented by the training data, i.e. totally overlapping. More specifically, learning by the FDA criterion optimises a model to the training identity classes given sufficient samples per class but it does not work well with small sample sizes, and more critically, it does *not necessarily* optimise the model for previously unseen test identity classes. This is because if the training identity population is relatively small, as in most re-id datasets, an unseen test person may not be similar to any of training people, That is, the distributions of the training and test population may differ significantly. Without any prior knowledge, a good representation of an unseen test class is some unique combination of all training persons *uniformly* without preference. Therefore, a feature embedding optimised uniformly without bias/weighting by the training class data sampling distribution is more likely to better cope with more diverse and unseen test classes, by better preserving class diversity in the training data *especially given the small sample size challenge* in re-id training data. This can be seen from the regularised properties of the OneHot Feature Coding in Sec. 3.

(III) Effect of Features. We evaluated three different features (WHOS, LOMO, and CNN) individually and also their combinations used in our IRS model with the OneHot Feature Coding in Table 7. When a single type of feature is used, it is found that CNN feature is the best except on VIPeR, and LOMO is more discriminative than WHOS in most cases. The advantage of CNN feature over hand-crafted LOMO and WHOS is significant given larger training data in CUHK03 and Market-1501, yielding a *gain* of 19.9% (CUHK03 (Manual)), 26.9% (CUHK03 (Detected)), and 15.0% (Market-1501) over LOMO in Rank-1. Without fine-tuning a CUHK03 trained model on the target domains, CNN feature still performs the best on CUHK01 due to the high similarity in view conditions between CUHK01 and CUHK03. CNN feature performs less well on VIPeR due to higher discrepancy in view conditions between VIPeR and CUHK03, i.e. the domain shift problem (Ma et al, 2013; Pan and Yang, 2010).

We further evaluated multi-feature based performance by score-level fusion. It is evident that most combinations lead to improved re-id accuracy, and fusing all three features often generate the best results. This confirms the previous findings that different appearance information can be encoded by distinct features and their fusion enhances re-id matching (Paisitkriangkrai et al, 2015; Zhang et al, 2016a; Matsukawa et al, 2016; Chen et al, 2016a).

(IV) Deep Learning Regression. Apart from the Ridge Regression (RR) algorithm (Hoerl and Kennard, 1970; Zhang et al, 2010), the IRS concept can be also realised in deep learning, i.e. Deep Learning Regression (DLR). We call this IRS implementation as **IRS(DLR)**. For this experiment, we adopted the DGD CNN model (Xiao et al, 2016) and the CUHK03 (Manual) dataset. In training IRS(DLR), we first

Table 7 Effects of feature choice in re-id performance using the IRS model with OneHot Feature Coding.

Dataset	VIPeR			
Rank (%)	R1	R5	R10	R20
WHOS (Lisanti et al, 2015)	44.5	**75.0**	**86.3**	**93.6**
LOMO (Liao et al, 2015)	**45.1**	74.6	85.1	93.3
CNN (Xiao et al, 2016)	33.1	59.9	71.5	82.2
WHOS+LOMO	53.0	79.8	89.6	95.5
CNN+LOMO	49.9	77.5	86.9	93.8
WHOS+CNN	49.7	78.0	87.9	94.4
WHOS+LOMO+CNN	**54.6**	**81.5**	**90.3**	**95.7**
Dataset	CUHK01 (486/485 split)			
WHOS (Lisanti et al, 2015)	48.8	73.4	81.1	88.3
LOMO (Liao et al, 2015)	68.3	86.7	92.6	96.2
CNN (Xiao et al, 2016)	**68.6**	**89.3**	**93.9**	**97.2**
WHOS+LOMO	71.2	90.0	94.4	97.3
CNN+LOMO	79.8	93.6	96.3	98.2
WHOS+CNN	76.1	92.9	96.1	98.2
WHOS+LOMO+CNN	**80.8**	**94.6**	**96.9**	**98.7**
Dataset	CUHK01 (871/100 split)			
WHOS (Lisanti et al, 2015)	77.0	92.8	96.5	99.2
LOMO (Liao et al, 2015)	80.3	94.2	96.9	99.5
CNN (Xiao et al, 2016)	**84.4**	**98.2**	**99.8**	**100**
WHOS+LOMO	83.6	95.4	98.8	**100**
CNN+LOMO	88.0	98.3	99.5	**100**
WHOS+CNN	**89.0**	98.5	**99.6**	**100**
WHOS+LOMO+CNN	88.4	**98.8**	**99.6**	**100**
Dataset	CUHK03 (Manually)			
WHOS (Lisanti et al, 2015)	59.6	87.2	92.8	96.9
LOMO (Liao et al, 2015)	61.6	87.0	94.6	98.0
CNN (Xiao et al, 2016)	**81.5**	**95.7**	**97.1**	**98.0**
WHOS+LOMO	65.2	92.2	96.8	**99.1**
CNN+LOMO	**82.6**	96.0	97.5	98.6
WHOS+CNN	80.4	95.7	98.0	98.4
WHOS+LOMO+CNN	81.9	**96.5**	**98.2**	98.9
Dataset	CUHK03 (Detected)			
WHOS (Lisanti et al, 2015)	50.6	82.1	90.4	96.1
LOMO (Liao et al, 2015)	53.4	83.1	91.2	96.4
CNN (Xiao et al, 2016)	**80.3**	**96.3**	**98.6**	**99.0**
WHOS+LOMO	59.9	89.4	95.5	98.5
CNN+LOMO	82.4	95.7	97.4	98.4
WHOS+CNN	81.1	95.4	97.5	**98.6**
WHOS+LOMO+CNN	**83.3**	**96.2**	**97.9**	**98.6**
Dataset	Market-1501			
Query Per Person	Single-Query		Multi-Query	
Metric (%)	R1	mAP	R1	mAP
WHOS (Lisanti et al, 2015)	55.2	27.5	60.3	33.5
LOMO (Liao et al, 2015)	57.7	29.0	68.0	37.8
CNN (Xiao et al, 2016)	**72.7**	**48.1**	**80.2**	**58.5**
WHOS+LOMO	62.4	33.6	69.0	41.0
CNN+LOMO	73.0	48.5	80.9	59.1
WHOS+CNN	72.8	48.3	80.3	58.7
WHOS+LOMO+CNN	**73.9**	**49.4**	**81.4**	**59.9**

trained the DGD to convergence with the softmax cross-entropy loss. Then, we added n(=1,2,3) new 512-dim FC layers (including ReLU activation) with random parameter initialisation on top of DGD. Finally, we frozen all original DGD layers and optimised the new layers only by L_2 loss.

In this test, we compared with the DGD (1) CNN Features and (2) Softmax Predictions (considered as some sort of IRS features although not strictly the same due to different modelling designs). We observed in Table 8 that: (1) IRS(DLR) outperforms both CNN Features and Softmax Prediction. This indicates the benefit of IRS in a deep learning context. (2) IRS(DLR) is relatively inferior to CNN+IRS(RR), suggesting that a deep learning model is not necessarily superior in regressing IRS when given limited training data. Moreover, IRS(RR) is superior on model learning efficiency, hence more suitable for incremental model update.

Table 8 Evaluation on deep learning regression (DLR) on CUHK03 (Manually). Deep model: DGD (Xiao et al, 2016). DLR$^{n\text{-FC}}$: $n \in \{1, 2, 3\}$ FC layers added in DLR. RR = Ridge Regression.

Rank (%)	R1	R5	R10	R20
CNN Feature	73.7	91.5	95.0	97.2
Softmax Prediction	73.3	91.0	93.9	96.4
IRS(DLR$^{1\text{-FC}}$)	75.1	92.7	95.3	97.5
IRS(DLR$^{2\text{-FC}}$)	76.6	93.1	95.9	**98.1**
IRS(DLR$^{3\text{-FC}}$)	74.2	92.5	94.8	97.1
CNN + **IRS**(RR)	**81.5**	**95.7**	**97.1**	98.0

Table 9 Evaluation on the complementary effect of deep model pre-training based transfer learning (TL) and IRS on VIPeR. Deep model: DGD (Xiao et al, 2016). $*$: Reported result in (Xiao et al, 2016).

Rank (%)	R1	R5	R10	R20
W/O TL*	12.3	-	-	-
W TL	34.1	66.3	76.2	83.7
TL + **IRS**(RR)	**39.9**	**70.6**	**79.3**	**86.2**

(V) Complementary of Transfer Learning and IRS. Transfer learning (TL) is another independent scheme for solving the SSS problem. We tested the benefit of deep learning pre-trained TL and IRS. We evaluated three methods based on the DGD Xiao et al (2016): (1) ***W/O TL***: Trained the DGD on VIPeR training data (632 images) only. (2) ***W TL***: First pre-trained the DGD on 26,246 CUHK03 images for knowledge transfer learning, then fine-tuned on the VIPeR training data. (3) ***TL + IRS(RR)***: First adopted the CUHK03 pre-trained and VIPeR fine-tuned DGD to extract CNN features, then deployed the ridge regression based IRS to train the final re-id feature embedding model. All three models were evaluated on the same VIPeR test data. Table 9 shows that: (1) Pre-training based TL significantly improves re-id performance. This demonstrates the benefit of TL in solving the SSS problem. (2) IRS clearly further improves the re-id accuracy. This verifies the additional benefits of IRS and the complementary advantage of TL and IRS to a deep learning model for solving the SSS challenge.

(VI) Comparisons to Subspace/Metric Learning Models. We performed comparative experiments on four subspace and metric learning models including KISSME (Koestinger et al, 2012), kLFDA (Xiong et al, 2014), XQDA (Liao et al, 2015), and NFST (Zhang et al, 2016a), using three different types of features (WHOS, LOMO, CNN) and identical training/test data. We utilised the same subspace dimension for XQDA and our IRS, i.e. the number of training person classes. We conducted this evaluation on VIReR and

Table 10 Comparing subspace learning models with different features.

Dataset - Feature	VIPeR - WHOS			
Rank (%)	R1	R5	R10	R20
KISSME (Koestinger et al, 2012)	28.7	57.2	72.6	86.1
kLFDA (Xiong et al, 2014)	40.1	68.5	81.2	91.7
XQDA (Liao et al, 2015)	35.1	63.9	74.9	86.0
NFST (Zhang et al, 2016a)	43.6	74.1	86.1	92.7
IRS	**44.5**	**75.0**	**86.3**	**93.6**
Dataset - Feature	VIPeR - LOMO			
KISSME (Koestinger et al, 2012)	22.1	53.4	68.8	83.8
kLFDA (Xiong et al, 2014)	38.6	69.2	80.4	89.2
XQDA (Liao et al, 2015)	40.0	68.1	80.5	91.1
NFST (Zhang et al, 2016a)	42.3	71.5	82.9	92.1
IRS	**45.1**	**74.6**	**85.1**	**93.3**
Dataset - Feature	VIPeR - CNN			
KISSME (Koestinger et al, 2012)	22.6	46.9	59.0	72.7
kLFDA (Xiong et al, 2014)	30.9	55.6	65.7	75.0
XQDA (Liao et al, 2015)	11.7	26.2	35.5	48.1
NFST (Zhang et al, 2016a)	31.2	56.0	67.2	78.4
IRS	**33.1**	**59.9**	**71.5**	**82.2**
Dataset - Feature	CUHK03(M) - WHOS			
Rank (%)	R1	R5	R10	R20
KISSME (Koestinger et al, 2012)	31.6	63.4	76.6	88.3
kLFDA (Xiong et al, 2014)	32.9	59.2	75.7	82.6
XQDA (Liao et al, 2015)	41.1	66.5	77.2	86.6
NFST (Zhang et al, 2016a)	34.4	59.7	68.2	77.6
IRS	**59.6**	**87.2**	**92.8**	**96.9**
Dataset - Feature	CUHK03(M) - LOMO			
KISSME (Koestinger et al, 2012)	32.7	68.0	81.3	91.4
kLFDA (Xiong et al, 2014)	45.8	77.1	86.8	93.1
XQDA (Liao et al, 2015)	52.2	82.2	92.1	96.3
NFST (Zhang et al, 2016a)	58.9	85.6	92.5	96.3
IRS	**61.6**	**87.0**	**94.6**	**98.0**
Dataset - Feature	CUHK03(M) - CNN			
KISSME (Koestinger et al, 2012)	73.8	94.0	96.2	**98.0**
kLFDA (Xiong et al, 2014)	76.0	92.3	96.0	**98.0**
XQDA (Liao et al, 2015)	70.8	92.0	96.2	97.9
NFST (Zhang et al, 2016a)	62.6	78.9	85.5	89.7
IRS	**81.5**	**95.7**	**97.1**	**98.0**

CUHK03 (Manual). Table 10 shows that the proposed IRS model consistently surpasses all the compared alternative models. This again suggests the advantages of IRS in learning discriminative re-id models.

(VII) Regularisation Sensitivity. We analysed the sensitivity of the only free parameter λ in Eq. (3) which controls the regularisation strength of our IRS model. This evaluation was conducted with the LOMO feature in the multi-query setting on Market-1501 (Zheng et al, 2015). Specifically, we evaluated the Rank-1 and mAP performance with λ varying from 0 to 1. Fig. 7 shows that our IRS model has a large satisfactory range of λ and therefore not sensitive. We set $\lambda = 0.1$ in all evaluations.

(VIII) Model Complexity. In addition to model re-id accuracy, we also examined the model complexity and computational costs, in particular model training time. We carried out this evaluation by comparing our IRS model with some strong metric learning methods including kLFDA (Xiong et al, 2014), XQDA (Liao et al, 2015), MLAPG (Liao and Li, 2015), and NFST (Zhang et al, 2016a). Given n training samples represented by d-dimensional feature vectors, it requires $\frac{3}{2}dnm + \frac{9}{2}m^3$ ($m = \min(d, n)$) floating point addition and multiplications (Penrose, 1955) to perform an eigen-decomposition for solving either a generalised eigen-problem (Xiong et al, 2014; Liao et al, 2015) or a null space (Zhang et al, 2016a), whereas solving the linear system of the IRS model (Eq. (3)) takes $\frac{1}{2}dnm + \frac{1}{6}m^3$ (Cai et al, 2008). Deep learning models (Ahmed et al, 2015; Xiao et al, 2016; Varior et al, 2016a) are not explicitly evaluated since they are usually much more demanding in computational overhead, requiring much more training time (days or even weeks) and more powerful hardware (GPU). In this evaluation, we adopted the LOMO feature for all datasets and all the models compared, the 485/486 people split on CUHK01, the manually labelled person bounding boxes on CUHK03, and the single-query setting on Market-1501.

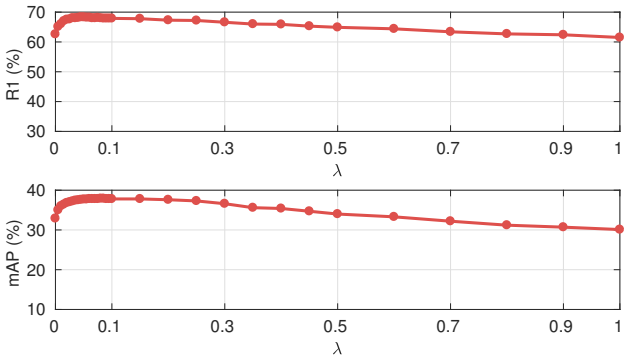


Fig. 7 Regularisation sensitivity on the Market-1501 dataset. The multi-query setting was used.

Table 11 Model complexity and training costs of person re-id models. *Metric*: Model training time (in seconds), smaller is better.

Dataset	VIPeR	CUHK01	CUHK03	Market-1501
Training Size	632	1940	12197	12936
MLAPG	50.9	746.6	4.0×10^4	-
kLFDA	5.0	45.9	2203.2	1465.8
XQDA	4.1	51.9	3416.0	3233.8
NFST	1.3	6.0	1135.1	801.8
IRS	**1.2**	**4.2**	**248.8**	**266.3**

For each model, we recorded and compared the average training time of 10 trials performed on a workstation with 2.6GHz CPU. Table 11 presents the training time of different models (in seconds). On the smaller VIPeR dataset, our IRS model training needed only 1.2 seconds, similar as NFST and 42.4 times faster than MLAPG. On larger datasets CUHK01, CUHK03 and Market-1501, all models took longer time to train and training the IRS model remains the fastest with speed-up over MLAPG enlarged to 177.8 / 160.8 times on CUHK01 / CUHK03, respectively[2]. This demonstrates

[2] The MLAPG model failed to converge on Market-1501.

the advantage of the proposed IRS model over existing competitors for scaling up to large sized training data.

5.2 Incremental Person Re-Id Evaluation

We further evaluated the performance of our IRS model using the incremental learning IRS$^{\text{inc}}$ algorithm (Sec. 4). This setting starts with a small number, e.g. 10 of labelled true match training pairs, rather than assuming a large pre-collected training set. Often, no large sized labelled data is available in typical deployments at varying scenes in advance. More labelled data will arrive one by one over time during deployment due to human-in-the-loop verification. In such a setting, a re-id model can naturally evolve through deployment life-cycle and efficiently adapt to each application test domain. In this context, we consider two incremental re-id model learning scenarios: **(I)** *Passive* incremental learning where unlabelled person images are randomly selected for human to verify; **(II)** *Active* incremental learning where person images are actively determined by the proposed JointE^2 active learning algorithm (Sec. 4.1).

Incremental Re-Id Evaluation Protocol. Due to the lack of access to large sized training samples in batch, incrementally learned models are typically less powerful than batch learned models (Poggio and Cauwenberghs, 2001; Ristin et al, 2014). Therefore, it is critical to evaluate how much performance drop is introduced by the Incremental Learning (IL) algorithm, IRS$^{\text{inc}}$, as compared to the corresponding Batch-wise Learning (BL) and how much efficiency is gained by IL.

We started with 10 labelled identities, i.e. cross-camera truth matches of 10 persons, and set the total labelling budget to 200 persons. For simplicity, we selected four test cases with $50, 100, 150, 200$ labelled identities respectively and evaluated their model accuracy and training cost. To compare the Accumulated Learning Time (ALT)[3], i.e. the summed time for training all the IRS models when the label number is increased from 50 to 200 one by one (in total 151 updates), we interpolated estimations on training time between these four measured test cases. A one-by-one model update is necessary particularly when deploying a pre-trained sub-optimal re-id model to a previously unseen camera network with weak starting performance.

[3] The BL model needs to be trained once only after all 200 person classes are labelled when we consider the batch-wise model learning setting (Sec. 5.1). However, here we consider instead the incremental learning setting with the aim to evaluate the proposed incremental learning algorithm in both training efficiency and effectiveness, as compared to the batch learning counterpart when deployed for model incremental update. Given the batch-wise learning strategy, incremental model update can only be achieved by re-training a model from scratch. Therefore, the accumulated learning time is a rational metric for efficiency comparison in this context.

We adopted the LOMO visual feature on all datasets. We utilised the 485/486 people split on CUHK01, the manually labelled person images on CUHK03, the single-query setting on Market-1501, and the same test data as the experiments in Sec.5.1. We conducted 10 folds of evaluations each with a different set of random unlabelled identities and reported the averaged results.

(I) Passive Incremental Learning. We compared the proposed incremental learning (IL) based IRS (IRS$^{\text{inc}}$) with the batch-wise learning (BL) based IRS in Table 12 for model training time and re-id Rank-1 performance. It is found that IRS model training speed can increase by one order of magnitude or more, with higher speed-up observed on larger datasets and resulting in more model training efficiency gain. Specifically, on VIPeR, BL took approximately 36.5 seconds to conduct the 151 model updates by re-training, while IL only required 3.28 seconds. When evaluated on Market-1501, BL took over 5.5 hours (1.9×10^4 seconds) to perform the sequential model updates, while IL was more than $20\times$ faster, only took 877.3 seconds. Importantly, this speed-up is at the cost of only $1\sim2\%$ Rank-1 drop. This suggests an attractive trade-off for the IRS$^{\text{inc}}$ algorithm between effectiveness and efficiency in incremental model learning.

(II) Active Incremental Learning. We further evaluated the effect of the proposed *JointE*^{2} active learning algorithm (Sec. 4.1) by random passive unlabelled image selection (*Random*). Also, we compared with a state-of-the-art density based active sampling method (Ebert et al, 2012) which prefers to query the densest region of unlabelled sample space (*Density*). For both active sampling methods, we used our IRS$^{\text{inc}}$ for re-id model training. We evaluated the four test cases ($50, 100, 150, 200$ labelled identities) as shown in Table 13.

It is evident from Table 13 that: **(1)** On all four datasets, our JointE^2 outperformed clearly both *Random* and *Density* given varying numbers of labelled samples. For example, when 50 identities were labelled, the proposed JointE^2 algorithm beats *Random* sampling in Rank-1 by $4.0\%(23.4-19.4)$, $9.1\%(29.9–20.8)$, $3.0\%(25.1–22.1)$, $9.0\%(36.5–27.5)$ on VIPeR, CUHK01, CUHK03 and Market-1501, respectively. **(2)** Our JointE^2 model obtained similar or even better performance with less human labelling effort. For example, on Market-1501, by labelling 150 identities, JointE^2 achieved Rank-1 rate of 54.8%, surpassed Random (54.3%) and Density (53.9%) with a greater budget of 200 identities.

In summary, the results in Tables 12 and 13 show clearly that the hybrid of our proposed IRS$^{\text{inc}}$ model and JointE^2 active sampling method provides a highly scalable active incremental re-id model training framework, with attractive model learning capability and efficiency from less labelling effort suited for real-world person re-id applications.

Table 12 Comparing passive Incremental Learning (IL) vs. Batch-wise Learning (BL) using the IRS model. ALT: Accumulated Learning Time, i.e. the summed time for training all the 151 IRS models when the label number is increased from 50 to 200 one by one.

Dataset		VIPeR					CUHK01					CUHK03					Market-1501				
Label #		50	100	150	200	*ALT*	50	100	150	200	*ALT*	50	100	150	200	*ALT*	50	100	150	200	*ALT*
Time	BL	0.23	0.23	0.25	0.26	36.5	1.43	1.51	1.57	1.66	232.8	20.4	21.7	22.4	24.5	3349.9	119.5	121.5	125.6	140.3	1.9×10^4
(sec.)	IL	**0.02**	**0.02**	**0.02**	**0.03**	**3.28**	**0.14**	**0.15**	**0.16**	**0.17**	**23.4**	**1.62**	**1.69**	**1.70**	**1.81**	**257.0**	**1.94**	**5.05**	**6.61**	**9.60**	**877.3**
R1	BL	**20.6**	**29.2**	**34.9**	**38.9**	-	**21.9**	**37.3**	**46.5**	**52.5**	-	**24.0**	**35.2**	**40.5**	**43.8**	-	**28.6**	**44.5**	**51.7**	**55.2**	-
(%)	IL	19.4	**29.2**	33.6	37.2	-	20.8	35.6	45.3	51.5	-	22.1	33.0	38.8	41.7	-	27.5	44.2	50.6	54.3	-

Table 13 Evaluation on the active incremental learning algorithm. *Metric*: Rank-1 rate (%).

Dataset	VIPeR				CUHK01				CUHK03				Market-1501			
Label #	50	100	150	200	50	100	150	200	50	100	150	200	50	100	150	200
Random	19.4	29.2	33.6	37.2	20.8	35.6	45.3	51.5	22.1	33.0	38.8	41.7	27.5	44.2	50.6	54.3
Density (Ebert et al, 2012)	18.4	26.8	33.5	37.5	23.3	37.0	44.5	50.0	23.7	34.8	40.2	42.7	32.3	46.2	51.5	53.9
JointE^2	**23.4**	**31.4**	**36.5**	**40.9**	**29.9**	**39.7**	**47.1**	**52.2**	**25.1**	**36.8**	**41.3**	**43.0**	**36.5**	**50.7**	**54.8**	**58.2**

6 Conclusion

In this work, we developed a novel approach to explicitly designing a feature embedding space for supervised batch-wise and incremental person re-identification model optimisation. We solved the re-id model learning problem by introducing an identity regression method in an Identity Regression Space (IRS) with an efficient closed-form solution. Furthermore, we formulated an incremental learning algorithm IRS$^{\text{inc}}$ to explore sequential on-line labelling and model updating. This enables the model to not only update efficiently the re-id model once new data annotations become available, but also allows probably early re-id deployment and improves adaptively the re-id model to new test domains with potential temporal dynamics. To better leverage human annotation effort, we further derived a novel active learning method JointE^2 to selectively query the most informative unlabelled data on-line. Extensive experiments on four benchmarks show that our IRS method outperforms existing state-of-the-art re-id methods in the conventional batch-wise model learning setting. Moreover, the proposed incremental learning algorithm increases significantly model training speed, over 10 times faster than batch-wise model learning, by only sacrificing marginal model re-id capability with $1{\sim}2\%$ Rank-1 drop. This labelling-while-deploying strategy has the intrinsic potential of helping reduce the cost of manual labelling in large scale deployments by structuring semantically the unlabelled data so to expedite the true match identification process. Additionally, our active learning method improves notably the human labelling quality w.r.t. the thus-far model, particularly when limited budget is accessible, providing over 3% Rank-1 improvement than Random sampling given 50 identities labelling budget. While person re-id has attracted increasing amount of efforts especially in the deep learning paradigm, model learning scalability, model incremental adaptation, and labelling effort minimisation in large scale deployments however are significantly underestimated although very critical in real-world applications. By presenting timely an effective solution in this work, we hope that more investigations towards these important problems will be made in the future studies. One interesting future direction is to develop incremental deep re-id learning algorithms.

Acknowledgments

This work was partially supported by the China Scholarship Council, Vision Semantics Ltd, Royal Society Newton Advanced Fellowship Programme (NA150459), and Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111-571149).

A Derivation of FDA Coding

In the following, we provide a detailed derivation of FDA coding (Eq. (4)) in our IRS method.

FDA Criterion. Specifically, the FDA criterion aims to minimise the intra-class (person) appearance variance and maximise inter-class appearance variance. Formally, given zero-centred training data $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$, we generate three scatter matrices defined as follows:

$$\begin{aligned} \boldsymbol{S}_w &= \frac{1}{n}\sum_{j=1}^{c}\sum_{l_i=j}(\boldsymbol{x}_i - \boldsymbol{u}_j)(\boldsymbol{x}_i - \boldsymbol{u}_j)^\top, \\ \boldsymbol{S}_b &= \frac{1}{n}\sum_{j=1}^{c} n_j \boldsymbol{u}_j \boldsymbol{u}_j^\top, \\ \boldsymbol{S}_t &= \boldsymbol{S}_w + \boldsymbol{S}_b = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top, \end{aligned} \tag{19}$$

where $\boldsymbol{S}_w$, $\boldsymbol{S}_b$, and $\boldsymbol{S}_t$ denote *within-class*, *between-class* and *total* scatter matrices respectively, $\boldsymbol{u}_j$ the class-wise centroids, and n_j the sample size of the j-th class (or person). The objective function of FDA aims at maximising $trace(\boldsymbol{S}_b)$ and minimising $trace(\boldsymbol{S}_w)$ simultaneously, where $\boldsymbol{S}_w$ can be replaced by $\boldsymbol{S}_t$ since $\boldsymbol{S}_t = \boldsymbol{S}_b + \boldsymbol{S}_w$. Hence,

an optimal transformation $\boldsymbol{G}^*$ by FDA can be computed by solving the following problem:

$$\boldsymbol{G}^* = \arg\max_{\boldsymbol{G}} \ trace\Big(\big(\boldsymbol{G}^\top \boldsymbol{S}_b \boldsymbol{G}\big)\big(\boldsymbol{G}^\top \boldsymbol{S}_t \boldsymbol{G}\big)^\dagger\Big). \tag{20}$$

Theorem 1. *With $\boldsymbol{Y}$ defined as Eq.* (4)*, the projection $\boldsymbol{P}^*$ learned by Eq.* (3) *is equivalent to $\boldsymbol{G}^*$, the optimal FDA solution in Eq.* (20)*.*

Proof. First, optimising the objective in Eq. (4) involves solving the following eigen-problem:

$$\boldsymbol{S}_t^\dagger \boldsymbol{S}_b \boldsymbol{G} = \boldsymbol{G}\boldsymbol{\Lambda}, \tag{21}$$

where $\boldsymbol{G} \in \mathbb{R}^{d\times q} = [\boldsymbol{g}_1, \cdots, \boldsymbol{g}_q]$ contains q eigenvectors of $\boldsymbol{S}_t^\dagger \boldsymbol{S}_b$, and $\boldsymbol{\Lambda} = diag(\alpha_1, \cdots, \alpha_q)$ with α_i the corresponding eigenvalue, and $q = rank(\boldsymbol{S}_b) \leq c-1$. From the definitions in Eq. (19) and Eq. (4), $\boldsymbol{S}_t$ and $\boldsymbol{S}_b$ can be further expanded as:

$$\boldsymbol{S}_t = \boldsymbol{X}\boldsymbol{X}^\top, \quad \boldsymbol{S}_b = \boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^\top\boldsymbol{X}^\top. \tag{22}$$

Here, the multiplier $\frac{1}{n}$ is omitted in both scatter matrices for simplicity. Now, we can rewrite the left-hand side of Eq. (21) as:

$$(\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger \boldsymbol{X}\boldsymbol{Y}\boldsymbol{Y}^\top\boldsymbol{X}^\top\boldsymbol{G} = \boldsymbol{G}\boldsymbol{\Lambda}. \tag{23}$$

Note that, the pseudo-inverse $\boldsymbol{S}_t^\dagger$ is calculated by $(\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger$. The reason is that in real-world problems such as person re-id where training data is often less sufficient, $\boldsymbol{S}_t$ is likely to be ill-conditioned, i.e. singular or close to singular, so that its inverse cannot be accurately computed.

By our solution $\boldsymbol{P}$ in Eq. (3), we can further rewrite Eq. (23):

$$\boldsymbol{P}\boldsymbol{Y}^\top\boldsymbol{X}^\top\boldsymbol{G} = \boldsymbol{G}\boldsymbol{\Lambda} \tag{24}$$

To connect the regression solution $\boldsymbol{P}$ and the FDA solution $\boldsymbol{G}$, we define a $c \times c$ matrix $\boldsymbol{R} = \boldsymbol{Y}^\top\boldsymbol{X}^\top\boldsymbol{P}$. According to the general property of eigenvalues (Horn and Johnson, 2012), $\boldsymbol{R}$ and $\boldsymbol{P}\boldsymbol{Y}^\top\boldsymbol{X}^\top$ share the same q non-zero eigenvalues. Also, if $\boldsymbol{V} \in \mathbb{R}^{c\times q}$ contains the q eigenvectors of $\boldsymbol{R}$, columns of the matrix $\boldsymbol{P}\boldsymbol{V}$ must be the eigenvectors of the matrix $\boldsymbol{P}\boldsymbol{Y}^\top\boldsymbol{X}^\top$. Therefore, the relation between $\boldsymbol{P}$ and $\boldsymbol{G}$ is:

$$\boldsymbol{G} = \boldsymbol{P}\boldsymbol{V} \tag{25}$$

Finally, we show in the following Lemma that $\boldsymbol{P}$ and $\boldsymbol{G}$ are equivalent in the aspect of re-id matching.

Lemma 1. *In the embedding provided by $\boldsymbol{P}$ and $\boldsymbol{G}$, the nearest neighbour algorithm produce same result. That is, $(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top\boldsymbol{P}\boldsymbol{P}^\top(\boldsymbol{x}_i - \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top\boldsymbol{G}\boldsymbol{G}^\top(\boldsymbol{x}_i - \boldsymbol{x}_j)$.*

Proof. The necessary and sufficient condition for Lemma 1 is $\boldsymbol{P}\boldsymbol{P}^\top = \boldsymbol{G}\boldsymbol{G}^\top$. As $\boldsymbol{V} \in \mathbb{R}^{c\times q}$, there must exist a matrix $\boldsymbol{V}_2 \in \mathbb{R}^{c\times(c-q)}$ such that $\hat{\boldsymbol{V}} = [\boldsymbol{V}, \boldsymbol{V}_2]$ is a $c \times c$ orthogonal matrix. Suppose the diagonal matrix $\boldsymbol{\Gamma}$ contains the non-zero eigenvalues of $\boldsymbol{R}$, then the eigen decomposition $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}^\top$ implies that $\boldsymbol{V}_2^\top\boldsymbol{R}\boldsymbol{V}_2 = 0$.

Recall that $\boldsymbol{R} = \boldsymbol{Y}^\top\boldsymbol{X}^\top\boldsymbol{P}$, and $\boldsymbol{P} = (\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger\boldsymbol{X}\boldsymbol{Y}$, then we obtain:

$$\boldsymbol{V}_2^\top\boldsymbol{Y}^\top\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger\boldsymbol{X}\boldsymbol{Y}\boldsymbol{V}_2 = 0 \tag{26}$$

As $(\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger$ is positive definite, the above equation implies that $\boldsymbol{X}\boldsymbol{Y}\boldsymbol{V}_2 = 0$, and hence $\boldsymbol{P}\boldsymbol{V}_2 = (\boldsymbol{X}\boldsymbol{X}^\top + \lambda\boldsymbol{I})^\dagger\boldsymbol{X}\boldsymbol{Y}\boldsymbol{V}_2 = 0$. Hence, we have:

$$\begin{aligned}\boldsymbol{P}\boldsymbol{P}^\top &= \boldsymbol{P}\hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^\top\boldsymbol{P}^\top \\ &= \boldsymbol{P}\boldsymbol{V}\boldsymbol{V}^\top\boldsymbol{P}^\top + \boldsymbol{P}\boldsymbol{V}_2\boldsymbol{V}_2^\top\boldsymbol{P}^\top \\ &= \boldsymbol{G}\boldsymbol{G}^\top + 0\end{aligned} \tag{27}$$

As such, the proof to Lemma 1 and Theorem 1 is complete.

References

Ahmed E, Jones MJ, Marks TK (2015) An improved deep learning architecture for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Akaike H (1998) Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike, Springer, pp 199–213

Cai D, He X, Han J (2008) Srda: An efficient algorithm for large-scale discriminant analysis. IEEE Transactions on Knowledge and Data Engineering 20(1):1–12

Cebron N, Berthold MR (2009) Active learning for object classification: from exploration to exploitation. Data Mining and Knowledge Discovery 18(2):283–299

Chen D, Yuan Z, Chen B, Zheng N (2016a) Similarity learning with spatial constraints for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Chen LF, Liao HYM, Ko MT, Lin JC, Yu GJ (2000) A new lda-based face recognition system which can solve the small sample size problem. Pattern Recognition 33(10):1713–1726

Chen SZ, Guo CC, Lai JH (2016b) Deep ranking for person re-identification via joint representation learning. IEEE Transactions on Image Processing 25(5):2353–2367

Chen YC, Zheng WS, Yuen PC, Lai J (2016c) An asymmetric distance model for cross-view feature mapping in person re-identification. In: IEEE Transactions on Circuits and Systems for Video Technology, vol PP, pp 1–1

Cheng D, Gong Y, Zhou S, Wang J, Zheng N (2016) Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: IEEE Conference on Computer Vision and Pattern Recognition

Das A, Panda R, Roy-Chowdhury A (2015) Active image pair selection for continuous person re-identification. In: IEEE International Conference on Image Processing

Duda RO, Hart PE, Stork DG (2012) Pattern Classification. John Wiley & Sons

Ebert S, Fritz M, Schiele B (2012) Ralf: A reinforced active learning formulation for object class recognition. In: IEEE Conference on Computer Vision and Pattern Recognition

Farenzena M, Bazzani L, Perina A, Murino V, Cristani M (2010) Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference on Computer Vision and Pattern Recognition

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2):179–188

Fukunaga K (2013) Introduction to statistical pattern recognition. Academic Press

Geng M, Wang Y, Xiang T, Tian Y (2016) Deep transfer learning for person re-identification. arXiv preprint arXiv:161105244

Gong S, Cristani M, Yan S, Loy CC (2014) Person re-identification. Springer

Gray D, Brennan S, Tao H (2007) Evaluating appearance models for recognition, reacquisition and tracking. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance

Guo YF, Wu L, Lu H, Feng Z, Xue X (2006) Null foley–sammon transform. Pattern Recognition 39(11):2248–2251

Hardoon DR, Szedmak S, Szedmak O, Shawe-taylor J (2007) Canonical correlation analysis; an overview with application to learning methods

Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 27(2):83–85

Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12(1):55–67

Horn RA, Johnson CR (2012) Matrix Analysis. Cambridge university press

Hospedales TM, Gong S, Xiang T (2012) A unifying theory of active discovery and learning. In: European Conference on Computer Vision

Joshi AJ, Porikli F, Papanikolopoulos N (2009) Multi-class active learning for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition

Käding C, Freytag A, Rodner E, Bodesheim P, Denzler J (2015) Active learning and discovery of object categories in the presence of unnameable instances. In: IEEE Conference on Computer Vision and Pattern Recognition

Kang C, Xiang S, Liao S, Xu C, Pan C (2015) Learning consistent feature representation for cross-modal multimedia retrieval. IEEE Transactions on Multimedia 17(3):370–381

Kang J, Ryu KR, Kwon HC (2004) Using cluster-based sampling to select initial training set for active learning in text classification. In: Advances in Knowledge Discovery and Data Mining

Karanam S, Gou M, Wu Z, Rates-Borras A, Camps O, Radke RJ (2016) A comprehensive evaluation and benchmark for person re-identification: Features, metrics, and datasets. arXiv e-print

Kodirov E, Xiang T, Fu Z, Gong S (2016) Person re-identification by unsupervised l1 graph learning. In: European Conference on Computer Vision

Koestinger M, Hirzer M, Wohlhart P, Roth PM, Bischof H (2012) Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition

Land EH, McCann JJ (1971) Lightness and retinex theory. Journal of the Optical Society of America 61(1):1–11

Layne R, Hospedales TM, Gong S (2013) Domain transfer for person re-identification. In: Workshop of ACM International Conference on Multimedia, Barcelona, Catalunya, Spain

Li W, Zhao R, Wang X (2012) Human reidentification with transferred metric learning. In: Asian Conference on Computer Vision

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

Li W, Zhu X, Gong S (2017) Person re-identification by deep joint learning of multi-loss classification. In: International Joint Conference of Artificial Intelligence

Li Z, Chang S, Liang F, Huang T, Cao L, Smith J (2013) Learning locally-adaptive decision functions for person verification. In: IEEE Conference on Computer Vision and Pattern Recognition

Liao S, Li SZ (2015) Efficient psd constrained asymmetric metric learning for person re-identification. In: IEEE International Conference on Computer Vision

Liao S, Zhao G, Kellokumpu V, Pietikäinen M, Li SZ (2010) Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: IEEE Conference on Computer Vision and Pattern Recognition

Liao S, Mo Z, Zhu J, Hu Y, Li SZ (2014) Open-set person re-identification. arXiv preprint arXiv:14080872

Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by local maximal occurrence representation and metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition

Lin Y, Lv F, Zhu S, Yang M, Cour T, Yu K, Cao L, Huang T (2011) Large-scale image classification: fast feature extraction and svm training. In: IEEE Conference on Computer Vision and Pattern Recognition

Lisanti G, Masi I, Del Bimbo A (2014) Matching people across camera views using kernel canonical correlation analysis. In: ACM International Conference on Distributed Smart Cameras

Lisanti G, Masi I, Bagdanov AD, Del Bimbo A (2015) Person re-identification by iterative re-weighted sparse ranking. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(8):1629–1642

Liu C, Loy CC, Gong S, Wang G (2013) Pop: Person re-identification post-rank optimisation. In: IEEE International Conference on Computer Vision, Sydney, Australia

Loy CC, Hospedales TM, Xiang T, Gong S (2012) Stream-based joint exploration-exploitation active learning. In: IEEE Conference on Computer Vision and Pattern Recognition

Ma AJ, Yuen PC, Li J (2013) Domain transfer support vector ranking for person re-identification without target camera label information. In: IEEE International Conference on Computer Vision

Ma B, Su Y, Jurie F (2012) Local descriptors encoded by fisher vectors for person re-identification. In: Workshop of European Conference on Computer Vision

Martinel N, Das A, Micheloni C, Roy-Chowdhury AK (2016) Temporal model adaptation for person re-identification. In: European Conference on Computer Vision

Matsukawa T, Okabe T, Suzuki E, Sato Y (2016) Hierarchical gaussian descriptor for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

McLaughlin N, Martinez Del Rincon J, Miller P (2015) Data-augmentation for reducing dataset bias in person re-identification. In: Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on, IEEE, pp 1–6

Mika S, Ratsch G, Weston J, Scholkopf B, Mullers KR (1999) Fisher discriminant analysis with kernels. In: Proceedings of IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing., pp 41–48

Moghaddam B, Jebara T, Pentland A (2000) Bayesian face recognition. Pattern Recognition 33(11):1771–1782

Osugi T, Kim D, Scott S (2005) Balancing exploration and exploitation: A new algorithm for active machine learning. In: IEEE International Conference on Data Mining

Paige CC, Saunders MA (1982) Lsqr: An algorithm for sparse linear equations and sparse least squares. ACM Transactions on Mathematical Software 8(1):43–71

Paisitkriangkrai S, Shen C, van den Hengel A (2015) Learning to rank in person re-identification with metric ensembles. In: IEEE Conference on Computer Vision and Pattern Recognition

Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10):1345–1359

Park CH, Park H (2005) A relationship between linear discriminant analysis and the generalized minimum squared error solution. SIAM Journal on Matrix Analysis and Applications 27(2):474–492

Pedagadi S, Orwell J, Velastin SA, Boghossian BA (2013) Local fisher discriminant analysis for pedestrian re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Peng P, Xiang T, Wang Y, Pontil M, Gong S, Huang T, Tian Y (2016) Unsupervised cross-dataset transfer learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1306–1315

Penrose R (1955) A generalized inverse for matrices. In: Proc. Cambridge Philos. Soc, Cambridge Univ Press, vol 51, pp 406–413

Poggio T, Cauwenberghs G (2001) Incremental and decremental support vector machine learning. In: Advances in Neural Information Processing Systems

Prosser B, Zheng WS, Gong S, Xiang T (2010) Person re-identification by support vector ranking. In: British Machine Vision Conference

Ristin M, Guillaumin M, Gall J, Van Gool L (2014) Incremental learning of ncm forests for large-scale image classification. In: IEEE Conference on Computer Vision and Pattern Recognition

Settles B (2012) Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning 6(1):1–114

Settles B, Craven M (2008) An analysis of active learning strategies for sequence labeling tasks. In: The Conference on Empirical Methods in Natural Language Processing

Sharma A, Kumar A, Daume H, Jacobs DW (2012) Generalized multiview analysis: A discriminative latent space. In: IEEE Conference on Computer Vision and Pattern Recognition

Shi H, Yang Y, Zhu X, Liao S, Lei Z, Zheng W, Li SZ (2016) Embedding deep metric for person re-identification: A study against large variations. In: European Conference on Computer Vision

Su C, Zhang S, Xing J, Gao W, Tian Q (2016) Deep attributes driven multi-camera person re-identification. In: European Conference on Computer Vision

Varior RR, Haloi M, Wang G (2016a) Gated siamese convolutional neural network architecture for human re-identification. In: European Conference on Computer Vision

Varior RR, Shuai B, Lu J, Xu D, Wang G (2016b) A siamese long short-term memory architecture for human re-identification. In: European Conference on Computer Vision

Wang F, Zuo W, Lin L, Zhang D, Zhang L (2016a) Joint learning of single-image and cross-image representations for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Wang H, Gong S, Xiang T (2014a) Unsupervised learning of generative topic saliency for person re-identification. In: British Machine Vision Conference

Wang H, Gong S, Xiang T (2016b) Highly efficient regression for scalable person re-identification. In: British Machine Vision Conference

Wang H, Gong S, Zhu X, Xiang T (2016c) Human-in-the-loop person re-identification. In: European Conference on Computer Vision

Wang T, Gong S, Zhu X, Wang S (2014b) Person re-identification by video ranking. In: European Conference on Computer Vision

Wang T, Gong S, Zhu X, Wang S (2016d) Person re-identification by discriminative selection in video ranking. IEEE Transactions on Pattern Analysis and Machine Intelligence 38(12):2501–2514

Wang X, Han TX, Yan S (2009) An hog-lbp human detector with partial occlusion handling. In: IEEE International Conference on Computer Vision

Wang Z, Du B, Zhang L, Zhang L, Fang M, Tao D (2016e) Multi-label active learning based on maximum correntropy criterion: Towards robust and discriminative labeling. In: European Conference on Computer Vision

Woodbury MA (1950) Inverting modified matrices. Memorandum Report 42:106

Wu S, Chen YC, Li X, Wu AC, You JJ, Zheng WS (2016) An enhanced deep feature representation for person re-identification. In: IEEE Winter Conference on Applications of Computer Vision

Xiao T, Li H, Ouyang W, Wang X (2016) Learning deep feature representations with domain guided dropout for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Xiong F, Gou M, Camps O, Sznaier M (2014) Person re-identification using kernel-based metric learning methods. In: European Conference on Computer Vision

Yan S, Xu D, Zhang B, Zhang HJ, Yang Q, Lin S (2007) Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(1):40–51

Yang L, Jin R (2006) Distance metric learning: A comprehensive survey. Michigan State Universiy 2(2)

Yang Y, Yang J, Yan J, Liao S, Yi D, Li SZ (2014) Salient color names for person re-identification. In: European Conference on Computer Vision

Yi D, Lei Z, Li SZ (2014) Deep metric learning for practical person re-identification. arXiv e-print

Yu H, Yang J (2001) A direct lda algorithm for high-dimensional datawith application to face recognition. Pattern Recognition 34(10):2067–2070

Zhang L, Xiang T, Gong S (2016a) Learning a discriminative null space for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Zhang Y, Shao M, Wong EK, Fu Y (2013) Random faces guided sparse many-to-one encoder for pose-invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition

Zhang Y, Li B, Lu H, Irie A, Ruan X (2016b) Sample-specific svm learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Zhang Z, Dai G, Xu C, Jordan MI (2010) Regularized discriminant analysis, ridge regression and beyond. The Journal of Machine Learning Research 11:2199–2228

Zhao R, Ouyang W, Wang X (2013) Unsupervised salience learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Zhao R, Ouyang W, Wang X (2014) Learning mid-level filters for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition

Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: A benchmark. In: IEEE International Conference on Computer Vision

Zheng L, Yang Y, Hauptmann AG (2016) Person re-identification: Past, present and future. arXiv e-print

Zheng W, Zhao L, Zou C (2005) Foley-sammon optimal discriminant vectors using kernel approach. IEEE Transactions on Neural Networks 16(1):1–9

Zheng WS, Gong S, Xiang T (2013) Reidentification by relative distance comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(3):653–668

Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: IEEE International Conference on Computer Vision

Zhu F, Xie J, Fang Y (2016) Heat diffusion long-short term memory learning for 3d shape analysis. In: European Conference on Computer Vision