Sample Specific Multi-Kernel Metric Learning for Person Re-identification

Jun FANG¹, Ru-fei ZHANG² and Feng JIANG²
¹University of Chinese Academy of Sciences, Beijing, China
²Beijing Institute of Control and Electronics Technology, Muxidi North Street, 100038, Beijing, China

Keywords: Person re-identification, Sample specific multi-kernel learning, Metric learning.

Abstract. The existing metric learning based Person Re-Identification are challenged with large appearance variations across non-overlapping cameras. In this paper, we propose to learn a similarity metric that incorporates robustness against cross views appearance variations. The robustness in metric learning is achieved by learning an adaptive sample specific multi-kernel space for each pair of cross-view images for each person in the training set, referred as SSMK. To this end, we first project features of each person into a multi-kernel feature space, where the subtle features of a person are highlighted to help differentiate this person from others. With the discriminative feature projection of all persons in the learned SSMK space, a robust global metric referred as SSMK-M is learned via Local Fisher Discriminant Analysis (LFDA). During testing, SSMK space is constructed in an incremental way to minimize appearance variations. Experiments show that the proposed approach out-performs most non-deep-learning based approaches on popular benchmarks including VIPeR, GRID and CUHK01.

Introduction

Person re-identification refers to identify a person observed in a large network of non-overlapping cameras which has practical applications in surveillance, security, image retrieval and non-temporal tracking [1] scenarios. However, high-performance person re-identification in the wild still remains difficult due to large variance in poses, viewpoints and illumination [2], as shown in Fig. 1. In many retrieval-based tasks, similarity metric learning plays a vital role which try to learning a subspace where the similarity between a bunch of images of the same person are higher than that of different persons. However, the large variance in poses, viewpoints and illuminations would cause it very difficult to learn, as shown in Fig.2b (i.e. RGB, HSV and LBP histograms).

In this paper, we target at reducing disparity in the feature space and improve the discriminative capability of similarity metric. Motivated by the fact that different features share different latent discriminant capabilities, we aim to highlight these latent salient and discriminant appearance features which differs from previous approaches that mainly design handcrafted feature, either by learning dictionary [7, 8, 9], or directly extracting raw features [10, 11]. However, as each person owns unique discriminative appearance characteristics. Therefore, highlighting the salient and discriminating features globally in latent common subspaces [12], or applying camera-view transformation [13], would significantly deteriorate the local differentiating features. Inspired by the sample based approaches [14, 15], we aim to minimize cross-views disparity locally for each person.
We propose to learn a sample(person) specific multi-kernel space, referred as SSMK by adapting multi-kernel learning [16]. The main idea behind of our idea is to minimize feature disparity and handle the complex non-linear feature space. SSMK for each person can highlight its subtle features by projecting all heterogeneous features into a weighted multi-kernel feature space. Which can be viewed well in Fig.2c, where features are projected into the learned SSMK for a given pair, and can be seen much congruous in different views compared to Fig.2b.

After projecting features into the SSMK feature space for each individual person, a global distance metric, referred as SSMK-M, is learned to perform re-identification. SSMK-M is learned with Local Fisher Discriminant Analysis [17] to gain robustness and capability of differentiation among different persons. The main contributions of this work are summarized as follows:

SSMK is learned for each person to highlight the subtle differentiating features.
SSMK can simultaneously weight features, minimize non-linearity in the feature space and integrate heterogeneous features for learning a robust metric, which can be integrated with most metric learning approaches for person re-identification.

Related Works
Our work is related to appearance-based person re-identification, including extracting discriminant features and metric learning. In the following, we make a brief introduction in these two fields.

Feature Extraction
Robust features are supposed to achieve high robustness between images of the same person. Based on this propose, features type can be classified into either low-level or mid-level. Low-level features are directly extracted from raw pixels of small patches, including LOMO [17]. [18, 19] proposes to acquire semantic color name for each pixel of each region to obtain robustness against light variations. Further, [20] acquire pose invariance by matching extracted patch features spatially [21] learned multiple Gaussian distributions of each image patch.
Mid-level features represent semantic information from body parts, attributes or related objects. In [22], SDALF takes advantage of human body symmetry to obtain both pose and illumination invariant
features. In [23], GaLF is formed by accumulating local color and texture features with the spatial information. Similarly, [24] accumulate different features from each body part to enhance discriminative information. Moreover, human attributes are learned for re-identification [25, 26].

**Metric Learning**

It is required to build relation between different features and reduce background noises or outliers in the feature space, a discriminant low dimension subspace is learned to perform re-identification.

[27] learns metric based on LMNN using similarity and dissimilarity constraints between pair of samples. Tao et al. [28] applied dual-regularization to further improve the performance and learned a robust metric. [29] used the locally threshold to learn a metric, referred as LADF. [30] utilized all the similarity values of a training sample with the rest of samples in the training set to learn a metric.

Since, re-identification aims to identify a particular person from the others. Hence, [31] took liberty and applied discriminant analysis in identifying each individual person by learning a metric using Local Fisher Discriminant Analysis. [32] utilized different linear and non-linear kernels to mitigate the non-linearity in the feature space. Similarly, [33] also explicitly learned polynomial kernel.

Though the above kernel based metric learning provide discriminant subspace for person re-identification, however a single kernel may not be discriminant enough to differentiate each person in the complex heterogeneous feature space. Multi-kernel learning (MKL) [16] has already demonstrated its inherent capabilities in many computer vision tasks including classification [34], event detection [35] and recognition [36, 37]. Inspired by the success of MKL, this work intends to adapt multi-kernel for metric learning, and then learn an adaptively weighted multi-kernel space for each individual person on the query set.

**Methodology**

The methodology of learning sample specific multi-kernel space SSMK, and learning metric SSMK-M using SSMK is visually described in Fig.3. SSMK is learned iteratively for each of the observed training pair as showed in Fig.3a. During training, different features are extracted from each image pair, and are projected into different linear and non-linear kernel spaces. These projected features are used to learn SSMK space for each training person iteratively. The learning procedure of SSMK is given in details in section 3.2. Finally, after getting SSMK space for each person in the training set, a global metric SSMK-M, denoted as $M^*$, is learned for re-identification.

During testing, multi-kernel space for each query is also learned in an incremental way. Initially, SSMK space $\beta_j$ is first learned for only query probe through a KNN approach, as shown in Fig.3b. Then with the learned $\beta_j$ space of the given probe, all the gallery samples are projected into SSMK space for initial re-identification. Taking top-$r$ ranked gallery samples from initial identification, $r$ probe-gallery pairs are formed to learn a more reliable pairwise SSMK space, as shown in SSMK learning for Probe-Gallery part in Fig.3b. Finally, an averaged SSMK space $\beta_j$ for each probe-gallery pair is obtained to perform final re-identification using metric $M^*$.

**Feature and Kernel**

Feature extraction procedure is shown in Fig.4. Similar to [40, 41, 42, 43, 44], the image is divided into six horizontal bands, both color and texture features are used. The former includes RGB, YUV, HSV, LAB and YCbCr. Dense SIFT from [45, 46]. In addition, color naming feature [19] and deep features [47] are used for each patch. We define $F$ to represent these features:

$$F = \{f_n, n = 1, ..., N\} \quad (1)$$
Figure 3. Methodology of the proposed Sample Specific Multi-Kernels and Metric SSMK-M along with Re-identification Approach

Figure 4. Feature Extraction (a) Image Pair from VIPeR, (b) Extract Patches Features, (c) Accumulate Patch Features to form F

Where $N$ represent the number of feature types, and each feature type $f_n$ is $l_1$-normalized.

All the features in the set $F$, are now projected into different kernel spaces. We adopt linear and non-linear kernels in this paper. For non-linear kernels, Chi-Square and RBF kernel are used:

$$K_{chi}(x^a_i, x^b_i) = \exp\left( -\frac{e^{-\alpha^2}}{2\alpha^2} \right), K_{rbf}(x^a_i, x^b_i) = \exp\left( -\alpha \sum (\frac{x^a_i - x^b_i}{\sigma})^2 \right).$$

(2)

Where $x^a_i$ and $x^b_i$ are samples of person $i$ in different views. Linear kernels including linear and poly-nominal is used, which are defined as:

$$K_l(x^a_i, x^b_i) = \langle x^a_i, x^b_i \rangle, \quad K_p(x^a_i, x^b_i) = \langle x^a_i, x^b_i \rangle + 1^m, m \in N.$$  

(3)

All these kernels are defined as $\Phi$ follows:

$$\Phi = \{\Phi_k, k = 1, ..., K\}$$

(4)

Where $K$ is the number of kernels and $\Phi_k$ is the $k$-th kernel.

**Sample Specific Multi-Kernel Space Learning**

After getting all the kernel projections of each feature, sample specific multi-kernel space SSMK is now learned by progressively computing weights of each kernel $\Phi_k$ for each person $i$ with its image.
pairs \((x^a_i, x^b_i)\) from different camera view. To learn the weight of \(\Phi_k\), its ranking error \(Err\) is obtained firstly, which is defined as:

\[
Err_{i, f_n, \Phi_k} = \frac{d_{\text{rank1}}(\Phi_k(x^a_i, \Phi_k(x^b_i, f_n)), d(\Phi_k(x^a_i, \Phi_k(x^b_i, f_n)))
\]

(5)

Where \(x^a_i, f_n\) represents the feature \(f_n\) for sample \(x^a_i\) and \(\Phi_k(x^b_i, f_n)\) is its kernel projection. In Eq.5, ranking error \(Err_{i, f_n, \Phi_k}\) in the \(t\)-th iteration is computed using the ratio between similarity values. \(d_{\text{rank1}}()\) gives similarity between \(x^a_i\) and the gallery sample \(x^b_{\text{rank1}, i, f_n}\) of which are ranked as top-1 match. While measures similarity between actual pair of samples \((x^a_i, x^b_i)\). Since, \(N\) features are used, the average ranking error denoted as \(Err_{\text{avg}}\) for kernel \(\Phi_k\) is given as:

\[
Err_{\text{avg}, i, \Phi_k, t} = \frac{\sum_{n=1}^{N} Err_{i, f_n, \Phi_k}}{N}
\]

(6)

Where \(Err_{\text{avg}, i, \Phi_k, t}\) is the average ranking error of \(\Phi_k\) in \(t\)-th iterations? Using \(Err_{\text{avg}, i, \Phi_k, t}\) the weight \(\alpha_{\Phi_k}\) of \(\Phi_k\) in \(t\)-th iteration is then computed as:

\[
\alpha_{i, \Phi_k, t} = \exp^{-Err_{\text{avg}, i, \Phi_k} \times w_{i, \Phi_k}}
\]

(7)

Where \(\alpha_{i, \Phi_k, t}\) is the newly updated weight of \(\Phi_k\) for pair \((x^a_i, x^b_i)\) in iteration \(t\), while \(w_{i, \Phi_k}\) is the initial weight for \(\Phi_k (= 1/K)\). The weights of all the kernels in the set \(\Phi\) associated to pair \((x^a_i, x^b_i)\) is then updated iteratively using Eq.7. Then, the SSMK space denoted as \(\beta_i\) of pair \((x^a_i, x^b_i)\) in \(t\)-th iteration as can be formed as:

\[
\beta_{i, j} = (\alpha_{i, \Phi_{k, t}}, \Phi_k)
\]

(8)

From the above equations, it is clear that \(\beta_i\) is dependent on \(Err_{\text{avg}}\) and pre-learned metric \(M\). As the pre-learned metric \(M\) is learned conventionally and lack of robustness against cross views appearance disparity, therefore, the obtained \(Err_{\text{avg}}\) may not be reliable. To overcome this shortcoming, different from the previous approaches [48, 49], where the kernel weights are learned once, the proposed SSMK learns and updates the weights of the kernels iteratively, by learning a new metric in each \(t\)-th iteration. Metric \(M\) is learned by projecting the features of all the training pairs into their respective SSMK spaces obtained from the previous \(t-1\) iteration (details in section 3.3), and then the learned metric \(M\) is used to compute the average ranking error in the next iteration. Then SSMK space for person \(i\) after \(T\) iterations is defined as:

\[
\beta_i = (\alpha_{i, \Phi_{k, T}}, \Phi_k)
\]

(9)

Sample Specific Multi-Kernel Based Metric Learning

Since sample specific multi-kernel space \(\beta_i\) for each training pair is already learned, we can now learn sample specific multi-kernel based metric SSMK-M, denoted as \(M^*\), \(M^*\) is learned via Local-Fisher Discriminant Analysis [32], which seeks a low dimension subspace by maximizing the similarity between images of the same identity person in different views.

\[
\max_{M^*} \text{tr}(M^{*T} S_{\text{int}}, M^*) / (M^{*T} S_{\text{intr}}, M^*)
\]

(10)
Where $S_{inter}$ and $S_{intra}$ are inter class and intra class covariance matrices respectively, which are obtained globally by projecting the features of each training pair into its corresponding **weighted multi-kernel space** $\beta_i$. However, $S_{intra}$ is computed globally which may lose significant intra-class information. Since intra-class co-variance of each person may differ from global $S_{intra}$, therefore, to incorporate the subtle distinct information of each person, $S_{intra}$ is redefined by integrating both the global and pair wise averaged class covariance between two persons [50] as,

$$S_{intra}^* = \gamma S' + (1 - \gamma)S_{intra}$$  \hspace{1cm} (11)

Where $S'$ is the pair-wised average co-variance between two different persons in the training set, while the $S_{intra}$ is the conventional global co-variance matrix. $\gamma$ ($0 < \gamma < 1$) is a factor that controls the balance between pair wise co-variance $S'$ and global co-variance $S_{intra}$.

Further, locality information is also incorporated in FDA learning by using Locality preserving projection [51]. Locality information is integrated using local scaling method given in [52], which learns an affinity matrix $A = [\alpha_{ij}]$ by measuring the distance between pair of samples. Each element $\alpha_{ij}$ in the affinity matrix $A$ is then given as:

$$\alpha_{ij} = \exp\left(-\frac{d_p(x_i^a, x_i^b)}{\sigma_i^a \sigma_i^b}\right)$$  \hspace{1cm} (12)

Where $\alpha_{ij}$ represents the affinity between samples $x_i^a$ and $x_i^b$, using the $d_p$ ($p$-norm distance). While $\sigma_i^a$ and $\sigma_i^b$ represent the local scales of samples $x_i^a$ and $x_i^b$ respectively, as defined in [53]. Affinity matrix $A$ obtained above is then combined with $S_{intra}$ and $S_{inter}$ matrices to learn the new modified $S_{intra-new}$ and $S_{inter-new}$ as:

$$S_{intra-new}^{*} = \gamma AS' + (1 - \gamma)AS_{intra},$$  \hspace{1cm} (13)

$$S_{inter-new} = AS_{inter}.$$  \hspace{1cm} (14)

**Algorithm 1 Training: SSMK and SSMK-M Learning**

**Given:** $F = \{f_1, f_2, \ldots, f_N\}$, $\Phi = \{\Phi_1, \Phi_2, \ldots, \Phi_K\}$

**Number of persons = n'**

Initialize $w_k(\Phi_k)$ (usually set to $1/K$ where $K$ is the total number of Kernels in set $\Phi$).

for $i = 1, \ldots, N$

for $k = 1, \ldots, K$

compute $Err_{ri, \Phi_k, t}$

$$= \sum_{j \neq i} d_{ri, \Phi_k} \phi_k(x_i^a, x_j^b)$$

where $d_{ri, \Phi_k} = (x_i - x_j)^T M (x_i - x_j)$

end for

compute $Err_{ri, \Phi_k} = \sum_{j \neq i} Err_{ri, \Phi_k}$

update $\alpha_{ij, \Phi_k} = \exp(-Err_{ri, \Phi_k}) \times w_i(\Phi_k)$

where $w_i(\Phi_k)$ is the initial weight of $\Phi_k$

end for

compute $\beta_{ij} = (\alpha_{ij, \Phi_k}, \Phi_k)$

where $\beta_{ij}$ is the sample specific multi-kernel space of person $i$

end for

**Output** Final smk $\beta = (\alpha_{ij, \Phi_k}, \Phi_k)$ and $\max_{M^*} Tr(M^* S_{inter-new} M^*) = 0$.

Now, the sample specific multi-kernel based metric SSMK-M in Eq.10 can be redefined as follows:

$$\max_{M^*} Tr\left(\frac{M^{*T} S_{intra} M^*}{M^{*T} S_{intra-new} M^*}\right)$$  \hspace{1cm} (15)
Finally, Eq.14 can be solved using eigen-value method of [54], to obtain the global SSMK-M M* for re-identification. In algorithm 1, the procedure for learning SSMK and SSMK-M is provided in detail.

**Re-Identification for Test Set**

SSMK space denoted as $\beta_j$ is also learned for each sample $j$ in the test set.

However, since learning SSMK requires labeled pair of samples belonging to the same person. Therefore, before learning SSMK space for each test sample, it is required to establish an initial label identity in an unsupervised way, such that the actual identity may not be revealed before performing re-identification. In Fig.3b, initially, a multi-kernel space referred as $\beta_j$ is first obtained for a given test probe only, by averaging the kernel weights of top-$u$ KNN samples of the given test probe, which are obtained from the corresponding view of the training set. Then, the learned $\beta_j$ is used to perform initial re-identification of the test probe to obtain top-$r$ potential matching candidates from the gallery set, as shown in SSMK learning for Probe-Gallery part in Fig.3b. These top-$r$ potential gallery candidates are now served as $r$ potential probe-gallery pairs to learn a labelled, averaged pair wise SSMK space. Finally, after obtaining SSMK space for each probe-gallery pair, the final re-identification can now be performed to acquire correct matching.

**Initialization of the Multi-Kernel Space for Each Query.** For a given test probe $x_j^a$, initially the $N$ kinds of features are projected into $K$ different kernel spaces. As it is required to learn SSMK space for probe $x_j^a$ without knowing its identity label, an unsupervised KNN approach is used to obtain its initial multi-kernel space $\beta_j$ . KNN approach uses the learned SSMK-M metric $M^*$ to find similar samples from the corresponding view in training set for each projected feature. Since there are $m$ number of persons in the gallery set, and each has its unique attributes, to eliminate outliers and select only those KNN samples that possess similar characteristic attributes with $x_j^a$, only top-$u$ KNN samples are selected and used to obtain $\beta_j$ of $x_j^a$. Average weight $\alpha_{j,\Phi_k}$ for each $\Phi_k$ is then computed by averaging over all top-$u$ samples for each $f_n$, as:

$$\alpha_{j,\Phi_k} = \frac{\sum_{f_n} \sum_{i=1}^u \alpha_{j,f_n,\Phi_k,i,r}}{N \times u},$$

(16)

Where $\alpha_{j,\Phi_k}$ is the average weight of $\Phi_k$ for probe $x_j^a$, while $\alpha_{j,f_n,\Phi_k}$ is the actual weight of $\Phi_k$ in each top-$u$ sample $i'$ for each $f_n$. Finally, multi-kernel space $\beta_j$ of probe $x_j^a$ is initialized as,

$$\beta_j = (\alpha_{j,\Phi_k}),$$

(17)

**Sample Specific Multi-Kernel Space for Test Probe-Gallery Pair.** Now, the computed $\beta_j$ can be used for initial matching between probe and gallery set, by projecting each gallery sample into the $\beta_j$ space of $x_j^a$. Which then generating a ranking list, that contains potential matching gallery samples arranged in ascending order of similarity. Since the label information is not available, the initially learned space $\beta_j$ for probe $x_j^a$ is not much reliable. Therefore, top-$r$ potential matching gallery samples from the ranking list are chosen as potential labeled samples, without knowing their identity labels. These, top-$r$ gallery samples and $x_j^a$ are then used to perform potential labeled pairs to learn a more robust, paired and reliable SSMK space for $x_j^a$ and gallery set, which will be used to perform final re-identification.

Forming each pair $(x_j^a, x_j^b, j',_{j=1...r})$, SSMK is then learned using the algorithm 1 in section 3.2. In total, there would be $r$ SSMK spaces are obtained for each $(x_j^a, x_j^b)$. Therefore, an average weight $\alpha_{j,\Phi_k}$ is computed for each kernel $\Phi_k$ using all the $r$ weights to form the final averaged SSMK space as:
And then the sample specific multi-kernel space $\beta_j$ for $x^a$ and gallery set is obtained as:

$$\beta_j = (\alpha_{j,\phi_1}, \Phi_k),$$

(19)

To our best knowledge, this is the very first work that learns the sample specific multi-kernel space for each test probe-gallery pair.

**Re-identification.** Finally, using $\beta_j$, re-identification can be performed as,

$$\text{Sim} = (\beta_j(x_j^a) - \beta_j(x_j^b))^T M^* (\beta_j(x_j^a) - \beta_j(x_j^b))$$

(20)

Where $M^*$ is the learned SSMK-M metric from section 3.3. In algorithm 2, the procedure of learning SSMK space $\beta_j$ using an incremental approach and then re-identifying a probe using $M^*$ is provided.

**Experiments**

**Datasets and Experimental Setup**

Three popular benchmark datasets including VIPeR, GRID and CUHK01, are used for the evaluation. VIPeR is a challenging two views dataset, with a total of 632 persons. GRID is also a two views dataset, where 250 persons are viewed in both camera views, while additional 775 persons are only viewed in single camera view. As a multi-shot dataset, CUHK01 captures 971 persons from two disjoint views. Each view contains two images per person, and thus there are total 3,884 images.

In the experiment, all the datasets are randomly partitioned into training and test sets respectively. Half split is randomly chosen for training, while the rest are used to form gallery set. GRID is an imbalanced dataset, thus, 125 persons viewed in both views are randomly chosen for training, and the remaining 125 persons together with additional 775 persons are used as the gallery set. Similar to the setting of single shot, CUHK01 is also randomly partitioned into training and test sets. For test set, 486 persons are randomly selected. In CUHK01, each person has already two samples per view, and hence, there is no need for random sampling of samples to form gallery set.

**Experimental Result Analysis**

VIPeR is one of the challenging datasets in re-identification, the results on VIPeR are obtained after averaging the results of twenty trials and then compared with [15, 20, 33, 57]. The ranking result are shown in CMC curve in Fig.5 and comparison table is provided in Table.1. In Fig.5, it is clear that the learned SSMK-M using weighted Polynomial kernels (referred as SSMK-M(Poly)) and
weighted RBF kernels (referred as SSMK-M(RBF)) demonstrated successful re-identification performance from rank-1 to rank-20. The learned metric based on weighted multi-kernels (referred as SSMK-M(Opt)) have also significantly outperformed several state of the art approaches including [20, 22, 32, 33]. Although rank-1 results of SSMK-M(Poly), SSMK-M(RBF) and SSMK-M(Opt) are lower than [56], the proposed SSMK based metrics have obtained much higher re-identification results at rank-5 and afterwards. In Fig.5, even with large view, posture and illumination variations among samples, learned SSMK-M is robust enough so that it can correctly match the probe with correct the gallery image.

CMC curve and comparison table for GRID are provided in Fig.6 and Table.2. Both graphical and analytical results of SSMK-M(Poly), SSMK-M(RBF) and SSMK-M(Opt) have demonstrated successful re-identification performance compared to several state of the art approaches, including [17, 33, 55, 58]. Although, proposed SSMK-M has lower rank-1 result compared to [4, 15, 21], SSMK-M(Poly), SSMK-M(RBF) and SSMK-M(Opt) obtained much higher reidentification performance at rank-5 and afterwards, showing its discriminant capabilities against drastically varying postures and illumination in GRID.
Table 1. Top rank comparisons with the VIPeR dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>VIPeR (P=316)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R=1</td>
</tr>
<tr>
<td>Zheng, PAMI2013 [55]</td>
<td>15.66</td>
</tr>
<tr>
<td>Bazzani, CVIU13 [22]</td>
<td>19.87</td>
</tr>
<tr>
<td>Li, CVPR13 [29]</td>
<td>29.11</td>
</tr>
<tr>
<td>Zhao, CVPR13 [20]</td>
<td>30.16</td>
</tr>
<tr>
<td>Xiong, ECCV14 [32] (K=1)</td>
<td>32.30</td>
</tr>
<tr>
<td>Xiong, ECCV14 [32] (Ensb.)</td>
<td>32.80</td>
</tr>
<tr>
<td>Chen, CVPR15 [33]</td>
<td>36.8</td>
</tr>
<tr>
<td>Wang, CVPR16 [3]</td>
<td>35.76</td>
</tr>
<tr>
<td>Zhang, CVPR16 [15]</td>
<td>42.66</td>
</tr>
<tr>
<td>Zhang, CVPR16 [57]</td>
<td>42.28</td>
</tr>
<tr>
<td>Proposed SSMK-M (Poly)</td>
<td>38.78</td>
</tr>
<tr>
<td>Proposed SSMK-M (RBF)</td>
<td>40.87</td>
</tr>
<tr>
<td>Proposed SSMK-M (Opt)</td>
<td>43.85</td>
</tr>
</tbody>
</table>

Table 2. Top rank comparisons on the GRID dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>GRID (P=125 and Gallery=900)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R=1</td>
</tr>
<tr>
<td>Zheng, PAMI2013 [55]</td>
<td>9.68</td>
</tr>
<tr>
<td>Prosser, BMVA10 [59]</td>
<td>10.2</td>
</tr>
<tr>
<td>Chen, ICPR14 [60]</td>
<td>10.68</td>
</tr>
<tr>
<td>Loy, ICIP13 [61]</td>
<td>12.2</td>
</tr>
<tr>
<td>Liao, CoRR [58]</td>
<td>15.2</td>
</tr>
<tr>
<td>Chen, CVPR15 [33]</td>
<td>16.3</td>
</tr>
<tr>
<td>Liao, CVPR15 [17]</td>
<td>16.56</td>
</tr>
<tr>
<td>Pribadi, PR16 [62]</td>
<td>15.2</td>
</tr>
<tr>
<td>Zhang, CVPR16 [15]</td>
<td>22.40</td>
</tr>
<tr>
<td>Chen, CVPR16 [4]</td>
<td>24.24</td>
</tr>
<tr>
<td>Matsukawa, CVPR16 [21]</td>
<td>24.7</td>
</tr>
<tr>
<td>Proposed SSMK-M (Poly)</td>
<td>19.4</td>
</tr>
<tr>
<td>Proposed SSMK-M (RBF)</td>
<td>19.9</td>
</tr>
<tr>
<td>Proposed SSMK-M (Opt)</td>
<td>20.4</td>
</tr>
</tbody>
</table>

SSMK-M is also evaluated multi-shot dataset CUHK01. All the three variants SSMK-M (Opt), SSMK-M (RBF) and SSMK-M (Poly) are learned and evaluated for twenty trials, and averaged results are shown in Fig.7 and Table.3. It can be observed that SSMK-M can perform much better re-identification by effectively projecting each person in a sample adaptive multi-kernel space that enhances similarity between pair of images of the same identity person. These variants have significantly outperformed several previous state-of-the-art approaches including [17, 21, 47, 56, 57]. However, SSMK-M (RBF) and SSMK-M (Poly) have lower rank-1 and rank-5 results compared to [57], which is mainly due to the complexity, non-linearity and modality in the feature space that cannot be easily modeled using only single type of weighted multi-kernel space.
Table 3. Top rank comparisons on CUHK01 dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>CUHK01(P=486)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R=1</td>
</tr>
<tr>
<td>Zhao, ICCV13 [20]</td>
<td>28.45</td>
</tr>
<tr>
<td>An, ICSVT15 [85]</td>
<td>31.10</td>
</tr>
<tr>
<td>Zhao, CVPR14 [86]</td>
<td>34.30</td>
</tr>
<tr>
<td>Ahmed, CVPR15 [47]</td>
<td>47.53</td>
</tr>
<tr>
<td>Liao, CVPR15 [17]</td>
<td>63.21</td>
</tr>
<tr>
<td>Cheng, CVPR16 [56]</td>
<td>53.7</td>
</tr>
<tr>
<td>Matsukawa, CVPR16 [21]</td>
<td>67.3</td>
</tr>
<tr>
<td>Zhang, CVPR16 [57]</td>
<td>69.09</td>
</tr>
<tr>
<td>Proposed SSMK-M (Poly)</td>
<td>67.78</td>
</tr>
<tr>
<td>Proposed SSMK-M (RBF)</td>
<td>68.60</td>
</tr>
<tr>
<td>Proposed SSMK-M (Opt)</td>
<td>69.2</td>
</tr>
</tbody>
</table>

Conclusion

Images in disjoint camera views possess large appearance variations, leading to the failure in finding correct matches in person re-identification. In this paper, the proposed Sample Specific Multi-Kernel (SSMK) approach addresses the problem of appearance variation in cross camera-views images, by adaptively projecting features into the weighted multi-kernel feature space, which is then used to learn a more robust and discriminative metric for re-identification. The obtained results on single-shot and multi-shot datasets have demonstrated the effectiveness of projecting features into SSMK and then learning metric SSMK-M for person re-identification.

References


[38] M. Ring, B. Eskofier, An approximation of the Gaussian RBF kernel for efficient classification with SVMs, Pattern Recognition Letters 84.


[74] A. Franco, L. Oliveira, Convolutional covariance features: Conception, integration and performance in person re-identification, Pattern Recognition (2016) –.


