Multi-scale Learning for Low-resolution Person Re-identification

Xiang Li[1], Wei-Shi Zheng*[1], Xiaojuan Wang[1], Tao Xiang[2], and Shaogang Gong[2]
[1]School of Information Science and Technology, Sun Yat-sen University, China
[2]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
lixiang651@gmail.com, wszheng@ieee.org, xiaojuanwang.cs@gmail.com
t.xiang@qmul.ac.uk, s.gong@qmul.ac.uk

Abstract

In real world person re-identification (re-id), images of people captured at very different resolutions from different locations need be matched. Existing re-id models typically normalise all person images to the same size. However, a low-resolution (LR) image contains much less information about a person, and direct image scaling and simple size normalisation as done in conventional re-id methods cannot compensate for the loss of information. To solve this LR person re-id problem, we propose a novel joint multi-scale learning framework, termed **j**oint m**u**lti-scale **d**iscriminant compon**e**nt **a**nalysis *(JUDEA). The key component of this framework is a* heterogeneous class mean discrepancy *(HCMD) criterion for cross-scale image domain alignment, which is optimised simultaneously with discriminant modelling across multiple scales in the joint learning framework. Our experiments show that the proposed JUDEA framework outperforms existing representative re-id methods as well as other related LR visual matching models applied for the LR person re-id problem.*

1. Introduction

Person re-identification (re-id) is a task of matching pedestrians observed from non-overlapping camera views in a surveillance system. A significant challenge for person re-id is that people are often captured in different camera views at significantly different distances to the cameras, resulting in very different image resolutions. An example is shown in Fig. 1. In the first camera view (top image) the target person walked close to the camera with his appearance details clearly visible in the captured normal resolution image, while the resolution of his image becomes much lower when he reappeared in a different view (bottom image) and was much further away from the camera. This difference in resolution between matching views, compounded by changes in lighting, pose and occlusion, makes re-identification extremely hard and unreliable.

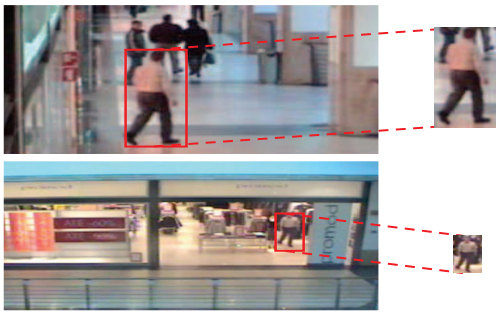


Figure 1. A typical person re-id scenario. The resolution of a person's images in two different camera views are significantly different, beyond the scope for simple image size normalisation by interpolation.

Although resolution difference is a common problem for re-id, it is largely ignored by existing approaches. In particular, existing methods focus on solving the challenges caused by view, pose, and lighting changes by exploring invariant and discriminant features [25, 8, 6, 33, 7, 18, 13, 21, 40, 16] or developing reliable and robust distance metrics [8, 11, 27, 41, 23, 31, 26, 17, 39, 38, 35]. When it comes to the low-resolution (LR) person re-id problem, that is, matching LR person images to normal (higher) resolution ones, most (if not all) methods would simply normalize input images to a uniform normal scale. However LR person images contain much less information than those of normal resolution and many appearance details have been lost. A simple image magnification by interpolation thus would not recover the lost information in the LR person images. Other Bag of Words (BoW) based methods [42, 21] do not explicitly require any normalisation of image size. Nevertheless, the small number of keypoints for BoW computation in LR person images will still lead to the loss of image details. Existing re-id models therefore do not offer a solution to the LR person re-id problem.

In this work, for the first time, a principled solution to L-

*Corresponding author

R person re-id problem is provided. Rather than re-scaling each LR image to a normal scale as in conventional re-id, or directly matching a pair of LR and normal (higher) resolution images, we aim to learn a discriminant model for LR person re-id jointly across different image scales to exploit the correlation of a person's appearance at different scales. More specifically, let us consider images of different scales belonging to different domains. We assume that images of the same person in significantly different scales shall distribute intrinsically in a similar structure in a latent space, provided that cross-scale common features can be extracted among these heterogeneous image domains. To that end, we propose a heterogeneous class mean discrepancy (HCMD) criterion. Minimising this criterion leads to the learning of the latent subspace which is capable of aligning the distributions of image features from significantly different image scales of the same person. Through this *cross-scale image domain alignment* process, the shared discriminant information can be propagated between the normal resolution person images and the LR ones. The HCMD-based cross-scale image domain alignment is optimised simultaneously with discriminant distance metric modelling in each scale in our joint learning framework, which is termed ***j**oint **m**ulti-scale **d**iscriminant compon**e**nt **a**nalysis* (JUDEA).

Our contributions are twofold: 1) To the best our knowledge, this is the first work focusing on solving the LR person re-id problem. Our learning-based framework is much more principled than existing approaches of image scale normalisation; 2) we introduce a new multi-scale discriminant distance metric learning model which simultaneously minimises a novel heterogeneous class mean discrepancy criterion (HCMD) for cross-scale image domain alignment, so they can benefit each other in such a joint learning model.

Extensive experiments are conducted on three datasets to validate the effectiveness of the proposed model. They include a LR person re-id dataset from the CAVIAR dataset [5] and two simulated LR datasets constructed from the VIPeR [8] and 3DPES datasets [1]. Our results demonstrate that the conventional approach of image scaling is not suitable for the LR person re-id problem, and the proposed approach is much more effective. In addition, the proposed JUDEA model outperforms a number of related alternative LR image matching methods designed for other visual recognition problems such as face recognition.

2. Related Work

Although LR image matching, particularly matching LR images against normal (higher) resolution images, has not been studied in person re-id, it has been investigated intensively in face recognition. Many LR face recognition methods exploit super-resolution (SR) techniques to obtain high-resolution (HR) images before matching, with numerous learning-based face SR algorithms been studied in the last decade [9, 4, 19, 22, 34, 36]. However, most of these methods require accurate and dense alignment of LR and HR images. It is possible for face images; but it is much more costly and difficult to obtain sufficient labelled and perfectly aligned pair-wise person images across non-overlapping camera views in order to cover the varied intra-class changes for learning effective SR models. This is due to the significantly greater degree of unknown changes in body parts between the probe and gallery images, e.g. matching between a LR person image with a backpack with a normal resolution one from the frontal view. These SR-based LR image matching methods are thus not suitable for the LR person re-id problem, as validated by our experiments (see Sec. 4.2.4).

In the last five years there are several coupled transformation based subspace models developed for LR face recognition [30, 29, 2, 24, 43, 14]. A basic idea of these works is to learn coupled transformations such that a LR image can directly match a HR image. Our approach differs from these transformation-based methods in that we do not explicitly match an LR body image with a normal resolution one. This is because, due to the misalignment problem in person re-id as mentioned above, in practice body images of different resolution always look more different as compared to face images and thus there is no direct correspondence between low and normal (higher) resolution body images. Instead, our model simultaneously extracts discriminant projections on different scale image spaces, and further aligns their distributions via cross-scale heterogeneous transfer modelling in a latent feature space.

The proposed HCMD criterion is related to the maximum mean discrepancy (MMD) [3] method. However, our cross domain alignment based on HCMD is under a heterogeneous setting. In contrast, the existing MMD-based domain adaptation methods [3] focus on the homogeneous case, where the dimensions of the two domains must be the same, which is not the case for our problem. They are thus not applicable for our problem. There are also related heterogeneous domain adaptation methods [15, 12, 32, 28]. However, for person re-id, in practice people in the training set will not appear in testing set. So some of the heterogeneous domain adaptation method [15, 28] cannot be applied for re-id since they assume that the training and test sets contain the same classes. The most closely related approach to ours is the manifold based alignment model in [32]. Our experiments show that the proposed model outperforms this manifold based alignment method.

3. Methodology

3.1. Problem Formulation

Our aim is to match a LR probe image of a person against normal even high resolution gallery images. Instead

of directly matching them, our solution is a joint multi-scale person re-id framework which compute a discriminant subspace where images of different scale can be matched more accurately. The basic idea is that, despite exhibited in different image scales, all images of the same person are assumed to be distributed intrinsically in similar structures in a latent low-rank dimensional space. Therefore, one is able to utilise information extracted from the normal resolution images in order to assist the learning of discriminant distance metrics for the LR images through a joint multi-scale learning model. In the following, we first present the idea on heterogeneous domain alignment and then it will be integrated into our joint multi-scale learning model.

Without loss of generality, we present a two-scale formulation (see Fig. 2). For convenience, we denote the two scales as normal scale and small scale respectively. A multi-scale formulation can be readily generalised.

Suppose a pairwise training set is given, $\boldsymbol{X}_h = \{(\boldsymbol{x}_i^h, y_i)\}_{i=1}^N$ and $\boldsymbol{X}_s = \{(\boldsymbol{x}_i^s, y_i)\}_{i=1}^N$, where $\boldsymbol{x}_i^h \in \mathbb{R}^{d_h}$ is the feature vector extracted from a person image of normal scale, and $\boldsymbol{x}_i^s \in \mathbb{R}^{d_s}$ is the feature vector extracted from the same image of small scale. N is the total number of samples in the training set. $\boldsymbol{x}_i^h$ and $\boldsymbol{x}_i^s$ are labelled as class y_i, and we have $d_h > d_s$, that is, we have two heterogeneous image domains with different feature representations.

3.2. Cross-scale Image Domain Alignment

We assume that different visual appearance variations of the same person at different image scales are similar in a latent low-rank dimensional space. In other words, we assume that the intrinsic structures of data distributions of a person's appearance across image scales are similar in a feature space. We wish to exploit the shared features across the image scales. To that end, one needs to measure the differences of data distributions of the same person across image scales after projecting the image features of different scales into a common low-rank subspace. We call this process the *cross-scale image domain alignment*. Since images of different scales are more likely to be described differently by their features, learning scale-specific projections is required in order to satisfy this cross-scale alignment. To that end, we define a *heterogeneous class mean discrepancy* (HCMD) for minimisation, as follows:

$$\min_{\boldsymbol{W}_h, \boldsymbol{W}_s} \text{HCMD}(\boldsymbol{W}_h, \boldsymbol{W}_s) = \frac{1}{C}\sum_{i=1}^{C} ||\boldsymbol{W}_h^T \boldsymbol{u}_i^h - \boldsymbol{W}_s^T \boldsymbol{u}_i^s||_2^2, \quad (1)$$

where $\boldsymbol{u}_i^h$ is the mean feature vector of images of the ith person/class in the normal scale domain $\boldsymbol{X}_h$, $\boldsymbol{u}_i^s$ is the corresponding mean feature vector of the images in the small scale domain $\boldsymbol{X}_s$, C is the number of classes, $\boldsymbol{W}_h \in \mathbb{R}^{d_h \times r}$ and $\boldsymbol{W}_s \in \mathbb{R}^{d_s \times r}$ denote the projections on two image scales, respectively, where r is the dimensionality of the projected low-rank subspace.

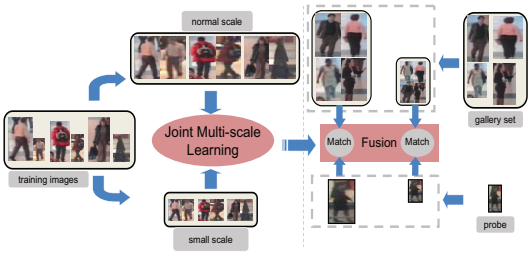


Figure 2. A joint multi-scale learning framework for low-resolution person re-id problem.

This HCMD criterion is inspired by the concept of the maximum mean discrepancy (MMD) [3], which measures distribution difference by computing the distance between total-class data means across domains, defined by

$$\min_{\phi} \text{Dist}(\boldsymbol{X}, \boldsymbol{Y}) = ||\frac{1}{n_1}\sum_{i=1}^{n_1}\phi(\boldsymbol{x}_i) - \frac{1}{n_2}\sum_{i=1}^{n_2}\phi(\boldsymbol{y}_i)||_2^2, \quad (2)$$

where the two domains are $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_1}\}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n_2}\}$, and function ϕ is a mapping.

However, HCMD is different from MMD in two aspects: First, HCMD considers the alignment between image spaces with different dimensions (e.g. $d_h \neq d_s$), while MMD is constrained to the alignment between two domains of the same dimension. Second, MMD pools data of two domains together in an unsupervised way by minimising the difference of the total-class data means of two domains; in contrast, HCMD pools the same class data of the two domains together in a supervised way by minimising the difference of the same class data means from the two domains. The distributions of the same person images across different scales can be similar, but the distributions between images of different classes/people across scales are not.

3.3. Multi-scale Discriminant Learning

For data distribution alignment across image scales in a low-rank dimensional space, we aim to learn a discriminant metric for each scale. The idea is to ensure on each scale, the intra-class distance is minimised whilst the inter-class distance is maximised during cross-scale data distribution alignment. These discriminant information can be described by the following inter-class scatter matrix $\mathbf{S}_b$ and intra-class scatter matrix $\mathbf{S}_w$:

$$\boldsymbol{S}_b = \sum_{i,j=1}^{N} \frac{\overline{\boldsymbol{A}}_{i,j}^b}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T, \; \boldsymbol{S}_w = \sum_{i,j=1}^{N} \frac{\overline{\boldsymbol{A}}_{i,j}^w}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T$$

Here we specifically incorporate weights for each pair of samples $\boldsymbol{x}_i, \boldsymbol{x}_j$ based on their affinity $\boldsymbol{A}_{i,j}$ [37], where $\overline{\boldsymbol{A}}_{i,j}^b = \frac{\boldsymbol{A}_{i,j}}{N} - \frac{\boldsymbol{A}_{i,j}}{N_c}$ and $\overline{\boldsymbol{A}}_{i,j}^w = \frac{\boldsymbol{A}_{i,j}}{N_c}$ if $\boldsymbol{x}_i, \boldsymbol{x}_j$ are in the same class, otherwise $\overline{\boldsymbol{A}}_{i,j}^b = \frac{1}{N}$ and $\overline{\boldsymbol{A}}_{i,j}^w = 0$, N_c is the number of samples in the corresponding class, and N is the

total number of samples in all classes. This aims to extract local data variation which has been proven to be useful [26]. For two image domains of different scales, we denote their inter-class and intra-class scatter matrices as $\boldsymbol{S}_b^h$ and $\boldsymbol{S}_w^h$ for the normal scale and $\boldsymbol{S}_b^s$ and $\boldsymbol{S}_w^s$ for the small scale.

Now, we aim to minimise $\text{HCMD}(\boldsymbol{W}_h, \boldsymbol{W}_s)$ whilst maximising both (1) the ratio between inter-class covariance and intra-class covariance for the normal scale images under projection $\boldsymbol{W}_h$ and (2) the ratio between inter-class covariance and intra-class covariance for the small scale images under projection $\boldsymbol{W}_s$. That is to simultaneously maximise the following three criteria:

$$\max_{\boldsymbol{W}_h, \boldsymbol{W}_s} \begin{cases} \text{HCMD}(\boldsymbol{W}_h, \boldsymbol{W}_s)^{-1}, \\ \frac{tr(\boldsymbol{W}_h^T \boldsymbol{S}_b^h \boldsymbol{W}_h)}{tr(\boldsymbol{W}_h^T \boldsymbol{S}_w^h \boldsymbol{W}_h)}, \\ \frac{tr(\boldsymbol{W}_s^T \boldsymbol{S}_b^s \boldsymbol{W}_s)}{tr(\boldsymbol{W}_s^T \boldsymbol{S}_w^s \boldsymbol{W}_s)}. \end{cases} \tag{3}$$

This cross-scale image domain alignment requires jointly achieving a discriminant optimisation in two image scales. However, it is nontrivial to simultaneously perform the above optimisation. To solve this problem, we consider instead a relaxed criterion that unifies all of them as follows:

$$\max_{\boldsymbol{W}_h, \boldsymbol{W}_s} \frac{\text{tr}(\boldsymbol{W}_h^T \boldsymbol{S}_b^h \boldsymbol{W}_h + \boldsymbol{W}_s^T \boldsymbol{S}_b^s \boldsymbol{W}_s)}{\text{tr}(\boldsymbol{W}_h^T \boldsymbol{S}_w^h \boldsymbol{W}_h + \boldsymbol{W}_s^T \boldsymbol{S}_w^s \boldsymbol{W}_s) + \alpha \text{HCMD}(\boldsymbol{W}_h, \boldsymbol{W}_s)} \tag{4}$$

where α is a parameter controlling the strength of HCMD.

For the joint learning on more than two scales, more domains pairs of different scales are used to model HCMD and more domain data are used to form $\boldsymbol{S}_b^s$ and $\boldsymbol{S}_w^s$.

We call the above model *joint multi-scale discriminant component analysis* (JUDEA). We will show that JUDEA model can be converted to a conventional eigenvalue decomposition problem, making it computationally tractable.

3.4. Optimisation

An intuitive way to optimise $\boldsymbol{W}_h$ and $\boldsymbol{W}_s$ in Eq. (4) is to learn each of them separately by fixing the other. It is, however, a computationally complex task. Fortunately, we show that it is possible to directly compute an optimal concatenated matrix $\boldsymbol{W} = [\boldsymbol{W}_h; \boldsymbol{W}_s]$. More specifically, we define $\boldsymbol{I}_d$ as the $d \times d$ identity matrix and $\boldsymbol{O}_{d \times m}$ as the $d \times m$ matrix of all zero. And let $\boldsymbol{\phi}_h = [\boldsymbol{I}_{d_h}, \boldsymbol{O}_{d_h \times d_s}]$, $\boldsymbol{\phi}_s = [\boldsymbol{O}_{d_s \times d_h}, \boldsymbol{I}_{d_s}]$. Hence $\boldsymbol{W}_h = \boldsymbol{\phi}_h \boldsymbol{W}$, $\boldsymbol{W}_s = \boldsymbol{\phi}_s \boldsymbol{W}$. Therefore, learning $\boldsymbol{W}_h$ and $\boldsymbol{W}_s$ is equal to learning $\boldsymbol{W}$:

$$\boldsymbol{W} = \arg\max \frac{\text{tr}(\boldsymbol{W}^T \boldsymbol{\Lambda}_b \boldsymbol{W})}{\text{tr}(\boldsymbol{W}^T \boldsymbol{\Lambda}_w \boldsymbol{W}) + \alpha \, \text{tr}(\boldsymbol{W}^T \boldsymbol{\Lambda}_{\text{HCMD}} \boldsymbol{W})}, \tag{5}$$

where

$$\boldsymbol{\Lambda}_b = \boldsymbol{\phi}_h^T \boldsymbol{S}_b^h \boldsymbol{\phi}_h + \boldsymbol{\phi}_s^T \boldsymbol{S}_b^s \boldsymbol{\phi}_s, \; \boldsymbol{\Lambda}_w = \boldsymbol{\phi}_h^T \boldsymbol{S}_w^h \boldsymbol{\phi}_h + \boldsymbol{\phi}_s^T \boldsymbol{S}_w^s \boldsymbol{\phi}_s,$$

$$\boldsymbol{\Lambda}_{\text{HCMD}} = \frac{1}{C} \sum_{i=1}^{C} (\boldsymbol{\phi}_h^T \boldsymbol{u}_i^h - \boldsymbol{\phi}_s^T \boldsymbol{u}_i^s)(\boldsymbol{\phi}_h^T \boldsymbol{u}_i^h - \boldsymbol{\phi}_s^T \boldsymbol{u}_i^s)^T.$$

Hence the optimization of Eq. (5) can be cast as a typical generalized eigenvalue problem:

$$\boldsymbol{\Lambda}_b \boldsymbol{W} = \lambda \boldsymbol{\Lambda} \boldsymbol{W}, \tag{6}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_w + \alpha \boldsymbol{\Lambda}_{\text{HCMD}}$. In this way, we can obtain the optimal $\boldsymbol{W}_h$ and $\boldsymbol{W}_s$ efficiently.

3.5. Matching Low-resolution Probe Images

The joint multi-scale learning framework is used for matching a LR probe image against a set of normal resolution or HR gallery images. Similar to the training process of JUDEA, for a LR probe image $\boldsymbol{x}_p$, we obtain two images by scaling the input to a small scale image $\boldsymbol{x}_p^s$ and scaling it to a normal scale image $\boldsymbol{x}_p^h$, where the normal scale and small scale conform to the training setting. Similarly, for each images $\boldsymbol{x}_g$ in the gallery , normal scale image $\boldsymbol{x}_g^h$ and small scale image $\boldsymbol{x}_g^s$ are also obtained in this way. We combine the two different scale distances as:

$$d(\boldsymbol{x}_p, \boldsymbol{x}_g) = \beta ||\boldsymbol{W}_h^T \boldsymbol{x}_p^h - \boldsymbol{W}_h^T \boldsymbol{x}_g^h||_2 + (1-\beta) ||\boldsymbol{W}_s^T \boldsymbol{x}_p^s - \boldsymbol{W}_s^T \boldsymbol{x}_g^s||_2 \tag{7}$$

where $\boldsymbol{W}_h$ and $\boldsymbol{W}_s$ are the optimal projection matrices for different scales (Sec. 2.3), and β is the weight for regulating the effects of normal and small scale distance. This fusion matching strategy further exploits the information from different scales in multi-scale framework.

4. Experiments

4.1. Datasets and Settings

Datasets. The CAVIAR dataset [5] is widely used for evaluating person re-id, containing images of 72 individuals captured from 2 cameras in a shopping mall. This dataset is suitable for testing LR person re-id, as the resolution of images captured from the second camera is much lower than that in the first camera (Fig. 1 bottom). Among the 72 people, 22 were only captured in a single camera view with no low resolution images, and they were thus removed. The remaining data were used in our experiments which include 1000 images of 50 people, with 10 normal resolution images and 10 LR images per person (see Fig. 3).

Two simulated LR person datasets LR-VIPeR and LR-3DPES were also used for evaluation. These are based on VIPeR [8] and 3DPES [1] respectively. The VIPeR dataset consists of 632 people captured outdoor with two images for each person. The 3DPES dataset includes 1011 images of 192 individuals captured from 8 outdoor cameras with significantly different viewpoints. In this dataset each person has 2 to 26 images. In order to make the two datasets suitable for evaluating the person re-identification in low resolution, we randomly selected half of all the images of each person from both datasets, and replaced them with the LR images which were sub-sampled to a quarter of their original image size. Examples of the generated LR-VIPeR dataset and LR-3DPES dataset are shown in Fig. 3.

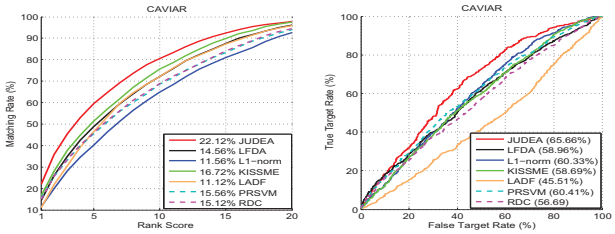
(a) CAVIAR (b) LR-VIPeR (c) LR-3DPES

Figure 3. Examples of the normal person images and the corresponding LR images on three LR datasets.

Settings. In our experiments, we adopted a single-shot experiment setting. All datasets were randomly divided into training set and testing set by half so that there are $p = 25$, $p = 316$ and $p = 96$ individuals in the testing set of CAVIAR, LR-VIPeR and LR-3DPES respectively. The probe set consists of all LR images per person in the testing set. One normal resolution image for each individual in the testing set was randomly selected to construct the gallery set. This procedure was repeated 10 times. For evaluation, we used the average cumulative match characteristic (CMC) curves to show the ranked matching rates.

The setting described above is the conventional closed-set setting, i.e. the gallery and probe sets contain exactly the same set of people. In a real-world application scenario, an open-set setting could be more appropriate, under which there is no one-to-one correspondence between the people appeared in the gallery and probe sets. For evaluation under the open-set setting, images of 50% of the gallery people were randomly removed and the probe set remain the same as the closed-set. For this setting, we used ROC curves instead of the CMC curves as the evaluation metric.

Compared Methods. To evaluate the proposed model, we compared it against a total of twelve different existing related models. We first compared it with six existing re-id methods, including a representative subspace learning method LFDA [26], a non-learning distance metric L1-norm, two discriminant distance learning methods KISSME [11] and LADF [17], and two ranking models PRSVM [27] and RDC [41]. Since none of the existing re-id approaches explicitly addresses the LR person re-id problem as we do, we further compared six other relevant methods. Specifically, five cross-domain learning methods were compared, including domain adaptation based on manifold alignment (DAMA) [32], canonical correlation analysis (CCA) [10], coupled marginal fisher analysis (CMFA) [29], maximum-margin coupled mappings (MMCM) [30] and the DTRSVM in [20]. Among the five methods mentioned above, DAMA is for general purpose, DTRSVM is for domain adaptation in person re-identification but not for LR image processing, and CCA, CMFA and MMCM are representative methods for LR face image matching. In addition, a popular super-resolution method based on sparse representation (SPARSE-SR) [36] was also compared.



(a) CMC (b) ROC

Figure 4. Comparison with related re-id methods in conventional setting on CAVIAR: CMC curves with rank-1 matching rate, and ROC curves with area-under-curve(AUC) values.

Feature Representation. The uniform normal scale and small scale were set to 128×48 and 64×24 respectively in our experiments (i.e. 1:4 scale ratio). We used appearance representations of pedestrians captured by a set of different basic features which are a mixture of color, LBP and HOG features. Specifically, we obtained overlapping patches of size 16×16 from each person image, defined with every 8 pixels in both the horizontal and vertical directions. We then extracted features in each patch and finally concatenated them to form the final feature of the image. The patch feature vectors were made of 16-bins histogram of 8 color channels (RGB, YCbCr, HS). To incorporate the texture patterns and shape information, uniform LBP histograms and HOG descriptors were also computed for each image patch. So each patch was represented by a 484-dimensional feature vector. For each image normalised to 128×48 and 64×24 pixels, a total of 75 and 14 patches were extracted respectively, forming 36300-dimensional and 6776-dimensional feature vectors for the two scales respectively.

4.2. Evaluation on the CAVIAR LR Dataset

4.2.1 Comparison with Existing RE-ID Methods

For matching LR images to a normal resolution image, existing re-id methods scale the LR image upto the normal scale. The results of JUDEA compared with 6 existing re-id methods are shown in Fig. 4 (a). The following observation can be made: (1) Compared to the subspace-based method LFDA, JUDEA outperforms LFDA notably, with JUDEA achieving 7% improvement over LFDA at rank-1. (2) Compared to the metric/ranking learning methods, it is evident that the proposed JUDEA improves matching significantly over all of them including KISSME, LADF, PRSVM and RDC. More specifically, the rank-1 matching rate is 22.12% for JUDEA, 16.72% for KISSME, 11.12% for LADF, 15.56% for PRSVM, and 15.12% for RDC. These results show that the existing re-id methods perform poorly for LR re-id. In particular, as the appearance details have been lost in LR images, these methods are not designed for coping with such loss in imagery information.

(a)

Methods	r =1	r =5	r =10	r =20
JUDEA	**22.12**	**59.56**	**80.48**	97.84
LFDA_F	17.68	53.76	76.60	97.36
L1-norm_F	12.40	43.44	67.88	94.36
KISSME_F	18.92	55.08	78.16	**98.00**
LADF_F	15.88	51.80	75.60	96.68
PRSVM_F	17.00	47.00	69.48	94.48
RDC_F	17.60	48.84	71.96	95.40

(b)

Methods	r =1	r =5	r =10	r =20
JUDEA	**22.12**	**59.56**	**80.48**	**97.84**
LFDA_C	16.40	50.08	72.52	96.56
L1-norm_C	11.84	41.52	65.72	93.20
KISSME_C	17.72	53.20	77.24	97.72
LADF_C	11.68	48.84	74.36	96.56
PRSVM_C	16.20	46.60	68.64	94.80
RDC_C	15.84	45.60	67.04	94.52

(c)

Methods	r =1	r =5	r =10	r =20
JUDEA	**22.12**	**59.56**	**80.48**	**97.84**
LFDA_H	14.00	47.56	72.28	96.72
L1-norm_H	10.40	38.32	64.44	92.92
KISSME_H	15.76	51.08	74.72	97.32
LADF_H	10.24	42.08	68.44	95.56
PRSVM_H	14.36	45.24	67.44	94.44
RDC_H	13.44	45.68	67.32	94.76

Table 1. Matching Rate (%): JUDEA vs. re-id methods under multi-scale settings on CAVIAR. "_F" indicates learning at two scales and combining, "_C" indicates learning on concatenated features of two scales, and "_H" indicates extracting features of the same dimension.

Methods	r =1	r =5	r =10	r =20
JUDEA	**22.12**	**59.56**	**80.48**	**97.84**
DAMA	19.08	52.68	76.04	97.52
CCA	12.12	40.52	62.40	92.12
CMFA	13.28	43.36	66.76	94.44
MMCM	15.24	46.64	68.84	95.76
DTRSVM	16.81	48.47	71.22	94.40
SPARSE-SR	15.12	49.36	72.84	96.60

Table 2. Matching Rate (%): JUDEA vs. others on CAVIAR.

4.2.2 Comparison under Multi-scale Settings

Since our JUDEA is learned using images of two scales in the experiments, whilst the existing re-id methods learn their models at a single scale. For a fair comparison, we now learn the six re-id models at two scales.

Learning at two image scales and combining. We re-scaled each image to both a normal scale image and a small scale image, learned them independently, and finally fused the distances when matching. The existing re-id methods learned in this way are denoted by adding a suffix "_F" after their names, e.g. LFDA_F. The results are shown in Table 1 (a). Compared to Fig. 4 (a), it is evident that all the methods performed better. This suggests that learning at separate scales is important for LR person re-id. However, it is also evident that our model still yields overall much better performance than other methods, especially at lower rank. This suggests even though existing methods were applied on two scales separately, they are still sub-optimal solutions without joint learning at different scales.

Learning on concatenated features of two image scales. We re-scaled each image to a normal scale image and a small scale image, extracted features from each image, concatenated the features of a pair of normal scale images and their small scale counterparts, and then tested the existing re-id methods on the concatenated features. In this case, we denote the existing re-id methods by adding a suffix "_C" after their names, e.g. LFDA_C. Although the performance of these methods (Table 1 (b)) has gained a slight improvement (compared to Fig. 4 (a)), this feature concatenation approach is less effective than our model.

Extracting features of the same dimension. Instead of re-scaling images, for each LR image, we directly extracted features in the same dimension as that of the normal resolution images by densely sampling as much as possible from the LR one. Then, existing re-id methods can be applied directly. In this experiment, we use the suffix "_H" after each name of existing re-id methods to denote this variation. From Table 1 (c), it is evident that JUDEA outperforms other methods. Compared to Fig. 4 (a), the performances of the existing methods are poorer. This suggests it is not a sensible solution by forcefully extracting equal amount of information in the LR images as in the normal resolution images, when the information has already been lost.

4.2.3 Comparison with Cross-domain Methods

Five cross-domain learning methods (DAMA [32], CCA [10], CMFA [29], MMCM [30] and DTRSVM [20]), which can cope with different scale domains, were also applied to LR re-id. However, none of them was designed for L-R re-id. In this context, DAMA can be considered as a joint learning model across different scales. In comparison, our JUDEA learns a locally discriminant metric so as to identify the appearance change locally. The advantage of JUDEA is validated by the experimental results shown in Table 2: at rank-1, JUDEA achieves approx. 3% performance advantage over DAMA; as rank increases, more improvements are observed. In addition, compared to the C-CA, CMFA and MMCM, the three coupled transformation methods used for LR face recognition, JUDEA also outperforms them significantly. This is because the assumption made by these methods on directly aligning low and high resolution face images is not applicable for LR person re-id problem. We also implemented DTRSVM for solving our multi-scale based LR re-id problem by treating different scales as different domains. Since DTRSVM requires the dimensions of all data must be the same, we have to extract the features of the same dimension from LR images as we do for the normal resolution images. As shown, DTRSVM is 5 and 11 matching rates lower than our method at rank 1 and rank 5 on CAVIAR, respectively. The inferior results show that domain adaptation by DTRSVM is not optimal for solving our multi-scale based LR re-id problem.

4.2.4 Comparison with Super-resolution Method

We also utilised a super-resolution method based on sparse representation (SPARSE-SR [36]) to generate normal reso-

Methods	Mixed resolution probe set				Normal resolution probe set			
	r =1	r =5	r =10	r =20	r =1	r =5	r =10	r =20
JUDEA	**45.35**	**74.29**	**88.17**	**98.82**	**71.29**	**90.84**	96.93	**99.96**
LFDA	39.33	66.86	83.05	97.77	67.96	89.47	96.62	99.82
L1-norm	36.46	60.78	77.60	95.45	64.13	83.78	91.69	98.62
KISSME	36.99	67.47	84.32	98.57	59.51	85.47	94.04	99.73
LADF	26.95	62.08	81.73	97.79	44.53	79.38	92.49	99.78
PRSVM	39.81	65.73	81.26	96.63	66.76	88.13	95.33	99.38
RDC	39.92	65.54	81.01	96.91	67.47	87.33	94.44	99.51
DAMA	40.72	69.47	84.95	98.53	64.76	88.13	94.84	99.64
CCA	33.83	60.06	76.04	95.54	57.96	81.78	91.20	99.33
CMFA	35.62	64.54	79.94	96.76	60.44	88.07	94.58	99.34
MMCM	36.71	64.82	80.42	96.90	60.57	85.02	93.29	98.17
DTRSVM	40.21	65.50	81.08	96.23	66.22	84.43	92.04	98.24
SPARSE-SR	39.94	68.80	84.36	98.19	67.51	90.40	**97.16**	99.96

Table 3. Matching Rate (%): JUDEA vs. related methods on CAVIAR with different probe sets. Mixed resolution probe set indicates probe set is composed of LR images and normal resolution images. Normal resolution probe set indicates probe set is composed of all normal resolution images.

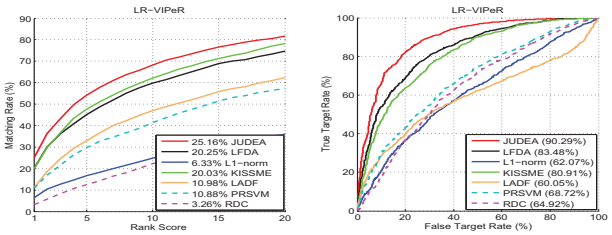
lution images from LR images and then applied LFDA for the re-id matching task. Table 2 shows that the performance of SPARSE-SR is 6% lower than that of the JUDEA at rank 1. Although SPARSE-SR is a popular method for image super-resolution, the results show that it does not solve the LR person re-id problem well. One reason is that it requires accurate and dense alignment across scales which is not available for person full body images. The other reason is that since there are only limited samples for training and each person's appearance varies significantly in the re-id datasets, most super-resolution methods tend to over-fit the training data and generalise poorly to the test data.

4.2.5 Effects of Probe Sets of Different Resolutions

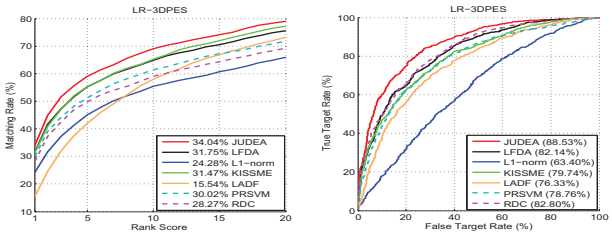
In a realistic re-id situation, a probe set may include both LR images and normal (higher) resolution images. To validate the effectiveness of our model for this situation experimentally, we kept similar percentages of LR images and normal resolution images in the probe set. The results in Table 3 show that the JUDEA model still outperforms other methods under this setting. In order to further verify the robustness of JUDEA, we considered an extreme case for which the probe set only has normal resolution images. As shown in Table 3, when the probe set has no LR images, our model can still obtain the best performance compared to the other methods, although as expected, the gap is smaller. This result shows that our model is competitive even when the LR person re-id problem does not exist.

4.3. Evaluation on Simulated LR Datasets

Our experiments were also conducted on two simulated LR datasets LR-VIPeR and LR-3DPES. Our results (Fig. 5 and Table 4) show clearly that JUDEA outperforms other methods on both datasets. The advantage is particularly significant on the LR-VIPeR, with JUDEA is 5% higher than the best of all the related methods at rank-1.



(a) CMC: LR-VIPeR (b) ROC: LR-VIPeR



(c) CMC: LR-3DPES (d) ROC: LR-3DPES

Figure 5. Comparison with related re-id methods in conventional setting on LR-VIPeR and LR-3DPES

Methods	LR-VIPeR				LR-3DPES			
	r =1	r =5	r =10	r =20	r =1	r =5	r =10	r =20
JUDEA	**25.16**	**54.27**	**68.10**	**81.71**	**34.04**	**58.99**	**69.07**	**78.98**
DAMA	18.01	43.01	56.80	72.88	30.24	52.47	61.33	73.02
CCA	9.37	24.68	35.09	47.88	26.22	45.91	54.89	64.36
CMFA	13.29	31.65	46.32	57.18	26.73	49.64	58.14	69.53
MMCM	14.74	34.43	49.03	62.85	28.43	51.54	61.28	70.86
DTRSVM	12.26	36.43	48.87	64.52	31.75	54.28	64.69	73.23
SPARSE-SR	20.70	45.76	58.73	73.99	32.83	55.62	66.45	76.52

Table 4. Matching Rate (%): JUDEA vs. other related methods on LR-VIPeR and LR-3DPES.

Methods	LR-VIPeR				LR-3DPES			
	r =1	r =5	r =10	r =20	r =1	r =5	r =10	r =20
JUDEA	**25.16**	**54.27**	**68.10**	**81.71**	**34.04**	**58.99**	**69.07**	**78.98**
LFDA_F	21.61	48.89	63.70	76.46	32.89	55.74	65.40	75.90
L1-norm_F	7.88	19.49	28.39	39.27	28.21	49.57	58.06	69.90
KISSME_F	21.84	50.09	65.98	79.05	32.35	56.08	65.88	77.60
LADF_F	11.33	33.89	47.44	63.80	16.06	42.72	58.81	73.33
PRSVM_F	11.36	30.32	42.82	57.66	33.03	53.22	62.50	71.57
RDC_F	10.60	29.40	41.08	56.49	31.31	53.30	62.51	72.01

Table 5. Matching Rate (%): JUDEA vs. re-id methods with fusion matching across two scales on LR-VIPeR and LR-3DPES.

Similarly, we have compared JUDEA with the six re-id methods under the "Multi-scale Settings" as what have been done in Sec. 4.2.2. Due to space limit, we only show the comparison results under the setting "Learning at two image scales and combining". The results in Table 5 show that our model also performs better than other re-id methods. Similar conclusions can be drawn for the other two settings.

4.4. Further Analysis

Contributions of HCMD in JUDEA. The HCMD criterion minimises the intra-class distribution differences of images of the same individual on different scale image domains. This reduces data redundancy and increases the availabili-

Methods	CAVIAR				LR-VIPeR				LR-3DPES			
	r =1	r =5	r =10	r =20	r =1	r =5	r =10	r =20	r =1	r =5	r =10	r =20
JUDEA	**22.12**	**59.56**	**80.48**	**97.84**	**25.16**	**54.27**	**68.10**	**81.71**	**34.04**	**58.99**	**69.07**	**78.98**
JUDEA-w/o	20.24	56.56	78.52	97.28	21.80	49.59	63.67	77.56	33.29	57.69	67.76	78.78
JUDEA$_{normal}$	19.32	54.64	76.32	96.40	23.92	52.18	67.41	81.17	32.50	56.77	67.53	78.27
JUDEA$_{small}$	19.76	56.64	78.08	97.76	18.57	46.01	60.89	75.89	30.76	56.01	66.84	77.89
JUDEA$_{three}$	**22.40**	**61.04**	**81.80**	**98.48**	**25.87**	**55.02**	**68.53**	**82.13**	**34.72**	**59.36**	**70.21**	**79.87**

Table 6. Matching Rate (%): Further Analysis of JUDEA on CAVIAR, LR-VIPeR, and LR-3DPES.

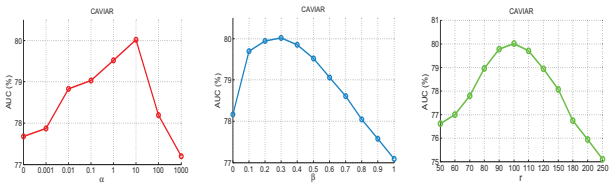
ty of details for LR images. In Table 6, JUDEA-w/o denotes JUDEA without HCMD. These results show clearly that JUDEA consistently outperforms JUDEA-w/o. The improvement is particularly significant on the LR-VIPeR.

Fusion matching vs. single matching in JUDEA. We also evaluated fusion matching in JUDEA. The fusion matching criterion is designed to improve the performance by combining similarity scores of different scales. We compared fusion matching with single matching which is adopted on the normal scale and the small scale, denoted as JUDEA$_{normal}$ and JUDEA$_{small}$ respectively. Table 6 shows that JUDEA with fusion matching obtains better results, even though JUDEA with single matching on traditional normal scale already gives a notable improvement over existing re-id methods as shown in Figs. 4 and 5.

Using more than two scales in JUDEA. The proposed JUDEA model uses two scales to achieve joint multi-scale learning in all previous experiments. One may wonder whether using more than two scales helps. To answer the question, we designed an even smaller scale (32×16) in addition to the existing two scales, resulting in a joint learning across three scales in JUDEA, which we call JUDEA$_{three}$. We performed experiments on three datasets and the results are reported in Table 6. It shows that the performance of JUDEA$_{three}$ has a slight improvement. This suggests that the benefit from using more scales is limited for LR person matching, with the added computational costs.

Effects of parameters. We implemented the JUDEA by selecting parameter $\alpha = 10$ and r=100 on all datasets, and β is set to 0.7 and 0.3 for LR-VIPeR and other datasets respectively. We varied the three parameters to evaluate JUDEA. Due to space limit, we only show results on the CAVIAR dataset here. Similar conclusions can be drawn from the results on the two datasets. We varied the value of one parameter whilst fixing the other. The AUC (Area under CMC curve) of α, β and r are plotted in Figs. 6 (a), (b) and (c), respectively. It can be seen that when α is around 10, β is around 0.3 and r is around 100, the model achieves the best result. But overall their effects are small.

Open-set testing. Due to limited space, we only can report the comparative results under the open-set testing in Figs. 4 (b), 5 (b) and 5 (d). It is also evident that the proposed model outperforms others under the open-set setting.



(a) Performances: α. (b) Performances: β. (c) Performances: r.

Figure 6. AUC of JUDEA with different parameters on CAVIAR.

4.5. Discussions

The key findings of the experiments are:

1) Existing re-id methods are not specifically designed for LR person re-id problem, and thus their performances degrade significantly as shown in Sec. 4.2.1 & 4.3.

2) Even when we modify the existing re-id methods to learn models at different scales in Sec. 4.2.2, there is still a clear margin between the performances of our method and theirs, showing the importance of joint multi-scale learning.

3) Compared to related cross-domain LR face recognition methods in Sec. 4.2.3 and 4.3, the proposed JUDEA does not explicitly align a pair of low and normal resolution images, but simultaneously learns discriminant metrics of different scales constrained by HCMD. The results suggest our strategy is more suitable for LR person re-id.

5. Conclusion

To address the low resolution (LR) person re-id problem, we proposed a **j**oint m**u**lti-scale **d**iscriminant compon**e**nt **a**nalysis (JUDEA) model by learning a shared subspace across different scales, which is the first specific work on solving such a challenge to our best knowledge. Extensive experiments were conducted to evaluate and compare the proposed model on three different LR person re-id datasets.

A number of conclusions can be drawn from the results. First, a multi-scale discriminant modelling unified with the proposed heterogeneous class mean discrepancy (HCMD) criterion for simultaneously learning metrics on image domains of different scales is more effective than single-scale based modelling followed by simple combination of the models. Second, LR is indeed a challenge to re-id. Although there exists LR recognition techniques in face recognition, we show that these techniques do not work well on the harder LR person re-id problem.

Acknowledgments

This work was supported partially by the National NSFC (Nos. 61472456, 61573387), NSFC for Excellent Young Scientist Programme (No. 61522115), Guangzhou Pearl River Science and Technology Rising Star Project (No. 2013J2200068), and in part by the Guangdong Natural Science Funds for Distinguished Young Scholar (No. S2013050014265).

References

[1] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *ACM workshop on Human gesture and behavior understanding*, 2011.

[2] S. Biswas, K. Bowyer, and P. Flynn. Multidimensional scaling for matching lowresolution face images. *IEEE TPAMI*, 34(10):2019–2030, 2012.

[3] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):49–57, 2006.

[4] A. Chakrabarti, A. Rajagopalan, and R. Chellappa. Super-resolution of face images using kernel pca-based prior. *IEEE TMM*, 9(4):888–892, 2007.

[5] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011.

[6] P. Dollár, Z. Tu, H. Tao, and S. Belongie. Feature mining for image classification. In *CVPR*, 2007.

[7] M. Farenzena, L. Bazzani, A. Perina, M. Cristani, and V. Murino. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[8] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 2008.

[9] P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *CVPR*, 2008.

[10] T.-K. Kim, J. Kittler, and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE TPAMI*, 29(6):1005–1018, 2007.

[11] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[12] B. Kulis, K. Saenko, and T. Darrell. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *CVPR*, 2011.

[13] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *IEEE TPAMI*, 35(7), 2013.

[14] B. Li, H. Chang, S. Shan, and X. Chen. Low-resolution face recognition via coupled locality preserving mappings. *IEEE Signal Processing Letters*, 17(1):20–23, 2010.

[15] W. Li, L. Duan, D. Xu, and I. Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE TPAMI*, 36(6):1134–1148, 2014.

[16] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.

[17] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013.

[18] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: what features are important? In *ECCV Workshop*, 2012.

[19] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *IJCV*, 75(1):115–134, 2007.

[20] A. J. Ma, P. C. Yuen, and J. Li. Domain transfer support vector ranking for person re-identification without target camera label information. In *ICCV*, 2013.

[21] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *ECCV Workshop*. Springer, 2012.

[22] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *PR*, 43(6):2224–2236, 2010.

[23] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*, 2012.

[24] P. Moutafis and I. A. Kakadiaris. Semi-coupled basis and distance metric learning for cross-domain matching: Application to low-resolution face recognition. In *IJCB*, 2014.

[25] U. Park, A. K. Jain, I. Kitahara, K. Kogure, and N. Hagita. Vise: Visual search engine using multiple networked cameras. In *ICPR*, volume 3, 2006.

[26] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, 2013.

[27] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, 2010.

[28] X. Shi, Q. Liu, W. Fan, P. S. Yu, and R. Zhu. Transfer learning on heterogenous feature spaces via spectral transformation. In *ICDM*, 2010.

[29] S. Siena, V. Boddeti, and B. Kumar. Coupled marginal fisher analysis for lowresolution face recognition. In *ECCV Workshops and Demonstrations*, 2012.

[30] S. Siena, V. Boddeti, and B. Kumar. Maximum-margin coupled mappings for cross-domain matching. In *BTAS*, 2013.

[31] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li. Person re-identification by regularized smoothing kiss metric learning. *IEEE TCSVT*, 23(10):1675–1685, 2013.

[32] C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *IJCAI*, 2011.

[33] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *ICCV*, 2007.

[34] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE TSMC-C*, 35(3):425–434, 2005.

[35] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*, 2014.

[36] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *TIP*, 19(11), 2010.

[37] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *NIPS*, 2004.

[38] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013.

[39] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.

[40] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014.

[41] W.-S. Zheng, S. Gong, and T. Xiang. Re-identification by relative distance comparison. *IEEE TPAMI*, 35(3).

[42] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *BMVC*, 2009.

[43] C. Zhou, Z. Zhang, D. Yi, Z. Lei, and S. Li. Low-resolution face recognition via simultaneous discriminant analysis. In *IJCB*, 2011.