Mirror Representation for Modeling View-Specific Transform in Person Re-Identification

Ying-Cong Chen[†], **Wei-Shi Zheng**[†,§,*], **Jianhuang Lai**[†,‡]
[†] School of Information Science and Technology, Sun Yat-sen University, China
[§] Guangdong Provincial Key Laboratory of Computational Science
[‡] Guandong Key Laboratory of Information Security Technology
chyingc@mail2.sysu.edu.cn, wszheng@ieee.org, stsljh@mail.sysu.edu.cn

Abstract

Person re-identification concerns the matching of pedestrians across disjoint camera views. Due to the changes of viewpoints, lighting conditions and camera features, images of the same person from different views always appear differently, and thus feature representations across disjoint camera views of the same person follow different distributions. In this work, we propose an effective, low cost and easy-to-apply schema called the Mirror Representation, which embeds the view-specific feature transformation and enables alignment of the feature distributions across disjoint views for the same person. The proposed Mirror Representation is also designed to explicitly model the relation between different view-specific transformations and meanwhile control their discrepancy. With our Mirror Representation, we can enhance existing subspace/metric learning models significantly, and we particularly show that kernel marginal fisher analysis significantly outperforms the current state-of-the-art methods through extensive experiments on VIPeR, PRID450S and CUHK01.

1 Introduction

In recent years, the close-circuit television (CCTV) has been widely deployed in public area such as hospitals, railway stations, office buildings, etc. Because of economical or privacy issue, there are always non-overlapping regions between camera views. Inevitably, re-identifying a pedestrian who disappears from a camera view and re-appears in another disjoint camera view is indispensable for some cross-camera analysis like tracking or activity prediction, and such a problem is called the *person re-identification*.

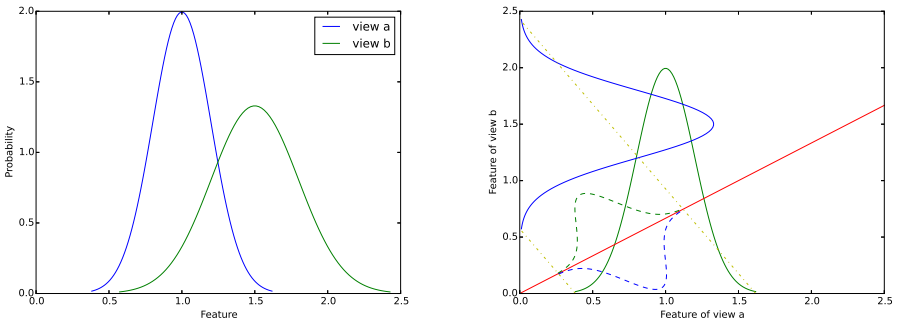
The non-overlapping cross-view re-identification differs conventional recognition in the same view due to the great changes of illumination, viewpoint or camera features, so that appearance of a pedestrian changes dramatically across camera views. Such environmental condition inconsistency results in view-specific feature distortion, i.e., the pedestrian's appearance itself from view a and view b is transformed to $T^a(x)$ and $T^b(x)$ respectively, where T^a and T^b are unknown distortion functions. Consequently, the features distributions of the two views are not well-aligned. Since T^a and T^b are usually of complex nonlinearity, it is the fact that $P(T^a(x)) \neq P(T^b(x))$ and $P(y|T^a(x)) \neq P(y|T^b(x))$.

In the field of person re-identification, a lot of methods have been developed including descriptor-based methods[Zhao *et al.*, 2014; Yang *et al.*, 2014; Kviatkovsky *et al.*, 2013] and subspace/metric learning methods[Weinberger *et al.*, 2006; Mignon and Jurie, 2012; Davis *et al.*, 2007; Zheng *et al.*, 2013; Li *et al.*, 2013; Pedagadi *et al.*, 2013; Xiong *et al.*, 2014]. Descriptor-based methods seek for more reliable descriptors across views. However, it is extremely difficult to extract robust descriptors against large appearance changes. Subspace learning seeks for discriminative combinations of features and enhances the performance, and metric learning aims to learn a variation insensitive and discriminant similarity measure between samples. However, these approaches all have an underlying assumption that all samples are drawn from the same distribution, which becomes invalid for matching person across *disjoint* camera views.

Previous approaches assume there is a uniform transformation to alleviate the feature distortions in any camera view. In this work, we relax this assumption by learning view-specific transform for each camera view. To derive our approach, we start from an intuitive feature transformation strategy called *zero-padding*, which augments the original features with zeros in view-specific entries. In this way, the view-specific knowledge is embedded in the augmented features and traditional subspace/metric learning can learn view-specific feature combinations. This idea is similar to [Daumé III, 2009], which aims at domain adaptation. However, as shown in Sec. 2.1, zero-padding may extract irrelevant feature transformations of different camera views and the discrepancy between view-specific feature transformations is not explicitly controlled. This ignores the fact that albeit captured from disjoint camera views, the appearances of two cross-view person images are still relevant.

To control the discrepancy between view-specific transformations explicitly, we first generalize the zero-padding strategy and introduce a novel augmented representation of a pedestrian's feature representation for a pair of camera views. This achieves the discrepancy modeling on feature-level. Secondly, we further introduce a transformation-level

[*]Corresponding Author



(a) Mismatched Distribution (b) Zero-padding

Figure 1: Illustration of the zero-padding augmentation. For visualization purpose, we only analyze one dimensional feature. (a) visualizes the distorted feature distributions of view a and view b. (b) shows the zero-padded features (the solid blue curve and the green curve) of both views and the aligned subspace (solid red line). The projected features are in dashed blue curve and dash green curve. We show that with *zero-padding* it is possible to find a subspace that align the distorted features. Note that in this example the distortion functions are assume to be linear, and for the non-linear case, kernel trick could be employed for non-linear mappings. (Best viewed in color)

discrepancy by explicitly measuring the difference between the transforms of two disjoint views. We call this the discrepancy regularization. By transforming the augmented feature to a new space, this regularization is involved in the new representation. We call this transformed augmented representation the Mirror Representation, since it is generated by both the original feature and the phantom feature. With the mirror representation, we demonstrate that existing metric learning methods can be greatly enhanced on person re-identification across two disjoint camera views.

Our work is related to domain adaptation in machine learning. Some domain adaptation methods like [Pan *et al.*, 2011; Si *et al.*, 2010; Geng *et al.*, 2011; Xiao and Guo, 2014; Li *et al.*, 2014] aim at diminishing the gap between source and target domains. However, these domain adaptation methods assume existence of the overlap between the classes of training set and testing set so that a classifier/metric learned from training set can be adapted to the testing one, while for person re-identification the training and testing classes are non-overlapping. Our work is also different from the transfer re-identification methods [Zheng *et al.*, 2012; Ma *et al.*, 2013]. We do not address the problem of incorporating any other source data for transferring knowledge to target task. Alternatively, we address a very different problem against these re-identification work by alleviating the mismatch of the feature distributions of different camera views.

In summary, our approach is efficient but not sophisticated, and it is low cost and easy-to-use. Our contributions are listed as follows:

- We propose a mirror representation to address the view-specific feature distortion problem in person re-identification, and then further extend it to the kernel version;
- Based on the mirror representation, we are able to significantly enhance existing metric learning/subspace models, and we demonstrate that such a generalization obtains significant improvement on three popular datasets.

Figure 2: Illustration of the feature distortion. Images of the two rows are from different camera views on VIPeR dataset. Note that both environment-specific distortion such as illumination conditions and person-specific distortions such as pose and backpacks exist.

2 Approach

In this section, we propose two types of discrepancy modeling to form the mirror representation. Before introducing our approach, we first present the analysis on the zero-padding which motivates our modeling.

2.1 Zero-Padding Augmentation

To deal with the view-specific distribution mismatch problem, learning view-specific mapping is desirable for a subspace model. An intuitive idea is to apply *zero-padding* for each view, i.e., transforming $\boldsymbol{X}^a$ to $[\boldsymbol{X}^a; \boldsymbol{0}]$ and $\boldsymbol{X}^b$ to $[\boldsymbol{0}; \boldsymbol{X}^b]$, where $\boldsymbol{0}$ denotes zero matrix whose size is $d \times n^a$ or $d \times n^b$. Applying subspace/metric learning methods (like MFA [Yan *et al.*, 2007]) over such augmented features results in a $2d \times c$ projection matrix $\boldsymbol{U} = [\boldsymbol{U}_1; \boldsymbol{U}_2]$. Projecting the augmented features into $\boldsymbol{U}$ results in $\boldsymbol{Y}^a = \boldsymbol{U}_1^T \boldsymbol{X}^a$ and $\boldsymbol{Y}^b = \boldsymbol{U}_2^T \boldsymbol{X}^b$. Since $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are different, they are free to

adjust according to the feature distortion. The zero-padding strategy can be visualized in Figure 1. Note that in this example we use the linearity of distortion functions for demonstration; for more complex functions, non-linear mapping is needed, which can be achieved by the kernel trick.

However, by using the zero-padding, one loses direct control of the relation between $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ on data of both views, since $\boldsymbol{U}_1$ is always on $\boldsymbol{X}^a$ and $\boldsymbol{U}_2$ is always on $\boldsymbol{X}^b$ as shown above. In person re-identification, the view-specific distortion consists of two types, namely the environment-specific and person-specific feature distortions (see Figure 2). The environment-specific feature distortion concerns environment change such as change of lighting, view point and camera features, and the pedestrian-specific feature distortion concerns the appearance's change of a person himself/herself. No matter which type of distortion, there is relation between the same distortions across camera views. Without explicitly embedding the relation between $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, the extracted view-specific transforms of two disjoint views may be irrelevant, and thus an inferior re-identification performance will be gained as shown in our experiment.

2.2 A Feature-level Discrepancy Modeling

Formally, the zero-padding augmentation could be rewritten as:

$$\boldsymbol{X}^a_{aug} = [\boldsymbol{I}, \boldsymbol{0}]^T \boldsymbol{X}^a, \quad \boldsymbol{X}^b_{aug} = [\boldsymbol{0}, \boldsymbol{I}]^T \boldsymbol{X}^b \tag{1}$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{0}$ is a zero matrix.

Note that the feature transformation matrices $[\boldsymbol{I}, \boldsymbol{0}]$ and $[\boldsymbol{0}, \boldsymbol{I}]$ are totally *orthogonal*, since $[\boldsymbol{I}, \boldsymbol{0}]^T[\boldsymbol{0}, \boldsymbol{I}] = \boldsymbol{0}$, and thus the mapping functions are irrelevant. To impose restriction on the discrepancy of two mapping functions, we generalize (1) to:

$$\boldsymbol{X}^a_{aug} = [\boldsymbol{R}, \boldsymbol{M}]^T \boldsymbol{X}^a, \quad \boldsymbol{X}^b_{aug} = [\boldsymbol{M}, \boldsymbol{R}]^T \boldsymbol{X}^b \tag{2}$$

Thus the mapping functions are also generalized to:

$$\begin{aligned} f_a(\boldsymbol{X}^a) &= \boldsymbol{U}^T \boldsymbol{X}^a_{aug} = (\boldsymbol{R}^T \boldsymbol{U}_1^T + \boldsymbol{M}^T \boldsymbol{U}_2^T) \boldsymbol{X}^a \\ f_b(\boldsymbol{X}^b) &= \boldsymbol{U}^T \boldsymbol{X}^b_{aug} = (\boldsymbol{M}^T \boldsymbol{U}_1^T + \boldsymbol{R}^T \boldsymbol{U}_2^T) \boldsymbol{X}^b \end{aligned} \tag{3}$$

Note that in this generalized version, the discrepancy of mapping functions is determined by either the discrepancy $\boldsymbol{R}$ and $\boldsymbol{M}$ or $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$. Note that the feature transformation matrices $[\boldsymbol{R}, \boldsymbol{M}]$ and $[\boldsymbol{M}, \boldsymbol{R}]$ do not need to be orthogonal. We can define $\boldsymbol{R}$ and $\boldsymbol{M}$ so that the discrepancy between $f_a(\cdot)$ and $f_b(\cdot)$ can be measured. In this way, the discrepancy of mapping functions is explicitly controlled/parameterized on the feature-level. Let r be the parameter that controls the discrepancy of $[\boldsymbol{R}, \boldsymbol{M}]$ and $[\boldsymbol{M}, \boldsymbol{R}]$: when $r = 0$, the discrepancy is minimized and $[\boldsymbol{R}, \boldsymbol{M}]$ is identical with $[\boldsymbol{M}, \boldsymbol{R}]$; when $r = 1$, the discrepancy is maximized and (2) is reduced to the zero-padding; when $0 < r < 1$, the discrepancy increases as r increases. Therefore, we design the $\boldsymbol{R}$ and $\boldsymbol{M}$ as follows:

$$\begin{aligned} \boldsymbol{R} &= \tfrac{1+r}{z} \boldsymbol{I} \\ \boldsymbol{M} &= \tfrac{1-r}{z} \boldsymbol{I} \end{aligned} \tag{4}$$

where z is the normalization term.

The choice of z affects the relation between r and the discrepancy of feature transformations when $0 < r < 1$. Empirically, we find that choosing z as the L_2 normalization, i.e., $z = \sqrt{(1-r)^2 + (1+r)^2}$, resulting in satisfying performance. Theorem 1 shows the relation between r and the discrepancy of the feature transformations, which is measured by the principle angles as detailed later. This theorem shows that the principle angle is a monotonic function of r.

Theorem 1. *Under the definition of* (4) *and let* $z = \sqrt{(1-r)^2 + (1+r)^2}$*, all the principle angles of* $[\boldsymbol{R}, \boldsymbol{M}]^T$ *and* $[\boldsymbol{M}, \boldsymbol{R}]^T$ *are*

$$\theta = \arccos(\frac{2}{r^2+1} - 1) \tag{5}$$

proof. Let $\boldsymbol{u}_k \in span([\boldsymbol{R}, \boldsymbol{M}]^T)$ and $\boldsymbol{v}_k \in span([\boldsymbol{M}, \boldsymbol{R}]^T)$, then the k.th principle angle is defined as:

$$\begin{aligned} \cos(\theta_k) \quad & = \max \boldsymbol{u}_k^T \boldsymbol{v}_k \\ s.t. \quad & \boldsymbol{u}_k^T \boldsymbol{u}_k = 1 \\ & \boldsymbol{v}_k^T \boldsymbol{v}_k = 1 \\ & \boldsymbol{u}_i^T \boldsymbol{u}_k = 0 \;\; \boldsymbol{v}_i^T \boldsymbol{v}_k = 0 \;\; \text{where } i \neq k \end{aligned} \tag{6}$$

Note that $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ can be represented as $\boldsymbol{u}_k = [\boldsymbol{R}, \boldsymbol{M}]^T \boldsymbol{h}_k$ and $\boldsymbol{v}_k = [\boldsymbol{M}, \boldsymbol{R}]^T \boldsymbol{l}_k$ where $\boldsymbol{h}_k$ and $\boldsymbol{l}_k$ are $d \times 1$ unit vectors, $\boldsymbol{u^T v}$ can be induced as follows:

$$\begin{aligned} \boldsymbol{u}_k^T \boldsymbol{v}_k &= \boldsymbol{h}_k^T (\boldsymbol{R}^T \boldsymbol{M} + \boldsymbol{M}^T \boldsymbol{R}) \boldsymbol{l}_k \\ &= \tfrac{2(1-r)(1+r)}{(1+r)^2 + (1-r)^2} \boldsymbol{h}_k^T \boldsymbol{l}_k \\ &= (\tfrac{2}{r^2+1} - 1) \boldsymbol{h}_k^T \boldsymbol{l}_k \end{aligned} \tag{7}$$

Since $\boldsymbol{h}_k$ and $\boldsymbol{l}_k$ are unit vectors, then $\max \boldsymbol{h}_k^T \boldsymbol{l}_k = 1$ for $k = 1, 2, \ldots, d$. □

2.3 A Transformation-level Discrepancy Modeling

In the last section, we measure the discrepancy of mapping functions in the feature level. From (3), we find that the discrepancy can be further controlled by restricting the difference between $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, which can be achieved by forming a regularization term for minimization. Such a regularization is complementary to the feature-level augmentation introduced in the last section, since it is on the transformation level when applied our approach to existing models.

More specifically, we form $||\boldsymbol{U}_1 - \boldsymbol{U}_2||^2$ as a regularization term for minimization, which can be further represented as $\boldsymbol{U}^T \boldsymbol{B} \boldsymbol{U}$ where $\boldsymbol{B} = [\boldsymbol{I}, -\boldsymbol{I}; -\boldsymbol{I}, \boldsymbol{I}]$. By further combining with the ridge regularization $\lambda' \boldsymbol{U}^T \boldsymbol{U}$, i.e. $\boldsymbol{U}^T \boldsymbol{B} \boldsymbol{U} + \lambda' \boldsymbol{U}^T \boldsymbol{U}$, the proposed regularization term can be re-expressed $\lambda \boldsymbol{U}^T \boldsymbol{C} \boldsymbol{U}$ where $\boldsymbol{C} = [\boldsymbol{I}, -\beta \boldsymbol{I}; -\beta \boldsymbol{I}, \boldsymbol{I}]$ and $\beta = \frac{1}{1+\lambda'} < 1$. We call this regularization term the *View Discrepancy Regularization*.

Now we show that one can perform a matrix transform to integrate this regularization, so that the subspace/metric learning methods need not to be modified except adding a simple ridge regularization. Note that with the view discrepancy regularization, many subspace/metric learning methods ([Pedagadi *et al.*, 2013; Xiong *et al.*, 2014; Mignon and Jurie, 2012] can be cast as:

$$\begin{aligned} \min_{\boldsymbol{U}} \quad & f(\boldsymbol{U}^T \boldsymbol{X}_{aug}) + \lambda \boldsymbol{U}^T \boldsymbol{C} \boldsymbol{U} \\ s.t. \quad & g_i(\boldsymbol{U^T X}_{aug}) \quad i = 1, 2, \cdots, c \end{aligned} \tag{8}$$

where f is the objective function and g_i are the constraints.

Since $\beta < 1$, $\boldsymbol{C}$ is full-rank and can be decomposed into $\boldsymbol{C} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$, and thus $\boldsymbol{P}^T\boldsymbol{C}\boldsymbol{P} = \boldsymbol{\Lambda}$. Let $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{H}$, and so $\boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U} = \boldsymbol{H}^T\boldsymbol{H}$. (8) is equivalent to:

$$\begin{array}{ll} \min\limits_{\boldsymbol{H}} & f(\boldsymbol{H}^T\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{X}_{aug}) + \lambda\boldsymbol{H}^T\boldsymbol{H} \\ s.t. & g_i(\boldsymbol{H}^T\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{X}_{aug}) \quad i = 1, 2, \cdots, c \end{array} \tag{9}$$

If we transform the augmented features $\boldsymbol{X}_{aug}$ to $\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{X}_{aug}$, $\boldsymbol{H}$ could be solved using standard subspace/metric learning with the normal ridge regression, and then $\boldsymbol{U}$ can be obtained by $\boldsymbol{U} = \boldsymbol{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{H}$.

In this work, the transformed feature, $\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{P}^T\boldsymbol{X}_{aug}$, is called **Mirror Representation**, since the original feature as well as its phantom are jointly transformed into a more robust space for better representation of the view-specific knowledge.

2.4 Kernelized Mirror Representation

The Mirror Representation discussed above provides *linear* view-specific mappings. However, the distortion functions and the intrinsic data distribution are usually non-linear. To overcome this problem, we borrow the idea of kernel modeling in machine learning to develop a kernel mirror representation.

In order to extend the feature augmentation approach to the kernel version, the anchor points are enriched to include both training data points and their phantoms, and the location of phantoms are view-specific. In our modeling, the anchor points are data points from training set. Hence the mapping functions are defined as below:

$$\begin{array}{rcl} f_a(\boldsymbol{x}^a) & = & \boldsymbol{\alpha}^T\boldsymbol{k}(\boldsymbol{X}[\boldsymbol{R}, \boldsymbol{M}], \boldsymbol{x}^a) \\ f_b(\boldsymbol{x}^b) & = & \boldsymbol{\alpha}^T\boldsymbol{k}(\boldsymbol{X}[\boldsymbol{M}, \boldsymbol{R}], \boldsymbol{x}^b) \end{array} \tag{10}$$

where $\boldsymbol{R}$ and $\boldsymbol{M}$ are $n \times n$ matrices defined in (4), and thus $\boldsymbol{X}[\boldsymbol{R}, \boldsymbol{M}]$ and $\boldsymbol{X}[\boldsymbol{M}, \boldsymbol{R}]$ are anchor points of view a and view b respectively. $\boldsymbol{k}(\boldsymbol{A}, \cdot)$ is defined as $[k(\boldsymbol{A}_1, \cdot), k(\boldsymbol{A}_2, \cdot), \cdots, k(\boldsymbol{A}_n, \cdot)]$ where $k(\cdot, \cdot)$ is the kernel function and $\boldsymbol{A}$ is the anchor points.

Compared to the linear case, for the kernel case, $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are replaced with $\boldsymbol{X}\boldsymbol{\alpha}_1$ and $\boldsymbol{X}\boldsymbol{\alpha}_2$, respectively. Hence the view discrepancy regularization $\boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U}$ is replaced with $\boldsymbol{\alpha}_1^T\boldsymbol{X}^T\boldsymbol{C}\boldsymbol{X}\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1^T\boldsymbol{C}'\boldsymbol{\alpha}_2$, respectively, where $\boldsymbol{C}' = [\boldsymbol{K}, -\beta\boldsymbol{K}, -\beta\boldsymbol{K}, \boldsymbol{K}]$.

3 A Principle Angle based Parameter Estimation Strategy

The Mirror Representation consists of feature-level discrepancy and transformation-level discrepancy. For those modeling, there are two parameters r and β to jointly control discrepancy.

Intuitively, when feature distributions of different views are largely mismatched, the optimal view-specific mapping functions are more discrepant; otherwise, samples of different views consist of more shared features and the mapping functions are likely to be similar. In this work, we propose to measure the degree of distribution mismatch by the principle angles of different views. Let $\boldsymbol{G}_a$ and $\boldsymbol{G}_b$ be the PCA-subspace of samples of view a and view b. The principle angles can be computed by the Singular Value Decomposition of $\boldsymbol{G}_a^T\boldsymbol{G}_b$ [Hamm and Lee, 2008]:

$$\boldsymbol{G}_a^T\boldsymbol{G}_b = \boldsymbol{V}_1\cos(\Theta)\boldsymbol{V}_2^T \tag{11}$$

where $\cos(\Theta) = diag(\cos(\theta_1), \cos(\theta_2), \cdots \cos(\theta_d))$ and θ_i are principle angles.

With the principle angles, r and β are set as:

$$r = \frac{1}{d}\sum_{k=1}^{d}(1 - \cos^2(\theta_k)), \quad \beta = \frac{1}{d}\sum_{k=1}^{d}\cos(\theta_k) \tag{12}$$

4 Experiment

4.1 Datasets and Settings

Datasets: Our experiments were conducted on three publicably available datasets: PRID450S [Roth *et al.*, 2014], VIPeR [Gray *et al.*, 2007] and CUHK01 [Li *et al.*, 2012]. All three datasets are with significant feature distortion. PRID450S contains 450 image pairs recorded from two different but static surveillance cameras. VIPeR contains 632 pedestrian image pairs captured outdoor with varying viewpoints and illumination conditions. CUHK01 contains 971 pedestrians from two disjoint camera views. Each pedestrian has two samples per camera view.

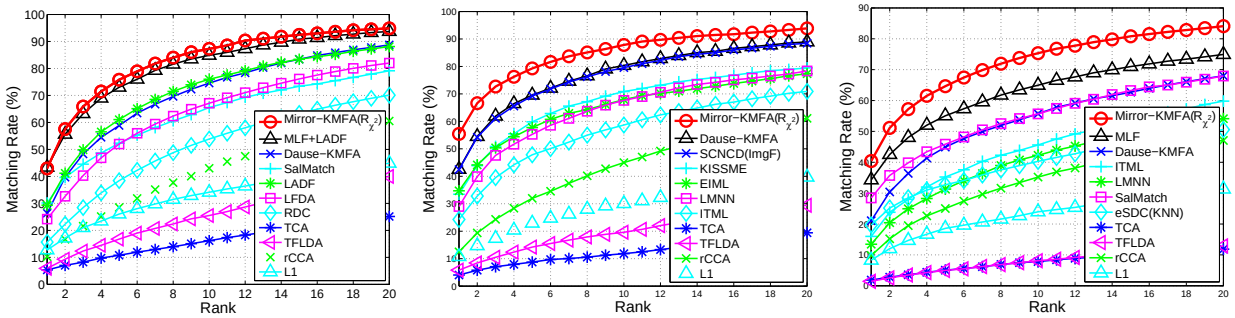
Features: We equally partitioned each image into 18 horizontal stripes, and RGB, HSV, YCbCr, Lab, YIQ and 16 Gabor texture features were extracted for each stripe. For each feature channel, a 16D histogram was extracted and then normalized by L_1-norm. All histograms were concatenated together to form a single vector. Since the PRID450S provides automatically generated foreground mask, our features were extracted from the foreground area in this dataset. For CUHK01 and VIPeR, our features were extracted on the whole image.

Experimental Protocol: Our experiments follow the same *single-shot* protocol: each time half of the pedestrians were selected randomly to form the training set, and the remaining pedestrian images were used to form the gallery set and testing set. For CUHK01, each pedestrian has 2 images for each view; we randomly selected *one* of them to form the gallery. The cumulative matching characteristic (CMC) curve is used to measure the performance of each method on each dataset. A rank k matching rate indicates the percentage of the probe image with correct matches found in the top k rank against the p gallery images. In practice, a high rank-1 matching rate is critical and the top k matching rank matching rate with a small k value is also important since the top matching images can be verified by human [Zheng *et al.*, 2013]. To obtain statistically significant results, we repeated the procedure 10 times and reported the average results.

Methods for Comparison: PCCA, KPCCA[Mignon and Jurie, 2012], MFA and KMFA [Yan *et al.*, 2007] are selected to evaluate the improvement of using our Mirror Representation. The kernels used in KPCCA and KMFA are χ^2 and R_{χ^2} (RBF-χ^2). The Zero-Padding representation is also evaluated for comparison. Besides, Mirror-KMFA is also compared with various person re-identification methods [Zhao *et al.*, 2014; 2013a; Li *et al.*, 2013; Pedagadi *et al.*, 2013; Zheng *et al.*, 2013; Yang *et al.*, 2014; Kostinger *et al.*, 2012; Hirzer *et al.*, 2012; Weinberger *et al.*, 2006; Davis *et al.*, 2007; Zhao

	Representation	Mirror Representation				Original Feature				Zero-Padding			
	Rank	1	5	10	20	1	5	10	20	1	5	10	20
VIPeR	$KMFA(R_{\chi 2})$	**42.97**	**75.82**	**87.28**	**94.84**	37.37	71.23	84.72	93.45	33.67	67.66	82.31	91.87
	$KMFA(\chi^2)$	**39.62**	**71.36**	**84.18**	**93.23**	35.57	67.34	81.14	91.74	30.28	63.54	77.88	89.15
	$KPCCA(R_{\chi 2})$	**32.88**	**67.91**	**82.03**	**91.77**	29.05	62.94	78.26	89.68	21.84	52.44	67.37	79.40
	$KPCCA(\chi^2)$	**29.37**	**64.11**	**78.96**	**90.63**	25.63	59.78	76.27	87.78	18.77	51.17	66.77	82.31
	MFA	**33.48**	**63.10**	**75.60**	**86.55**	30.76	59.43	73.61	85.41	21.87	52.06	66.58	81.39
	$PCCA$	**27.56**	**60.57**	**75.66**	**87.37**	25.47	56.96	71.08	85.25	22.53	55.60	71.30	86.36
CUHK01	$KMFA(R_{\chi 2})$	**40.40**	**64.63**	**75.34**	**84.08**	34.98	60.16	71.27	81.50	33.53	59.00	70.20	80.24
	$KMFA(\chi^2)$	**37.31**	**61.11**	**71.36**	**81.25**	32.34	56.14	67.52	77.73	31.35	56.71	67.56	78.18
	$KPCCA(R_{\chi 2})$	**29.57**	**56.53**	**69.21**	**79.40**	25.30	52.40	64.61	76.76	17.84	41.53	53.95	67.83
	$KPCCA(\chi^2)$	**26.69**	**54.40**	**66.88**	**77.87**	22.79	48.65	62.10	74.06	17.84	41.53	53.95	67.83
	MFA	**25.47**	**48.38**	**58.86**	**69.19**	20.71	41.51	52.42	63.21	14.13	33.12	43.10	54.07
	$PCCA$	**19.74**	**40.96**	**52.44**	**65.00**	16.79	38.13	49.29	61.35	3.89	9.02	12.32	16.28
PRID450S	$KMFA(R_{\chi 2})$	**55.42**	**79.29**	**87.82**	**93.87**	52.76	77.56	84.71	91.56	46.18	74.13	84.31	92.40
	$KMFA(\chi^2)$	**53.42**	**77.29**	**85.82**	**91.51**	51.02	75.29	82.80	89.47	41.82	71.29	81.82	90.04
	$KPCCA(R_{\chi 2})$	**41.51**	**71.51**	**81.42**	**91.24**	40.09	68.76	79.73	90.13	33.60	65.78	78.18	88.00
	$KPCCA(\chi^2)$	**39.82**	**68.31**	**80.22**	**89.82**	37.60	66.18	78.49	88.62	28.27	58.71	72.40	85.60
	MFA	**40.58**	**77.56**	**67.47**	**86.58**	38.22	63.42	73.87	83.64	21.16	50.00	62.98	76.84
	$PCCA$	**38.40**	**68.40**	**79.51**	**88.31**	36.76	65.69	76.22	85.16	32.80	64.62	76.98	87.38

Table 1: Top Matching Rank(%) on VIPeR, CUHK01 and PRID450S. Three representations, the Mirror Representation, the original feature and the zero-padding are compared. Among the three representations, we mark the best performance for each dataset, each method and each rank in **red** color and mark the second best in blue.



(a) VIPeR (b) PRID450S (c) CUHK01

Figure 3: Comparison to the state-of-the-art on VIPeR, PRID450S and CUHK01

et al., 2013b], domain adaptation methods [Daumé III, 2009; Pan *et al.*, 2011; Si *et al.*, 2010] as well as the baselines a regularized CCA (rCCA) and L_1-norm method. In order to balance the scale of the objective function when applying our mirror representation Eq. 9), λ is set as 10^{-6}, 10^{-2}, 10^{-2} and 0.5 for KPCCA, KMFA, PCCA and MFA respectively.

rank	1	5	10	20
Mirror-KMFA($R_{\chi 2}$)	42.97	**75.82**	**87.28**	**94.84**
MLF+LADF [Zhao *et al.*, 2014]	**43.39**	73.04	84.87	93.70
Daumé-KMFA($R_{\chi 2}$))[Daumé III, 2009]	26.36	58.86	74.56	88.73
MLF [Zhao *et al.*, 2014]	29.11	52.34	65.95	79.87
SalMatch [Zhao *et al.*, 2013a]	30.16	52.31	65.54	79.15
LADF [Li *et al.*, 2013]	29.34	61.04	75.98	88.10
LFDA [Pedagadi *et al.*, 2013]	24.18	52.00	67.12	82.00
RDC [Zheng *et al.*, 2013]	15.66	38.42	53.86	70.09
TFLDA [Si *et al.*, 2010]	5.92	16.77	25.51	40.00
TCA [Pan *et al.*, 2011]	5.35	10.79	16.17	25.16
rCCA(baseline)	11.80	31.01	45.66	63.10
L_1(baseline)	12.15	26.01	32.82	42.47

Table 2: Top Matching Rank (%) on VIPeR compared to the state-of-the-art.

rank	1	5	10	20
Mirror-KMFA($R_{\chi 2}$)	**55.42**	**79.29**	**87.82**	**91.56**
Daumé-KMFA($R_{\chi 2}$))[Daumé III, 2009]	42.53	69.56	80.36	88.93
SCNCD(ImgF) [Yang *et al.*, 2014]	42.44	69.22	79.56	88.44
KISSME [Kostinger *et al.*, 2012]	33.47	59.82	70.84	79.47
EIML [Hirzer *et al.*, 2012]	34.71	57.73	67.91	77.33
LMNN [Weinberger *et al.*, 2006]	28.98	55.29	67.64	78.36
ITML [Davis *et al.*, 2007]	24.27	47.82	58.67	70.89
TFLDA [Si *et al.*, 2010]	5.82	14.22	19.60	29.42
TCA [Pan *et al.*, 2011]	4.04	8.76	11.82	19.42
rCCA(baseline)	12.49	31.96	44.98	61.11
L_1(baseline)	10.76	22.93	30.04	39.69

Table 3: Top Matching Rank (%) on PRID450S compared to the state-of-the-art.

4.2 Effectiveness of Mirror Representation

Table 1 shows the comparison among the top rank matching performance of MFA, KMFA, PCCA, and KPCCA across 3 representations: the Mirror Representation, the original feature and the zero-padding. It is clear that the Mirror Representation enhances the performance of all the methods we tested, while the zero-padding does not. As we discussed in Sec. 2.1, zero-padding is likely to loses the control of the discrepancy of view-specific transformations, probably leading

rank	1	5	10	20
Mirror-KMFA(R_{χ^2})	**40.40**	**64.63**	**75.34**	**84.08**
Daumé-KMFA(R_{χ^2}))[Daumé III, 2009]	21.08	44.92	55.72	68.02
MLF [Zhao *et al.*, 2014]	34.30	55.06	64.96	74.94
ITML [Davis *et al.*, 2007]	15.98	35.22	45.60	59.81
LMNN [Weinberger *et al.*, 2006]	13.45	31.33	42.25	54.11
SalMatch [Zhao *et al.*, 2013a]	28.45	45.85	55.67	67.95
eSDC(KNN) [Zhao *et al.*, 2013b]	19.67	32.72	40.29	50.58
TFLDA [Si *et al.*, 2010]	1.44	4.97	8.10	12.89
TCA [Pan *et al.*, 2011]	1.84	5.03	7.84	11.86
rCCA(baseline)	8.21	24.19	34.87	47.76
L_1(baseline)	4.45	12.97	19.80	29.94

Table 4: Top Matching Rank (%) on CUHK01 compared to the state-of-the-art.

to a serious overfitting to training data and therefore generalizing not very well on unseen testing data, since in person re-identification the people for training will not appear in the testing stage.

Our Mirror Representation enhances the performance in both linear case and non-linear case. It is not surprising that using the kernel trick boosts the performance, since the distortion functions are non-linear and non-linear mappings are more suitable. Note that a more complex kernel R_{χ^2} usually achieves better results than the simpler one χ^2, which may also indicates the complexity of the feature distortion.

4.3 Comparison to Related Methods

From tables 2-4, we find that the KMFA(R_{χ^2}) with Mirror Representation achieves promising results on the three datsets. We compare this method with the other state-of-the-art methods. Tables 2-4 and Figure 3 show that our method significantly outperforms other methods except MLF+LADF [Zhao *et al.*, 2014] at rank-1 on VIPeR. However, MLF+LADF combines MLF [Zhao *et al.*, 2014] and LADF [Li *et al.*, 2013], and neither of the these two methods outperforms ours. We also compare another feature augmentation method [Daumé III, 2009] with exactly the same KMFA setting denoted as Daumé-KMFA(R_{χ^2}). Since it also does not control the discrepancy of view-specific mappings, the performance is unsatisfactory.

We additionally compare two popular domain adaptation methods TCA[Pan *et al.*, 2011] and TFLDA[Si *et al.*, 2010]. As shown, they do not perform well. This is probably because both TCA and TFLDA assume $P(Y_s|X_s) = P(Y_t|X_t)$ or $P(X_s) = P(X_t)$, which is not suitable for person re-identification, since the people to train are the ones to test.

4.4 Parameter Sensitivity Analysis

The Mirror Representation has two parameters: r and β. Using either the feature-level discrepancy or the transformation-level discrepancy could enhance the performance. Due to space limit, we take KMFA on CUHK01 for illustration. To demonstrate the effectiveness of the feature-level discrepancy, we disable the transformation-level discrepancy by setting $\beta = 0$ and evaluate the parameter sensitivity of r. Figure 4 (a) shows that turning r can significantly enhance the performance by nearly 5%. When $r = 1$ and $\beta = 0$, the performance is unsatisfactory, since the two learned view-specific transformation may be irrelevant. The demonstration of the transformation-level discrepancy is similar: we disable the feature-level discrepancy by setting $r = 1$ and evaluate the sensitivity of β. Figure 4 (b) shows that turning β can also boost the performance significantly. Note that the best result achieved by turning r or β exhaustively is 40.90% rank-1 matching rate, while in our experiments we turn r and β automatically and achieve 40.40%. This demonstrates the effectiveness of our parameter estimation strategy. Figure 4 (c) shows the parameter sensitivity of λ, which is the ridge regularizer in (9). As shown, there is a clear trend that the performance will go down as λ increase.

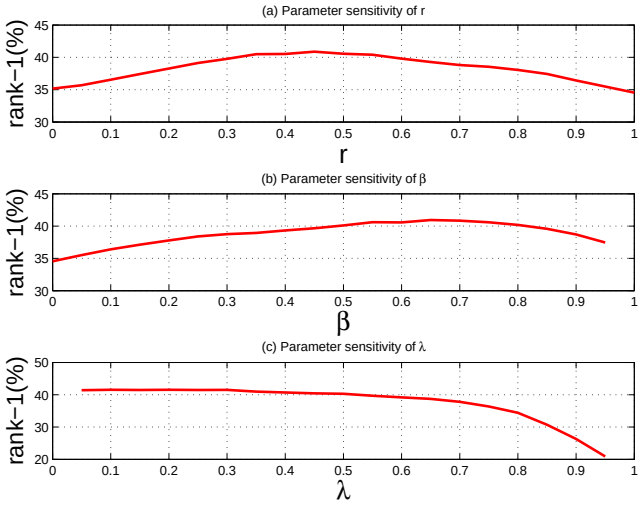


Figure 4: Parameter Sensitivity Analysis with KMFA on CUHK01.

5 Conclusion

In this paper, we have proposed the Mirror Representation to alleviate the view-specific feature distortion problem for person re-identification. The Mirror Representation consists of two complementary components: feature augmentation and view discrepancy regularization, designed for feature-level discrepancy and transformation-level discrepancy respectively. Principles angles of different camera views are used to estimate the parameters. With our Mirror Representation, the kernel marginal fisher discriminant analysis significantly outperforms the state-of-the-art on all three popular datasets.

6 Acknowledgements

This project was partially supported by National Science & Technology Pillar Program (No. 2012BAK16B06), Natural Science Foundation Of China (No. 61472456), Guangzhou Pearl River Science and Technology Rising Star Project under Grant 2013J2200068, the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant S2013050014265, Guangdong Provincial Government of China through the Computational Science Innovative Research Team Program, and GuangZhou Program (2014J4100114).

References

[Daumé III, 2009] Hal Daumé III. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815*, 2009.

[Davis *et al.*, 2007] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning. In *ICML*, 2007.

[Geng *et al.*, 2011] Bo Geng, Dacheng Tao, and Chao Xu. Daml: Domain adaptation metric learning. *TIP*, 2011.

[Gray *et al.*, 2007] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *VS-PETS Workshop*. Citeseer, 2007.

[Hamm and Lee, 2008] Jihun Hamm and Daniel D Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *ICML*, 2008.

[Hirzer *et al.*, 2012] Martin Hirzer, Peter M Roth, and Horst Bischof. Person re-identification by efficient impostor-based metric learning. In *AVSS*, 2012.

[Kostinger *et al.*, 2012] M Kostinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[Kviatkovsky *et al.*, 2013] Igor Kviatkovsky, Amit Adam, and Ehud Rivlin. Color invariants for person reidentification. *PAMI*, 35(7):1622–1634, 2013.

[Li *et al.*, 2012] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *ACCV*. 2012.

[Li *et al.*, 2013] Zhen Li, Shiyu Chang, Feng Liang, Thomas S Huang, Liangliang Cao, and John R Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013.

[Li *et al.*, 2014] Wen Li, Lixin Duan, Dong Xu, and Ivor W Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *PAMI*, 2014.

[Ma *et al.*, 2013] Andy J Ma, Pong C Yuen, and Jiawei Li. Domain transfer support vector ranking for person re-identification without target camera label information. In *ICCV*, 2013.

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

[Pan *et al.*, 2011] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 2011.

[Pedagadi *et al.*, 2013] Sateesh Pedagadi, James Orwell, Sergio Velastin, and Boghos Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, 2013.

[Roth *et al.*, 2014] Peter M. Roth, Martin Hirzer, Martin Koestinger, Csaba Beleznai, and Horst Bischof. Mahalanobis distance learning for person re-identification. In Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen C. Loy, editors, *Person Re-Identification*, pages 247–267. Springer, 2014.

[Si *et al.*, 2010] Si Si, Dacheng Tao, and Bo Geng. Bregman divergence-based regularization for transfer subspace learning. *TKDE*, 2010.

[Weinberger *et al.*, 2006] Kilian Weinberger, John Blitzer, and Lawrence Saul. Distance metric learning for large margin nearest neighbor classification. *NIPS*, 2006.

[Xiao and Guo, 2014] Min Xiao and Yuhong Guo. Feature space independent semi-supervised domain adaptation via kernel matching. *PAMI*, 2014.

[Xiong *et al.*, 2014] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*. 2014.

[Yan *et al.*, 2007] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *PAMI*, 2007.

[Yang *et al.*, 2014] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z Li. Salient color names for person re-identification. In *ECCV*. 2014.

[Zhao *et al.*, 2013a] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Person re-identification by salience matching. In *ICCV*, 2013.

[Zhao *et al.*, 2013b] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.

[Zhao *et al.*, 2014] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Learning mid-level filters for person re-identfiation. In *CVPR*, 2014.

[Zheng *et al.*, 2012] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Transfer re-identification: From person to set-based verification. In *CVPR*. IEEE, 2012.

[Zheng *et al.*, 2013] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Reidentification by relative distance comparison. *PAMI*, 35(3):653–668, 2013.