Advances in Engineering Research, volume 185
2nd International Conference on Electrical and Electronic Engineering (EEE 2019)

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

Jun FANG[1], Ru-fei ZHANG[2] and Feng JIANG[2]
[1]University of Chinese Academy of Sciences, Beijing, China
[2]Beijing 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.

Figure 1. Images from VIPeR dataset in two views (each column contains images of the same person)

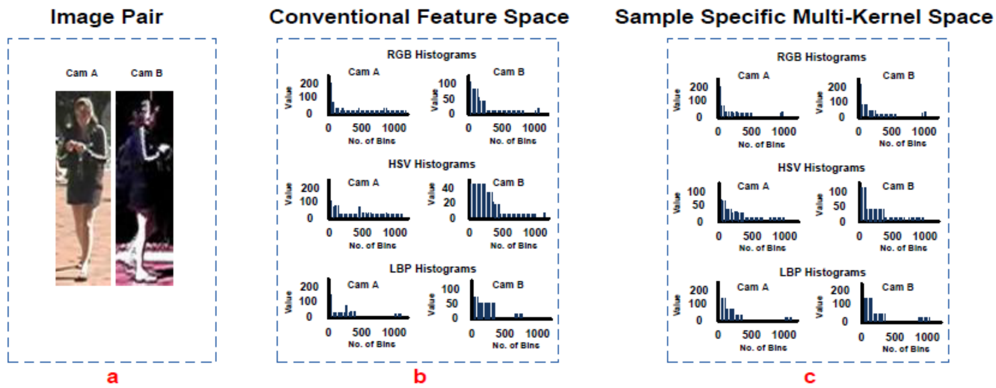

Figure 2. Feature Space of Pair of images (a) Pair of images from VIPeR, (b) Conventional Feature Space after feature extraction, (c) Feature Space after Projection via Sample Specific Multi-Kernel

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 β_j^j is first learned for only query probe through a KNN approach, as shown in Fig.3b. Then with the learned β_j^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 β_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:

$$F = \{f_n, n = 1, ..., N\} \tag{1}$$

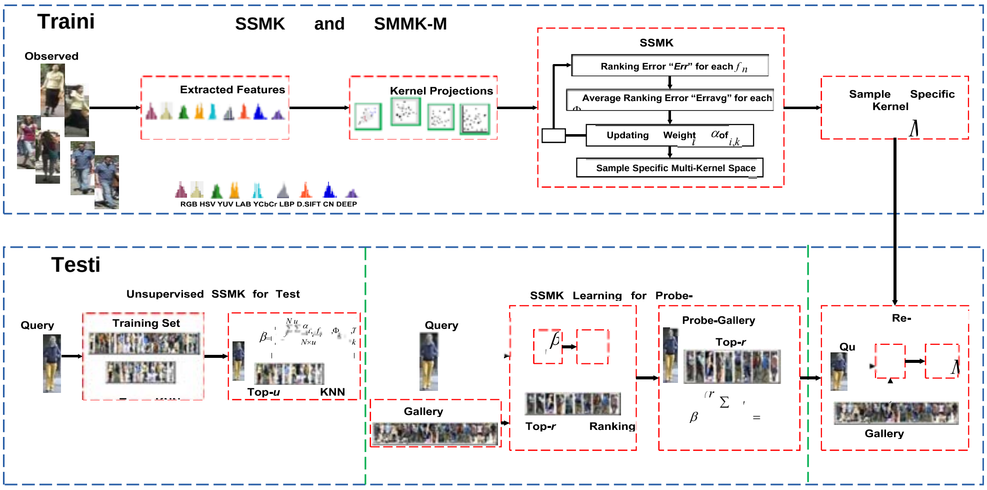

Figure 3. Methodology of the proposed Sample Specific Multi-Kernels and Metric SSMK-M along with Re-identification Approach

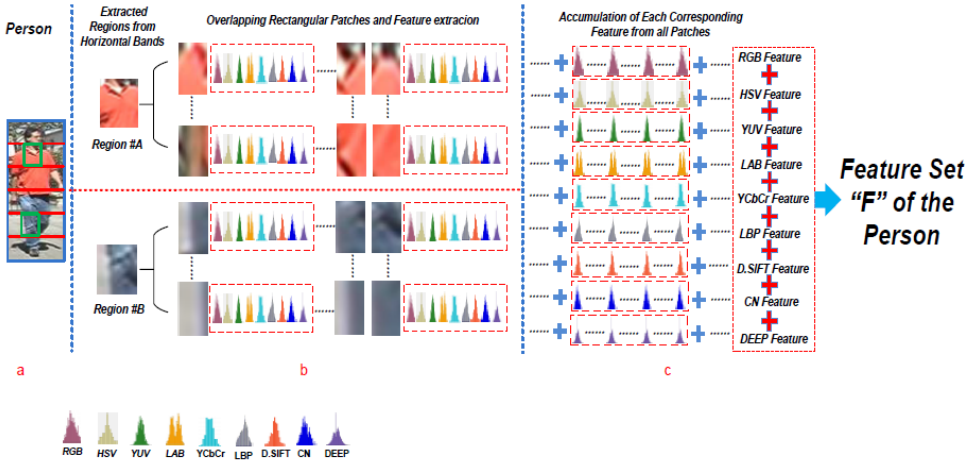

Figure 4. Feature Extraction (a) Image Pair from VIPeR, (b) Extract Patches Features, (c) Accumulate Patch Features to form F

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:

$$K_{rbf}(x_i^a, x_i^b) = \exp^{(-\frac{\|x_i^a - x_i^b\|^2}{2\alpha^2})}, K_{chi^2}(x_i^a, x_i^b) = \exp^{(-\alpha\sum_i \frac{(x_i^a - x_i^b)^2}{x_i^a + x_i^b})}. \quad (2)$$

Where x_i^a and x_i^b are samples of person i in different views. Linear kernels including linear and poly-nominal is used, which are defined as:

$$K_l(x_i^a, x_i^b) = <x_i^a, x_i^b>, \; K_p(x_i^a, x_i^b) = (<x_i^a, x_i^b> + 1)^m, m \in N. \quad (3)$$

All these kernels are defined as Φ follows:

$$\Phi = \{\Phi_k, k = 1, \ldots, K\} \quad (4)$$

Where K is the number of kernels and Φ_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 Φ_k for each person i with its image pairs (x_i^a, x_i^b) from different camera view. To learn the weight of Φ_k, its ranking error *Err* is obtained firstly, which is defined as:

$$Err_{i,f_n,\Phi_{k,t}} = \frac{d_{rank1}(\Phi_k(x_{i,f_n}^a), \Phi_k(x_{rank1,i,f_n}^b))}{d(\Phi_k(x_{i,f_n}^a), \Phi_k(x_{i,f_n}^b))} \tag{5}$$

Where x_{i,f_n}^a represents the feature f_n for sample x_i^a and $\Phi_k(x_{i,f_n}^a)$ is its kernel projection. In Eq.5,

ranking error $Err_{i,f_n,\Phi_{k,j}}$ in the *t*-th iteration is computed using the ratio between similarity values.

$d_{rank1}()$ gives similarity between x_i^a and the gallery sample x_{rank1,i,f_n}^b of which are ranked as top-1

match. While measures similarity between actual pair of samples (x_i^a, x_i^b) . Since, N features are used,

the average ranking error denoted as Erravg for kernel Φ_k is given as:

$$Erravg_{i,\Phi_k,t} = \frac{\sum_{n=1}^{N} Err_{i,f_n,\Phi_{k,t}}}{N} \tag{6}$$

Where $Erravg_{i,\Phi_k,t}$ is the average ranking error of Φ_k in *t*-th iterations? Using $Erravg_{i,\Phi_k,t}$ the weight α_{Φ_k} of Φ_k in *t*-th iteration is then computed as:

$$\alpha_{i,\Phi,t} = \exp^{-Erravg_{i,\Phi_k,t}} \times w_{i,\Phi_k} \tag{7}$$

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

$$\beta_{i,j} = (\alpha_{i,\Phi_{k,t}}, \Phi_k), \tag{8}$$

From the above equations, it is clear that β_i is dependent on $Erravg$ 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 $Erravg$ 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:

$$\beta_i = (\alpha_{i,\Phi_{k,T}}, \Phi_k) \tag{9}$$

Sample Specific Multi-Kernel Based Metric Learning

Since sample specific multi-kernel space β_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.

$$\max_{M^*} Tr(\frac{M^{*T} S_{inter} M^*}{M^{*T} S_{intra} M^*}) \tag{10}$$

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* β_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,

$$S^*_{intra} = \gamma S^i + (1-\gamma) S_{intra} \tag{11}$$

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. γ ($0 < \gamma < 1$) is a factor that controls the balance between pair wise co-variance S^j 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 $A = [a_{ij}]$ by measuring the distance between pair of samples. Each element a_{ij} in the affinity matrix A is then given as:

$$\alpha_{ij} = \exp^{(-\frac{d_p(x_i^a, x_i^b)}{\sigma_i^a \sigma_i^b})} \tag{12}$$

Where α_{ij} represents the affinity between samples x_i^a and x_i^b, using the d_p (p-norm distance). While σ_i^a and σ_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:

$$S^*_{intra-new} = \gamma A S' + (1-\gamma) A S_{intra}, \tag{13}$$

$$S_{inter-new} = A S_{inter}. \tag{14}$$

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 $w_{\Phi_k}(\Phi_k)$ (usually set to $1/K$ where K is the total number of Kernels in set Φ).
for t=1, ..., T **do**
  for i=1, ..., n' **do**
    for Φ=1, ..., K **do**
      for F=1, ..., N **do**
        compute $Err_{i,f_n,\Phi_k,t} = \frac{d_{rank1}(\Phi_k(x^a_{i,f_n}), \Phi_k(x^{tb}_{i,f_n}))}{d(\Phi_k(x^a_{i,f_n}), \Phi_k(x^b_{i,f_n}))}$
          where $d(.) = (x_i - x_j)^T M (x_i - x_j)$
      end for
      compute $Erravg_{i,\Phi_k,t} = \sum_{n=1}^{N} Err_{i,f_n,\Phi_k,t}$
      update $\alpha_{i,\Phi_k,t} = exp(-Erravg_{i,\Phi_k,t}) \times w_{i,\Phi_k}(\Phi_k)$
        where $w_{i,\Phi_k}(\Phi_k)$ is the initial weight of Φ_k
    end for
    compute $\beta_{i,t} = (\alpha_{i,\Phi_k,t}, \Phi_k)$
      where $\beta_{i,t}$ is the sample specific multi-kernel space of person i
  end for
  compute $max_M Tr(\frac{M^T \beta_i (S_{inter-new}) M}{M^T \beta_i (S^*_{intra-new}) M})$
    where M is used in next iteration
end for
Output Final ssmk $\beta_i = (\alpha_{i,\Phi_k,T}, \Phi_k)$ and
$max_{M^*} Tr(\frac{M^{*T} \beta (S_{inter-new}) M^*}{M^{*T} \beta (S^*_{intra-new}) M^*}) = 0$

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

$$\max_{M^*} Tr(\frac{M^{*T} S_{inter} M^*}{M^{*T} S_{intra} M^*}) \tag{15}$$

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 β_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 β_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 β_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_i^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_i^a without knowing its identity label, an unsupervised KNN approach is used to obtain its initial multi-kernel space β_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 β_j' of x_j^a . Average weight α_{j,Φ_k}' for each Φ_k is then computed by averaging over all top-*u* samples for each f_n, as:

$$\alpha'_{j,\Phi_k} = \frac{\sum_{f_{n=1}}^{N} \sum_{i'=1}^{u} \alpha_{i^{i},f_n,\Phi_k,T}}{N \times u}, \tag{16}$$

Where α_{j,Φ_k}' is the average weight of Φ_k for probe x_j^a, while α_{j,Φ_k}' is the actual weight of Φ_k in each top-*u* sample i' for each f_n . Finally, multi-kernel space β_j' of probe x_j^a is initialized as,

$$\beta'_j = (\alpha'_{j,\Phi_k,} \Phi_k), \tag{17}$$

Sample Specific Multi-Kernel Space for Test Probe-Gallery Pair. Now, the computed β_j' can be used for initial matching between probe and gallery set, by projecting each gallery sample into the β_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 β_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 $(x_j^a, x_{j',j'=1\ldots r}^b)$, 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 α_{j,Φ_k} is computed for each kernel Φ_k using all the *r* weights to form the final averaged SSMK space as:

$$\alpha_{j,\ \Phi_k} = \frac{\sum_{j'}^{r} \alpha_{j',\Phi_k}}{r}, \tag{18}$$

And then the sample specific multi-kernel space β_j for x^a and gallery set is obtained as:

$$\beta_j = (\alpha_{j,\Phi_k}, \Phi_k), \tag{19}$$

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 β_j, re-identification can be performed as,

$$Sim = (\beta_j(x_j^a) - \beta_j(x_j^b))^T M^* (\beta_j(x_j^a) - \beta_j(x_j^b)) \tag{20}$$

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

Algorithm 2 Re-Identification after Learning Unsupervised SSMK

Given: $F = [f_1, f_2, \ldots, f_N]$, $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_K]$, $X_j = [x^a_{j,j=1,\ldots,n}]$, $G = [x^b_{j,j=1,\ldots,n}]$
***SSMK* for Test Probe:**
compute $\alpha'_{j,\Phi_k} = \frac{\sum_{fn=1}^{N} \sum_{i'=1}^{u} \alpha_{i',fn,\Phi_k,T}}{N \times u}$
  where α'_{j,Φ_k} is the weight of each Φ_k of probe x^a_j, computed using top-u KNN samples of each f_n
compute $\beta'_j = (\alpha'_{j,\Phi_k}, \Phi_k)$
***SSMK* for Probe-Gallery Set:**
Given: Top-r potential Gallery matches
Repeat Algorithm 1 for each $(x^a_j, x^b_{j',j'=1,\ldots,r})$ pair
compute Average weight α_{j,Φ_k} of each Φ_k, using all r pairs, as, $\alpha_{j,\Phi_k} = \frac{\sum_{j'=1}^{r} \alpha_{j',\Phi_k}}{r}$
compute Probe-Gallery SSMK β_j using α_{j,Φ_k}, as $\beta_j = (\alpha_{j,\Phi_k}, \Phi_k)$
Re-Identification:
compute $Sim = (\beta_j(x^a_j) - \beta_j(x^b_j))^T M^* (\beta_j(x^a_j) - \beta_j(x^b_j))$ =0

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.

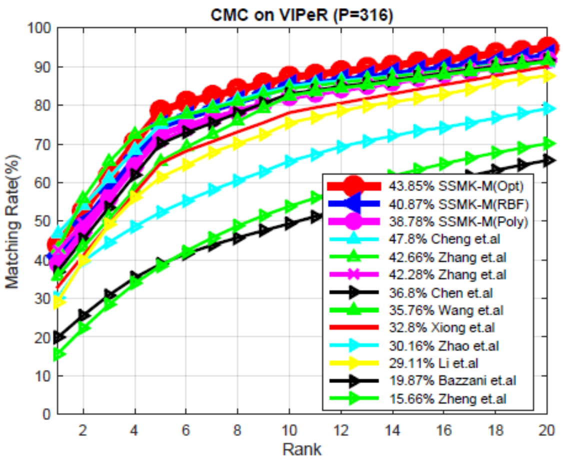

Figure 5. Comparison of CMC curves (VIPeR, *P*=316)

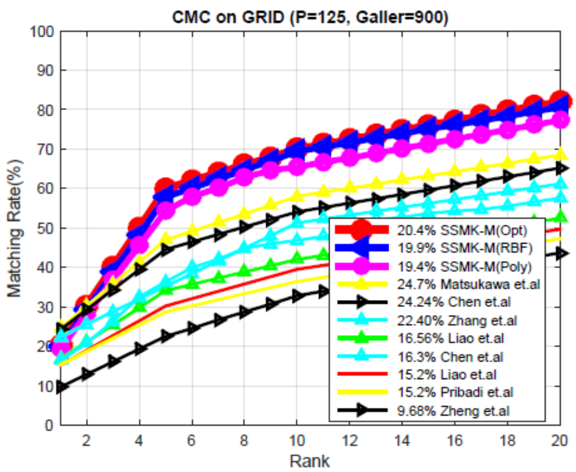

Figure 6. Comparison of CMC curves (GRID, P=125 and Gallery Set=900)

Table 1. Top rank comparisons wd the VIPeR dataset

Method	**VIPeR(P=316)**				
	R=1	**R=5**	**R=10**	**R=20**	**R=50**
Zheng, PAMI2013 [55]	15.66	38.42	53.86	70.09	80.00
Bazzani, CVIU13 [22]	19.87	38.89	49.37	65.73	84.84
Li, CVPR13 [29]	29.11	61.39	75.63	87.66	97.15
Zhao, CVPR13 [20]	30.16	52.31	65.54	79.15	91.49
Xiong, ECCV14[32] (K=1)	32.30	65.80	79.70	90.9	96.51
Xiong, ECCV14[32] (Ensb.)	32.80	65.50	79.10	90.0	97.46
Chen, CVPR15 [33]	36.8	70.4	83.7	91.7	97.8
Wang, CVPR16 [3]	35.76	66.6	82.8	-	-
Cheng, CVPR16 [56]	47.8	74.7	84.8	91.1	-
Zhang, CVPR16 [15]	42.66	-	84.27	91.93	-
Zhang, CVPR16 [57]	42.28	71.46	82.94	92.06	-
Proposed SSMK-M (Poly)	38.78	72.9	82.1	91.8	97.87
Proposed SSMK-M (RBF)	40.87	74.72	84.9	92.78	98.15
Proposed SSMK-M (Opt)	43.85	79.19	87.2	94.9	99.23

Table 2. Top rank comparisons on the GRID dataset

Method	**GRID (P=125 and Gallery=900)**				
	R=1	**R=5**	**R=10**	**R=20**	**R=50**
Zheng, PAMI2013 [55]	9.68	22.00	32.96	44.32	-
Prosser, BMVA10 [59]	10.2	24.6	33.3	43.7	-
Chen, ICPR14 [60]	10.68	25.76	35.04	46.48	-
Loy, ICIP13 [61]	12.2	27.8	36.3	46.6	-
Liao, CoRR [58]	15.2	30.1	39.2	49.3	-
Chen, CVPR15 [33]	16.3	35.8	46.0	57.6	-
Liao, CVPR15 [17]	16.56	33.84	41.84	52.40	-
Pribadi, PR16 [62]	15.2	28.40	36	46.4	-
Zhang, CVPR16 [15]	22.40	-	51.28	61.20	-
Chen, CVPR16 [4]	24.24	44.56	54.08	65.20	-
Matsukawa, CVPR16 [21]	24.7	47.0	58.4	69.0	-
Proposed SSMK-M (Poly)	19.4	53.79	65.42	77.09	88.01
Proposed SSMK-M (RBF)	19.9	57.88	69.09	80.36	90.15
Proposed SSMK-M (Opt)	20.4	59.41	69.89	81.96	91.02

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.

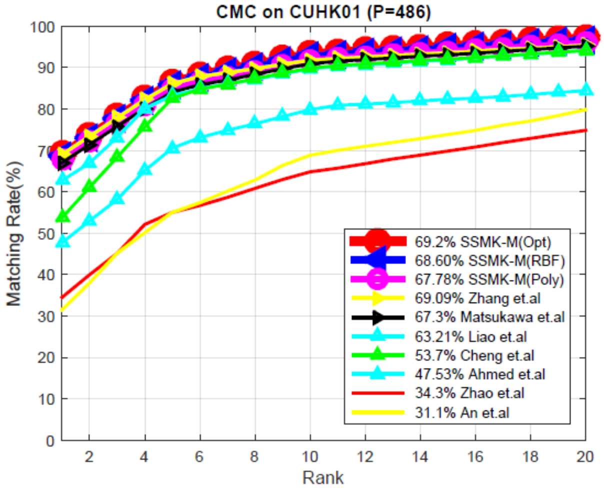

Figure 7. Comparison of CMC curves (CUHK01, *P*=486)

Table 3. Top rank comparisons on CUHK01 dataset

Method	CUHK01(P=486)				
	R=1	**R=5**	**R=10**	**R=20**	**R=50**
Zhao, ICCV13 [20]	28.45	45.85	55.67	67.95	-
An, ICSVT15 [85]	31.10	-	68.55	79.18	-
Zhao, CVPR14 [86]	34.30	55.06	64.96	74.94	-
Ahmed, CVPR15 [47]	47.53	71.2	80.02	86.6	-
Liao, CVPR15 [17]	63.21	83.89	90.04	94.16	
Cheng, CVPR16 [56]	53.7	84.3	91.0	96.3	99
Matsukawa, CVPR16 [21]	67.3	86.9	91.8	95.9	-
Zhang, CVPR16 [57]	69.09	86.87	91.77	95.39	-
Proposed SSMK-M (Poly)	67.78	85.21	92.46	95.9	99.25
Proposed SSMK-M (RBF)	68.60	86.45	92.84	96.71	99.50
Proposed SSMK-M (Opt)	69.2	87.78	93.19	97.06	99.82

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

[1] A. Bedagkar-Gala, S. K. Shah, A survey of approaches and trends in person re-identification, Image and Vision Computing 32 (4) (2014) 270 – 286.

[2] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, J. Wang, Person re-identification with correspondence structure learning, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3200–3208.

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

[4] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[5] Y.-C. Chen, W.-S. Zheng, J. Lai, Mirror representation for modeling view-specific transform in person re-identification (2015).

[6] Y. C. Chen, W. S. Zheng, J. H. Lai, P. Yuen, An asymmetric distance model for cross-view feature mapping in person re-identification, IEEE Transactions on Circuits and Systems for Video Technology PP (99) (2016) 1–1.

[7] S. Li, M. Shao, Y. Fu, Cross-view projective dictionary learning for person re-identification, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, 2015, pp. 2155–2161.

[8] S. Karanam, Y. Li, R. J. Radke, Person re-identification with discriminatively trained viewpoint invariant dictionaries, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4516–4524.

[9] M. A. Syed, J. Jiao, Multi-kernel metric learning for person re-identification, in: ICIP, 2016, pp. 784–788.

[10] D. Gray, H. Tao, Computer Vision – ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I, 2008, Ch. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, pp. 262–275.

[11] I. Kviatkovsky, A. Adam, E. Rivlin, Color invariants for person reidentification, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1622–1634.

[12] L. An, S. Yang, B. Bhanu, Person re-identification by robust canonical correlation analysis, IEEE Signal Processing Letters 22 (8) (2015) 1103–1107.

[13] A. Datta, L. M. Brown, R. Feris, S. Pankanti, Appearance modeling for person re-identification using weighted brightness transfer functions, in: Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 2367–2370.

[14] K. Liu, Z. Zhao, A. Cai, Parametric local multi-modal metric learning for person re-identification, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 2578–2583.

[15] Y. Zhang, B. Li, H. Lu, A. Irie, X. Ruan, Sample-specific svm learning for person re-identification, in: The IEEE Confer- ence on Computer Vision and Pattern Recognition (CVPR), 2016.

[16] M. Gönen, E. Alpaydın, Localized algorithms for multiple kernel learning, Pattern Recognition 46 (3, 2013) (2013) 795– 807.

[17] S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 2015, pp. 2197–2206.

[18] V. Shet, S. Khamis, C.-H. Kuo, Person re-identification using semantic color names and rankboost, in: WACV, 2013, pp. 281–287.

[19] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S. Z. Li, Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I, Springer International Publishing, 2014, Ch. Salient Color Names for Person Re-identification, pp. 536–551.

[20] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, in: Computer Vision (ICCV), 2013 IEEE International Conference on, 2013, pp. 2528–2535.

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

[22] L. Bazzani, M. Cristani, V. Murino, Symmetry-driven accumulation of local features for human characterization and re-identification, Computer Vision and Image Understanding 117 (2, 2013) 130–144.

[23] B. Ma, Q. Li, H. Chang, Computer Vision - ACCV 2014 Workshops: Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part III, Springer International Publishing, 2015, Ch. Gaussian Descriptor Based on Local Features for Person Re-identification, pp. 505–518.

[24] Y. Hu, S. Liao, Z. Lei, D. Yi, S. Z. Li, Exploring structural information and fusing multiple features for person re- identification, in: CVPR, 2013, pp. 794–799.

[25] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, W. Gao, Multi-task learning with low rank attribute embedding for person re-identification, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3739–3747.

[26] A. Li, L. Liu, K. Wang, S. Liu, S. Yan, Clothing attributes assisted person reidentification, IEEE Transactions on Circuits and Systems for Video Technology 25 (5) (2015) 869–878.

[27] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[28] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, Y. Y. Tang, Person re-identification by dual-regularized kiss metric learning, IEEE Transactions on Image Processing 25 (6) (2016) 2726–2738.

[29] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, J. R. Smith, Learning locally-adaptive decision functions for person verification, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 3610–3617.

[30] J. Chen, Z. Zhang, Y. Wang, Relevance metric learning for person re-identification by exploiting listwise similarities, IEEE Transactions on Image Processing 24 (12) (2015) 4741–4755.

[31] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 3318–3325.

[32] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: Com- puter Vision–ECCV 2014, Springer, 2014, pp. 1–16.

[33] D. Chen, Z. Yuan, G. Hua, N. Zheng, J. Wang, Similarity learning on an explicit polynomial kernel feature map for person re-identification, in: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, 2015, pp. 1565–1573.

[34] S. Yan, X. Xu, D. Xu, S. Lin, X. Li, Beyond spatial pyramids: A new feature extraction framework with dense spatial sampling for image classification, in: Proceedings of the 12th European Conference on Computer Vision - Volume Part IV, ECCV'12, 2012, pp. 473–487.

[35] A. Vahdat, K. Cannons, G. Mori, I. Kim, S. Oh, Compositional models for video event detection: A multiple kernel learning latent variable approach, in: International Conference on Computer Vision (ICCV), 2013.

[36] F. Yan, K. Mikolajczyk, M. Barnard, H. Cai, J. Kittler, lp norm multiple kernel fisher discriminant analysis for object and image categorisation, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 3626–3632.

[37] W. Yang, Y. Wang, A. Vahdat, G. Mori, Kernel latent svm for visual recognition, in: Advances in Neural Information Processing Systems (NIPS), 2012.

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

[39] B. Mocanu, R. Tapu, T. Zaharia, Using Computer Vision to See, 2016, pp. 375–390.

[40] W. S. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 649–656.

[41] A. Mignon, F. Jurie, Pcca: A new approach for distance learning from sparse pairwise constraints, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2666–2672.

[42] B. Fernando, E. Fromont, D. Muselet, M. Sebban, Discriminative feature fusion for image classification, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 3434–3441.

[43] Y. Hu, S. Liao, Z. Lei, D. Yi, S. Z. Li, Exploring structural information and fusing multiple features for person re- identification, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 794–799.

[44] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2360–2367.

[45] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 3586–3593.

[46] L. An, M. Kafai, S. Yang, B. Bhanu, Reference-based person re-identification, in: AVSS, 2013, pp. 244–249.

[47] E. Ahmed, M. Jones, T. K. Marks, An improved deep learning architecture for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[48] Y. Han, K. Yang, Y. Ma, G. Liu, Localized multiple kernel learning via sample-wise alternating optimization, IEEE Transactions on Cybernetics 44 (1) (2014) 137–148.

[49] X. Liu, L. Wang, J. Zhang, J. Yin, Sample-adaptive multiple kernel learning, 2014.

[50] D. Kong, C. H. Q. Ding, Pairwise-covariance linear discriminant analysis, in: AAAI, 2014, pp. 1925–1931.

[51] X. He, Locality preserving projections, Ph.D. thesis, Chicago, IL, USA, aAI3195015 (2005).

[52] H. Wang, F. Nie, H. Huang, learning robust locality preserving projection via p-order minimization, in: AAAI, 2015, pp. 3059–3065.

[53] L. Zelnik-manor, P. Perona, Self-tuning spectral clustering, in: NIPS, 2004, pp. 1601–1608. [54] Y. Ying, P. Li, S. Sonnenburg, F. Bach, C. S. Ong, Distance metric learning with eigenvalue optimization, in: JMLR, 2012, pp. 1–26.

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

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

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

[58] S. Liao, Y. Hu, S. Z. Li, Joint dimension reduction and metric learning for person re-identification, CoRR abs/1406.4216.

[59] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by support vector ranking, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2010, pp. 21.1–21.11.

[60] J. Chen, Z. Zhang, Y. Wang, Relevance metric learning for person re-identification by exploiting global similarities, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 1657–1662.

[61] C. C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, in: Image Processing (ICIP), 2013 20th IEEE International Conference on, 2013, pp. 3567–3571.

[62] R. C. H. Pribadi, H.-K. Pao, Sparse tree structured representation for re-identification, Pattern Recognition 60 (2016) 394– 404.

[63] W. Li, Y. Wu, J. Li, Re-identification by neighborhood structure metric learning, Pattern Recognition 61 (2017) 327 – 338.

[64] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2288–2295.

[65] M. Dikmen, E. Akbas, T. S. Huang, N. Ahuja, Computer Vision – ACCV 2010: 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part IV, Springer Berlin Heidelberg, 2011, Ch. Pedestrian Recognition with a Learned Metric, pp. 501–512.

[66] N. Martinel, C. Micheloni, G. L. Foresti, Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III, Springer International Publishing, 2015, Ch. Saliency Weighted Features for Person Re-identification, pp. 191–208.

[67] Y. Li, Z. Wu, R. J. Radke, Multi-shot re-identification with random-projection-based random forests, in: Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on, 2015, pp. 373–380.

[68] N. Martinel, C. Micheloni, G. L. Foresti, Kernelized saliency-based person re-identification through multiple metric learn- ing, IEEE Transactions on Image Processing 24 (12) (2015) 5645–5658.

[69] J. Garcia, N. Martinel, G. L. Foresti, A. Gardel, C. Micheloni, Person orientation and feature distances boost re- identification, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 4618–4623.

[70] Y.-J. Cho, K.-J. Yoon, Improving person re-identification via pose-aware multi-shot matching, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[71] T. Matsukawa, T. Okabe, Y. Sato, Person re-identification via discriminative accumulation of local features, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 3975–3980.

[72] M. Zeng, Z. Wu, C. Tian, L. Zhang, L. Hu, Efficient person re-identification by hybrid spatiogram and covariance descriptor, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on, 2015, pp. 48–56.

[73] F. Wang, X. Wang, D. Zhang, C. Zhang, T. Li, marginface: A novel face recognition method by average neighborhood margin maximization, Pattern Recognition 42 (11) (2009) 2863 – 2875.

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

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

[76] L. Bazzani, M. Cristani, A. Perina, V. Murino, Multiple-shot person re-identification by chromatic and epitomic analyses, Pattern Recognition Letters 33 (7) (2012) 898 – 903.

[77] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, in: British Machine Vision Conference (BMVC), 2011.

[78] S. Bak, G. Charpiat, E. Corvee, F. Bremond, M. Thonnat, Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III, Springer Berlin Heidelberg, 2012, Ch. Learning to Match Appearances by Correlations in a Covariance Metric Space, pp. 806–820.

[79] S. Bak, E. Corvee, F. Bremond, M. Thonnat, Multiple-shot human re-identification by mean riemannian covariance grid, in: Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on, 2011, pp. 179– 184.

[80] J. You, A. Wu, X. Li, W.-S. Zheng, Top-push video-based person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[81] N. McLaughlin, J. Martinez del Rincon, P. Miller, Recurrent convolutional network for video-based person reidentification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[82] Y. Li, Z. Wu, S. Karanam, R. J. Radke, Multi-shot human re-identification using adaptive fisher discriminant analysis, in: Proceedings of the British Machine Vision Conference (BMVC), BMVA Press, 2015, pp. 73.1–73.12

[83] T. Wang, S. Gong, X. Zhu, S. Wang, Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV, Springer International Publishing, 2014, Ch. Person Re-identification by

Video Ranking, pp. 688–703.

[84] S. Karanam, Y. Li, R. J. Radke, Sparse re-id: Block sparsity for person re-identification, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on, 2015, pp. 33–40.

[85] L. An, M. Kafai, S. Yang, B. Bhanu, Person re-identification with reference descriptor, IEEE Transactions on Circuits and Systems for Video Technology PP (99) (2015) 1–1.

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