Person Re-Identification Using Multiple Experts with Random Subspaces

Sai Bi, Guanbin Li, and Yizhou Yu
Department of Computer Science, the University of Hong Kong, Hong Kong
Email: {fsbi, gbli, yzyug}@cs.hku.hk

***Abstract*—This paper presents a simple and effective multi-expert approach based on random subspaces for person re-identification across non-overlapping camera views. This approach applies to supervised learning methods that learn a continuous decision function. Our proposed method trains a group of expert functions, each of which is only exposed to a random subset of the input features. Each expert function produces an opinion according to the partial features it has. We also introduce weighted fusion schemes to effectively combine the opinions of multiple expert functions together to form a global view. Thus our method overall still makes use of all features without losing much information they carry. Yet each individual expert function can be trained efficiently without overfitting. We have tested our method on the VIPeR, ETHZ, and CAVIAR4REID datasets, and the results demonstrate that our method is able to significantly improve the performance of existing state-of-the-art techniques.**

***Index Terms*—person re-identification, multi-experts, random subspaces**

I. Introduction

Person re-identification is the problem of identifying the same person that appears in non-overlapping cameras. It can find applications in modern surveillance systems, either for online tracking of an individual over a network of cameras or offline retrieval of all videos containing a person of interest. This problem is challenging because when matching images of the same person captured with non-overlapping cameras, there may exist huge discrepancies in terms of human poses, illumination, camera views and photometric settings, and so on. In addition, the lack of sufficient resolution in surveillance cameras makes it infeasible to identify a person using face verification.

There are two major approaches to person re-identification, namely unsupervised matching of image features [2], [3], [4] and training a decision function to assess the similarity of features in two images. These two approaches are developed for different application scenarios. The unsupervised approach is better suited for scenarios where it is impractical to obtain training images that capture the same group of people from all cameras involved. However, whenever such training images can be obtained, the second supervised approach is more appropriate since it typically achieves a better performance.

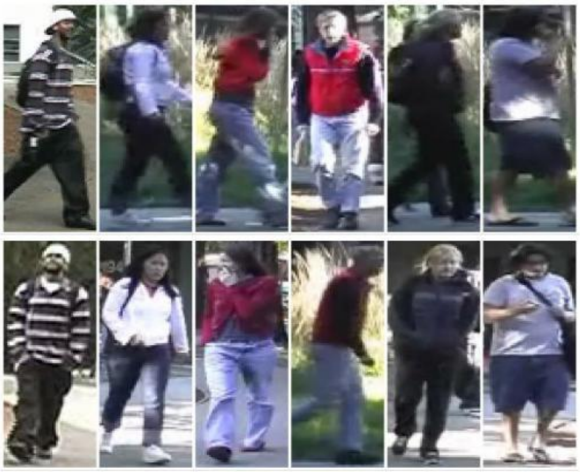

Figure 1. Sample images from the VIPeR dataset [1]. Images of the same person appear in the same column.

In this paper we focus on person re-identification based on supervised learning. For all methods in this category, highly discriminative features, such as LBP, Gabor and color histograms, are necessary to achieve good accuracy. Such feature descriptors are usually of high dimension, and to fully capture the information from one image, usually multiple features are combined, resulting in a dimension generally higher than 10k [5], [6].

High-dimensional features are critical to high performance [7], but they also give rise to various issues. For instance, to learn a Mahalanobis distance metric for feature descriptors of n dimensions, $O(n^2)$ parameters need to be optimized, making the training process both time-consuming and susceptible to local minima. More importantly, optimizing a large number of parameters amplifies the problem of overfitting due to the existence of noise and outliers in images, leading to poor generalization capability.

Therefore, dimension reduction methods such as PCA and CCA are often applied to aggressively reduce the dimension of feature descriptors. However, in this process subtle but highly discriminative information may be overlooked, especially when many different types of features are combined together, which decreases the discriminative power of the new features after dimension reduction.

Manuscript received September 9, 2014; revised December 12, 2014.

doi: 10.12720/joig.2.2.151-157

In this paper, we propose a novel multi-expert approach with random subspace to person re-identification. Considering the high dimension of feature descriptors, random subspace is applied because it can fully exploit the discriminative information that spreads across the feature descriptors. For our approach, we train multiple expert functions with each focusing on a subset of feature blocks randomly selected from images. In addition, we also introduce a weighted fusion scheme to combine the opinions of multiple expert functions trained together. Although each expert function only focuses on a subset of features, our fusion scheme can effectively combine individual recommendations to form a final conclusion. Thus, our method overall still makes use of all features without losing much information they carry. Yet each individual expert function can be trained efficiently without overfitting. We have tested our method on three public datasets, VIPeR [1], ETHZ [8], and CAVIAR4REID [9]. Our experimental results demonstrate that our method significantly outperforms other state-of-the-art techniques for person re-identification.

II. Related Work

Multiple unsupervised learning methods for person reidentification have been presented. Farenzena *et al.* [3] exploits the symmetry property of human figures, and extracts features from human body parts. These features are weighted by their distance to the symmetry axis of the human body. *Zhao et al.* [2] makes use of salient regions in pedestrian images and applies a weighted matching of salient patches. *Kviatkovsky et al.* [4] introduces the invariant property of the internal structure of a color distribution, and applies this signature to matching a pair of images.

With respect to supervised learning, two types of methods, metric learning and SVMs, have been investigated. Metric learning [10], [11], [12], [13], [5] has attracted much attention in recent years as a natural choice for person re-identification because it can effectively calibrate some of the discrepancies, such as illumination and camera photometric settings, among multiple cameras. Among the available choices, learning a Mahalanobis distance metric is mostly used, and it tries to learn a metric which projects extracted feature descriptors to a different space using a linear transformation. Euclidean distance can then be used in the new space to measure the dissimilarity between the persons appearing in a pair of images. Dikmen *et al.* [14] has applied Large Margin Nearest Neighbour [10] to person reidentification and introduced a uniform threshold to determine whether a given image pair is a match. Koestinger *et al.* [12] learns a distance metric from equivalence constraints from a statistical inference perspective. Mignon *et al.* [15] introduces pairwise constrained component analysis to project high dimensional data into low dimensional space, where distances between data points complies with a set of sparse training pairwise constraints. SVM-based methods [16], [6] treats person re-identification as a classification of consistent matches against inconsistent matches. Prosser *et al.* [16] reformulates person re-identification as a ranking problem and learns a subspace where a correct match is given the highest rank. Li *et al.* [6] learns a discriminant function which is equivalent to a second-order polynomial SVM classifier, which decides whether two images match or not by checking the value of the discriminant function. In addition to these two types of methods, AdaBoost [17] has also been applied to person re-identification.

Random subspaces have been used in a variety of computer vision techniques, including image classification [18], face recognition [19] and human detection [20]. Random subspaces are adopted for building a classifier ensemble that is robust against partial occlusions in [20]. Random sampling is used in the face recognition system in [19], which integrates shape, texture, and Gabor responses. Nonetheless, random subspaces have not been used in the context of person re-identification.

Note that multiple experts have been used to solve person re-identification. Li *et al.* [21] trains multiple experts and each expert is trained using subsets of image pairs with similar cross-view transforms. In comparison, experts in our method focus on different subsets of features from all training images, instead of different subsets of training images. And experimental results have shown the superiority of our method based on random subspaces.

III. Feature Description

We use color histograms and local binary patterns as image features. As described by Fig. 2, each image is divided into a rectangular grid with the size of each grid cell set to 4×4. Two horizontally adjacent grid cells form a window with 8×4 pixels. When a window is sliding over an image, its corners are always grid points. That means the step size of the sliding window is 4 pixels in both axes. We further define a block as a 16×8 rectangular region. As a result, each block contains 2 rows of windows with 3 overlapping windows on each row.

We use a window as a spatial unit for feature extraction. For LBP, the values over a window are deposited into a histogram with 35 bins. And for HSV and YUV color spaces, we define an 8-bin histogram for each color channel. In addition, the moments of each color channel are also added as features. This includes the mean, standard deviation, and skewness of each color channel. Among these, the skewness measures the degree of asymmetry of a distribution. Therefore, for each window, we have a 35-dimensional LBP feature, and a 66-dimensional HSV-YUV feature. These two features are simply concatenated together to form the feature descriptor of a window. The feature descriptor of a block is formed by stacking together the descriptors of the six windows inside the block.

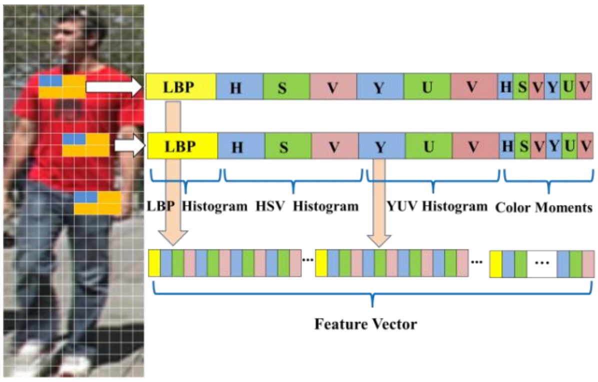

Figure 2. Our feature extraction process

IV. Re-Identification Using Multiple Experts

Our multi-expert approach trains a group of expert functions, each of which only takes a randomly chosen subset of the image features as input, and produces an estimated dissimilarity of the persons appearing in a pair of images. We use the following scheme to randomly choose a subset of features for training and testing every expert function. Suppose an input image is completely covered by a set of non-overlapping blocks. We randomly choose a subset of the blocks. Since there are two types of features (color histograms and LBPs) we use, we also randomly choose one of these two features to represent each chosen block. That is, we might choose different features to represent different blocks. Feature descriptors of all chosen blocks are stacked together to form a single descriptor for the input images. For a given pair of input images, every trained expert function returns a dissimilarity value. We further introduce weighted fusion schemes to combine the dissimilarity values returned by all expert functions, and compute a final dissimilarity score.

A. *Expert Functions*

Suppose x_i and x_j are feature descriptors of dimension *d*, an expert function *f* is a continuous function that maps a pair of feature descriptors to a real value representing the dissimilarity between them.

$$f:(x_i \in R^d, x_j \in R^d) \to R \quad (1)$$

Given the feature descriptor x of an image, let x^+ be the feature descriptor of another image of the same person, and x^- be the feature descriptor of an image of a different erson. An expert function should satisfy the following inequality since our re-identification results are based on ranking the dissimilarity scores returned by the expert functions.

$$f(x,x^+) < f(x,x^-) \quad (2)$$

An expert function can take many different forms. In this paper we primarily take learned Mahalanobis distance metrics or the discriminant function of support vector machines as expert functions. When we learn a Mahalanobis distance metric to measure dissimilarity, the expert function is defined as follows.

$$f_{Mdist}(x_i,x_j) = (x_i - x_j)^T M (x_i - x_j) \quad (3)$$

where M is a positive semidefinite matrix. There exist many metric learning algorithms [10], [11], [12], [13] for obtaining such a matrix from labeled training data.

When we train a linear or polynomial **SVM** to measure dissimilarity, the expert function is defined as follows.

$$f_{SVM}(x_i,x_j) = c^T B + b \quad (4)$$

where *c* and *b* are respectively the learned coefficient vector and bias of the **SVM**, and **B** is the polynomial basis vector. Let $z^T = [x_i^T x_j^T] = [z_1\ z_2 \ldots z_{2d}]$ be a vector formed by concatenating x_i and x_j. Then **B** = *z* for the linear basis, and
$\mathbf{B} = [z_1^2\ z_2^2\ \cdots\ z_{2d}^2\ z_1z_2\ z_1z_3\ \cdots\ z_{2d-1}z_{2d}\ z_1\ \cdots\ z_{2d}]^T$ for the second-order polynomial basis. In the case of a second order polynomial basis, the expert function can be rewritten as

$$f_{QSVM}(x_i,x_j) = z^T Q z + a^T z + b \quad (5)$$

where *Q* is a symmetric 2*d* ×2*d* matrix, and $a,b \in R^d$.

B. *Multi-Expert Fusion*

With multiple subsets of features $C_1, C_2, \ldots, C_m$, we train a group of expert functions $f_1, f_2, \ldots, f_m$. We introduce a weighted fusion to harness this group of expert functions. When they are applied to evaluate the dissimilarity between a pair of feature vectors, x_i and x_j, the final dissimilarity score is computed as follows.

$$F(x_i,x_j) = \sum_{k=1}^{m} w_k f_k(x_i,x_j) \quad (6)$$

where w_k is the fusion weight of the k-th expert function. During person re-identification, given x_i, the dissimilarity scores between x_i and every gallery image are sorted, and the pair with the smallest score is regarded as a correct match.

If all experts are considered equally important, all w_k's are set to 1. In general, different expert functions should be given different weights in the final result. This is because different feature subsets have varying capabilities in terms of identifying individual people. For instance, expert functions trained using features extracted mostly from the background of the input image may have little discrimination ability [22]. Given this, we have made use of visual saliency to assign different weights to the expert functions.

C. *Saliency Weights*

Visual saliency has been proven to provide useful cues in person re-identification. And many algorithms have been developed for estimating pixel-wise visual saliency. For our problem, we apply the method in [23] to obtain a per-pixel saliency map for every image. Expert functions that are trained on features with greater visual saliency should be given a larger weight. Suppose a expert f is trained using features extracted from a subset of image blocks $b_1, b_2, \cdots, b_L$. The saliency weight of f is calculated as follows,

$$w(f) = \frac{\sum_{i=1}^{L} S(b_i)}{L} \quad (7)$$

where $S(b_i)$ is the sum of the saliency value at all pixels in block b_i. Once the saliency weight of all expert functions has been computed, the sum of all weights is normalized to one.

Remark: Although each expert function is trained on a random subset of features, our method is not a feature selection technique because our method does not evaluate the quality of each feature subset and the union of features used by all expert functions is likely to cover all original features. Our method bears more resemblance to boosting [24] and bagging [25], but still maintains significant differences from them. This is because the new classifier in each iteration of boosting or bagging is trained using the complete feature vector of every training sample albeit the distribution over the training samples may change while each of our expert function is trained using only partial feature vectors and the distribution over the entire training dataset is kept fixed.

V. Experiments

We have conducted experiments on three public datasets for person re-identification, VIPeR [1], CAVIAR4REID [9], and ETHZ [8]. In our experiments, we adopted as expert functions two types of learned distance metrics, LMNN [14] and KISSME [12], and the discriminant function of symmetric quadratic SVM (SQSVM) from [6]. We set two goals for our experiments. First, verify that our multi-expert approach can bring significant performance gain to existing methods for person re-identification that are based on a decision function. Second, demonstrate that by fusing the results from multiple expert functions, our method can outperform existing state-of-the-art methods in terms of recognition accuracy.

A. VIPeR

The VIPeR dataset contains 632 pedestrians, with each pedestrian having two 128 × 48 images taken with a pair of disparate cameras. The challenging part of VIPeR is the difference in the viewing directions, which ranges from 45 to 180 degrees. In addition, the variations in pedestrian pose, illumination, image quality, and chrominance also add extra difficulty to person re-identification.

In our experiments, we follow the same protocol as in [12], and the dataset is split into training and testing subsets each containing 316 pedestrians randomly sampled non-repeatedly from the original dataset. For each image, we extract a feature descriptor according to Section 3. In our method, 30 expert functions are trained each using features extracted from a subset of 60 blocks randomly chosen from those covering the image, and for each block, the type of feature is also randomly chosen. Feature descriptors of all chosen block are stacked together to form a descriptor for the image. To compare results with and without multiple expert functions, we also extract features from all blocks of the image and train a single decision function. To avoid overfitting and speed up the training process, we follow the convention that each feature descriptor is reduced to 600 dimensions using PCA. To evaluate the performance, we use the cumulative matching characteristic (CMC) curve, which shows the probability that a correct match is found in top n matches. In this paper, person matching accuracy is always reported as an average of ten runs.

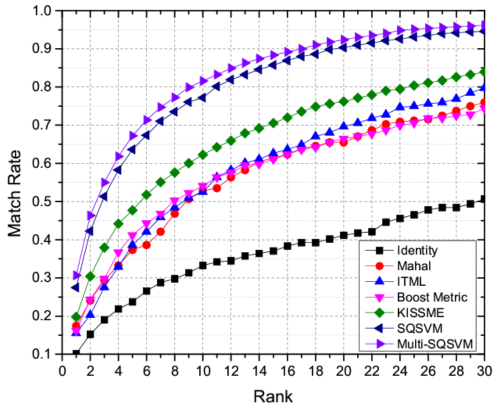

Figure 3. Cumulative matching characteristic (CMC) curves for the VIPeR dataset

In Fig. 3, we present the CMC curves of a few methods for person re-identification, including BoostMetric [26], ITML [11], KISSME [12], SQSVM [6], and our multi-expert SQSVM. When comparing our technique against existing methods, such as KISSME [12], we take the feature vector we developed in Section as the input to such methods rather than using the feature originally designed for them. Also note that, when used for person re-identification, the performance of LMNN varies with the type of features used [12], [27]. In Table I, we compare the performance of our method with existing state-of-theart methods, and the results for top 50 ranks are shown. As we can see, our method achieves best results over all ranks compared with other state-of-the-art methods for person re-identification such as SQSVM, KISSME, LAFT [21], LF [5], PS [9] and SDALF [3]. And in the important range between top 5 and top 10 ranks, our methods outperform others by more than 4%.

In addition, we also compare the results from single KISSME, LMNN and SQSVM, trained using features from all blocks of an image, with the results from their respective multi-expert versions. The detailed comparison of matching accuracy for top 50 ranks are also given in Table II.

TABLE I. Comparison with State-of-the-Art Methods on VIPER.

RANK	1	10	25	50
Multi-SQSVM	30.6%	81.7%	95.1%	98.7%
SQSVM [6]	27.2%	77.2%	93.1%	97.6%
LAFT [21]	29.6%	69.3%	88.7%	96.8%
LF [5]	24.2%	67.1%	85.1%	94.1%
PS [9]	21.8%	57.2%	76.0%	88.1%
SDALF [3]	19.9%	49.4%	70.5%	84.8%

TABLE II. COMPARISON OF RESULTS WITH AND WITHOUT MULTIPLE EXPERTS ON VIPER.

RANK	1	10	25	50
Multi-SQSVM	30.6%	81.7%	95.1%	98.7%
SQSVM	27.2%	77.2%	93.1%	97.6%
Multi-LMNN	30.2%	75.1%	89.8%	96.7%
LMNN	23.1%	67.5%	85.9%	95.7%
Multi-KISSME	32.1%	77.4%	91.3%	97.3%
KISSME	19.8%	66.2%	80.4%	92.3%

We can see that the results of KISSME, LMNN and SQSVM have been significantly improved using multiple experts. In particular, for KISSME and LMNN, more than 7% improvements have been achieved. The reason is that with multiple expert functions, although each expert knows only part of the information about the pedestrians, we can still combine the opinions of multiple experts to obtain a more complete understanding, whereas a single decision function that is trained using complete features overlooks subtle and discriminative information in the training process.

We further verify the performance of the proposed weighted fusion scheme, and the results are presented in Table III. We can see that our multi-expert approach with the proposed weighted fusion schemes outperforms the one with uniform weighting because experts that capture insignificant information such as the background have little discrimination ability, thus negatively affecting the accuracy of the fused evaluation of dissimilarity.

B. CAVIAR4REID

CAVIAR4REID was extracted from the CAVIAR datasets, and consists of 72 pedestrians, 50 of which have images from two cameras, and the remaining 22 have images from one camera only. For each person, the set of images were chosen to maximize variations in resolution, lighting condition, occlusion and pose. Different from VIPeR, images in CAVIAR4REID have a greater variation of resolution, ranging from 17 × 39 to 72 × 144. In our experiments, we follow the same protocol as in [21]. That is, we do not split the pedestrians into a raining set and a testing set because there are too few of them. Instead, if a person has images from two camera views, we randomly choose one pair of images from different cameras for training, and then another random pair of images of the same person from different cameras for testing. We follow the same feature extraction steps specified in Section except that images are resized to 36 ×80 and 50 blocks are chosen for each expert function. Both distance metrics learned using LMNN and discriminant functions of SQSVM have been tested as expert functions on this dataset, and the final identification results are an average of ten runs.

TABLE III. COMPARISON OF MULTI-EXPERT LMNN RESULTS WITH AND WITHOUT WEIGHTED FUSION. THE TWO ROWS SHOW RESPECTIVELY THE MATCHING ACCURACY OF MULTI-EXPERT FUSION WITH SALIENCY WEIGHTS, AND UNIFORM WEIGHTS.

RANK	1	5	10	15
Saliency	30.2%	60.7%	75.0%	82.2%
Uniform	28.8%	59.1%	73.3%	81.3%

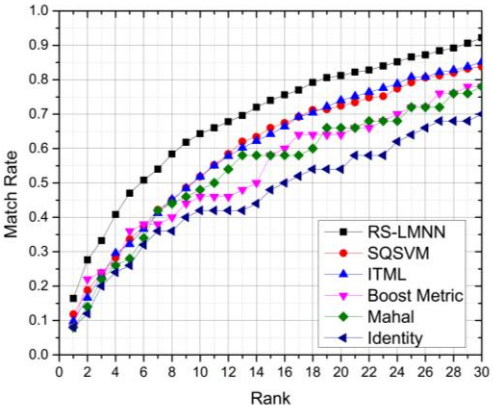

Figure 4. Cumulative matching characteristic (CMC) curve on the CAVIAR4REID dataset.

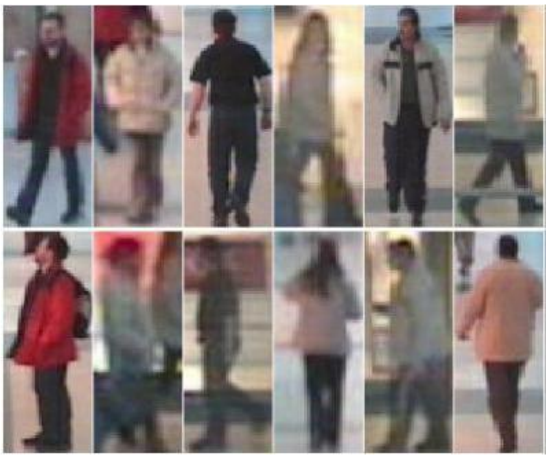

Figure 5. Sample image pairs from CAVIAR4REID. Images of the same person appear in the same column. Multiwitness LMNN can match images from the same column correctly while other methods cannot.

We have compared multi-expert LMNN with a few state-of-the-art methods, such as SQSVM, LAFT, PS, Boost- Metric and SDALF, as shown in Fig. 4 and Table IV. The results show that our multi-expert approach achieves significant performance improvements over other methods across all ranks. For top 5 and top 10 ranks in particular, more than 10 percentage points of improvements have been achieved. Fig. 5 shows sample image pairs that multi-expert LMNN can match correctly while other methods cannot. We have also compared the performance of LMNN and SQSVM with and without multiple experts on CAVIAR4REID, and reported the results in Table V. The substantial improvements achieved there further consolidate the effectiveness of our approach.

TABLE IV. COMPARISON WITH STATE-OF-THE-ART METHODS ON CAVIAR4REID.

RANK	1	5	10	30
Multi-LMNN	16.1%	47.0%	64.7%	92.3%
LAFT [21]	10.2%	39.0%	59.0%	88.0%
SQSVM [6]	11.8%	33.6%	51.8%	83.8%
PS [9]	8.5%	32.0%	48.0%	86.0%
SDALF [4]	6.8%	25.0%	45.0%	83.0%

TABLE V. COMPARISON OF RESULTS WITH AND WITHOUT MULTIPLE EXPERTS ON CAVIAR4REID.

RANK	1	5	10	30
Multi-LMNN	16.1%	47.0%	64.7%	92.3%
LMNN	10.2%	37.0%	51.0%	88.0%
Multi-SQSVM	15.8%	40.4%	55.2%	89.8%
SQSVM	11.8%	33.6%	51.8%	83.8%

C. ETHZ

TABLE VI. COMPARISON WITH STATE-OF-THE-ART METHODS ON ETHZ.

RANK	1	2	3	4	5	6	7
Multi-KISSME	80%	86%	88%	90%	92%	94%	94%
KISSME [12]	72%	80%	82%	86%	88%	88%	90%
Pairwise Metric [28]	77%	83%	87%	91%	92%	92%	92%
PLS [29]	79%	85%	86%	87%	88%	89%	90%
SDALF[4]	65%	73%	77%	79%	81%	82%	84%
Boost Metric [26]	63%	74%	77%	78%	79%	80%	83%

Images in the ETHZ dataset were captured from moving cameras. The dataset contains three sequences which have 4857 images of 83 pedestrians, 1936 images of 35 pedestrians, and 1762 images of 28 pedestrians respectively. All images were resized to 32×64. We chose the first sequence, which contains most people, to conduct our experiments. We follow the same steps as in Section to extract features for each image, and select 40 blocks for each expert function. Similar to what we did on CAVIAR4REID, we do not split the pedestrians into non-overlapping training and testing sets, instead we randomly choose one pair of images for each person for training, and another pair of images for the same person for testing. Multi-expert KISSME has been tested on this dataset, and the results are given in Table IV, which presents the CMC result for top 7 ranks. The experimental results show the superiority of our approach. Especially the improvement multi-expert KISSME achieves over the original KISSME reinforces the advantages of multiple experts with partial information over one single expert with complete information.

VI. CONCLUSIONS

In this paper we have developed a novel multi-expert approach based on random subspace for person re-identification, where multiple expert functions are trained with each function focusing on a subset of features. Compared to traditional methods where a single decision function is trained based on complete feature information, our approach has proven to be able to fully exploit the discriminant information in feature descriptors, and bring substantial improvements, and achieve higher accuracy. Our approach has great applicability, and can be widely adopted in methods for person re-identification that learn a decision function with continuous value to evaluate the similarity between two persons.

ACKNOWLEDGMENT

This work was partially supported by Hong Kong Innovation and Technology Fund (GHP/060/11).

REFERENCES

[1] D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking,” in *Proc. the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2007.

[2] R. Zhao, W. Ouyang, and X. Wang, “Unsupervised salience learning for person re-identification,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[3] M. Farenzenai, L. Bazzanil, A. Perinal, V. Murino, and M. Cristani, “Person re-identication by symmetry-driven accumulation of local features,” in *Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2010.

[4] I. Kviatkovsky, A. Adam, and E. Rivlin, “Color invariants for person re-identification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[5] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, “Local fisher discriminant analysis for pedestrian re-identification,” in *Proc. CVPR*, 2013.

[6] Z. Li, S. Y. Chang, F. Liang, T. S. Huang, L. L. Cao, and J. R. Smith, “Learning locally-adaptive decision functions for person verification,” in *Proc. CVPR*, 2013.

[7] D. Chen, X. D. Cao, F. Wen, and J. Sun, “Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification,” in *Proc. CVPR*, 2013.

[8] W. R. Schwartz and L. S. Davis, “Learning discriminative appearance-based models using partial least squares,” in *Proc. the XXII Brazilian Symposium on Computer Graphics and Image Processing*, 2009.

[9] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, “Custom pictorial structures for re-identification,” in *BMVC*, pp. 1–11, 2011.

[10] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *Journal of Machine Learning Research*, 2010.

[11] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” in *Proc. the 24th International Conference on Machine Learning*, 2007.

[12] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *Proc. IEEE Intern. Conf. on Computer Vision and Pattern Recognition*, 2012.

[13] M. Guillaumin, J. Verbeek, and C. Schmid, “Is that you? metric learning approaches for face identification,” in *Proc. ICCV*, 2009.

[14] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, “Pedestrian recognition with a learned metric,” in *Proc. ACCV, Part 4*, vol. 6495, 2010, pp. 501-512.

[15] A. Mignon and F. Jurie, “PCCA: A new approach for distance learning from sparse pairwise constraints,” in *Proc. CVPR. IEEE*, 2012.

[16] B. Prosser, W. S. Zheng, S. G. Gong, and T. Xiang, “Person re-identification by support vector ranking,” in *BMVC*, 2010.

[17] D. Gray and H. Tao, “Viewpoint invariant pedestrian recognition with an ensemble of localized features,” in *Proc. ECCV*, 2008.

[18] L. I. Kuncheva, J. J. Rodr´ıguez, C. O. Plumpton, D. E. J. Linden, and S. J Johnston, “Random subspace ensembles for fmri classification.,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 531–542, 2010.

[19] X. G. Wang and X. O. Tang, “Random sampling for subspace face recognition,” *International Journal of Computer Vision*, vol. 70, no. 1, pp. 91–104, 2006.

[20] J. Mar´ın, D. V ´azquez, A. M. L ´opez, J. Amores, and L. I. Kuncheva, “Occlusion handling via random subspace classifiers for human detection,” *IEEE Transactions on Cybernetics*, vol. 44, no. 3, pp. 342–354, 2014.

[21] W. Li and X. G. Wang, “Locally aligned feature transforms across views,” in *Proc. CVPR*, 2013.

[22] T. Kozakaya, S. Ito, and S. Kubota, “Random ensemble metrics for object recognition,” in *Proc. IEEE International Conference on Computer Vision*, Barcelona, Spain, November 6-13, 2011.

[23] X. D. Hou, J. Harel, and C. Koch, “Image signature: Highlighting sparse salient regions,” *IEEE Trans. Pattern Anal. Mach. Intell*, 2012.

[24] R. E. Schapire, “The boosting approach to machine learning: An overview,” in *Proc. MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[25] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, pp. 123–140, 1996.

[26] C. H. Shen, J. Kim, L. Wang, and A. V. D. Hengel, “Positive semidefinite metric learning with boosting,” in *Advances in Neural Information Processing Systems (NIPS’09)*, 2009, pp. 1651–1659.

[27] X. Liu, M. L. Song, D. C. Tao, X. C. Zhou, C. Chen, and J. J. Bu, “Semi-supervised coupled dictionary learning for person re-identification,” *IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 3550–3557. 2014.

[28] M. Hirzer, P. M. Roth, M. K ̈ostinger, and H. Bischof, “Relaxed pairwise learned metric for person reidentification,” in *Proc. ECCV* (*Part* 6), 2012.

[29] W. R. Schwartz and L. S. Davis, “Learning discriminative appearance-based models using partial least squares,” in *Proc. SIBGRAPI*, 2009.

Sai Bi received the BEng degree in computer science from the University of Hong Kong in 2014. He is a recipient of HK-SAR Government Scholarship, HKU Foundation Scholarship, and HKU Undergraduate Research Fellowship. His current research interests include computer graphics and computer vision.

Guanbin Li received the BE and Master’s degree in computer science from Sun-Yat Sen University in 2009 and 2012. He is currently a PhD candidate in the Department of Computer Science, the University of Hong Kong. He is a recipient of Hong Kong Postgraduate Fellowship. His current research interests include computer vision, image processing, and deep machine learning.

Yizhou Yu received the PhD degree in computer science from University of California at Berkeley in 2000. He is currently a professor in the Department of Computer Science at the University of Hong Kong. Prof. Yu is a senior member of IEEE. He has served as an associate editor of Computer Graphics Forum and the Visual Computer, and is on the editorial board of International Journal of Software and Informatics. His current research interests include computer graphics, computer vision, digital geometry processing, video analytics and biomedical data analysis.