Multi-Shot Re-Identification with Random-Projection-Based Random Forests

Yang Li, Ziyan Wu, Richard J. Radke
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
liy21@rpi.edu, ziyan@alum.rpi.edu, rjradke@ecse.rpi.edu

Abstract

Human re-identification remains one of the fundamental, difficult problems in video surveillance and analysis. Current metric learning algorithms mainly focus on finding an optimized vector space such that observations of the same person in this space have a smaller distance than observations of two different people. In this paper, we propose a novel metric learning approach to the human re-identification problem, with an emphasis on the multi-shot scenario. First, we perform dimensionality reduction on image feature vectors through random projection. Next, a random forest is trained based on pairwise constraints in the projected subspace. This procedure repeats with a number of random projection bases, so that a series of random forests are trained in various feature subspaces. Finally, we select personalized random forests for each subject using their multi-shot appearances. We evaluate the performance of our algorithm on three benchmark datasets.

1. Introduction

Recognizing the same person across a network of cameras with non-overlapping views, also known as re-identification (re-id), is a major issue in surveillance applications. This is a fairly challenging task, since the appearance of the same person may vary significantly in different camera views. Inter/intra camera illumination changes, occlusion, and pose changes aggravate the already difficult problem.

Most of the previous research on re-id is focused on the single-shot scenario [34, 22, 19, 20]. That is, there is only one image available for both the tagged person and each matching candidate. However, real-world re-id problems map much more naturally onto a multi-shot problem. That is, there are multiple images available to describe both the target and the matching candidates. For example, after a person of interest is detected in the field of view of one camera, he/she is usually tracked until leaving the current view. Similarly, candidates detected in other views will generally be tracked continuously. Thus, re-id in practice is actually a multi-shot probe and multi-shot gallery problem.

Current metric learning algorithms for the re-id problem are hindered by several difficulties. First, the dimension of the feature vectors used to represent images is usually very large, which makes the training process both time- and space-consuming. Moreover, most of these algorithms solve an optimization problem constrained by sparse pairwise distances of samples, and it is difficult to avoid overfitting. Lastly, they are only designed for single shot re-id problems, and are not able to take advantage of multiple images of the same person. That is, they represent unified models for all the targets and are unable to incorporate intraclass information extracted from a multi-shot scenario.

In this paper, we propose a novel metric learning method based on random forests [1, 10] to specifically address the multi-shot re-id problem. The intuition is that the voting nature of a random forest's output makes decision isolation and aggregation very easy. We build an ensemble of random forests, trained by feature vectors randomly projected onto different subspaces. By taking into account the outputs from discriminatively selected random forests, a similarity measurement is computed for each testing sample pair.

The main contribution of this paper is a robust and efficient algorithm to address the multi-shot re-id problem. This algorithm not only inherits the advantages of random forests such as interpretability, scalability and robustness to over-fitting, but also addresses the difficulty of high-dimensional feature vectors via random projection. Since it operates in a very low dimension, the training process is extremely fast and tremendous storage space is saved compared to current techniques. More importantly, the random projection enhances the classifier diversity of the random forests, which makes it outperform other dimen-

sionality reduction approaches such as Principal Component Analysis (PCA). Based on this framework, we further developed a person-specific random forest method to exploit the discriminative information from the multi-shot samples. Our algorithm demonstrates superior performance compared with state-of-the-art algorithms on multi-shot re-id benchmark datasets.

2. Related Work

Feature-based approaches aim to construct descriptive and distinctive representations from images. Gray et al. [16] learned an ensemble of color and texture features via the AdaBoost algorithm. The same technique has also been used to learn Haar-based discriminative features [3]. Schwartz et al. [27] selected discriminative features using partial least squares (PLS). Farenzena et al. [7] developed an unsupervised method to build an appearance-based descriptor. They exploited perceptual symmetry principles and tried to augment maximally stable color regions (MSCR). Zhao et al. [33] proposed an unsupervised learning method based on salient information to extract discriminative features. Ma et al. [21] combined Gabor filters and covariance descriptors to improve the robustness to illumination variation. Recently, Wu et al. [31] proposed an appearance model integrating camera viewpoint and human pose information.

In addition to feature-based methods, many researchers also approach the re-id problem through metric learning. Many of these methods focus on learning a Mahalanobis-like distance [13, 22, 19, 2]. Another set of approaches employs k-Nearest Neighbors (k-NN) strategies, such as Large Margin Nearest Neighbor (LMNN) and its variants [9, 14, 18]. Prosser et al. [24] proposed an ensemble of RankSVMs to rank pairwise similarity. Zheng et al. [34] built a probabilistic model (PRDC) to estimate distance. Zheng et al. [35] also reformulated this problem as a set-based verification task. Mignon et al. [22] introduced an algorithm that learns distance metrics from pairwise constraints (PCCA). Bak et al. [4] proposed another similarity measurement model based on Mean Riemannian Covariance (MRC) patches extracted from a tracked person. Recently, Pedagadi et al. [23] applied Local Fisher Discriminant Analysis to explore local, discriminative spaces for dimensionally reduced features, improving metric learning outcomes.

Random forests [1, 10] are a well-known decision tree based classifier ensemble. Random projection [8] has been widely applied as a dimensionality reduction method [15]. However, the combination of the two techniques, to the best of our knowledge, has never been evaluated. While decision tree classifiers based on feature extraction [25] or manifold learning based on random projection [6] have been previously proposed, this paper reveals that random projection

makes it possible to efficiently build a low-dimensional random forest ensemble while preserving data information. It is resistant to over-fitting and acquires even higher classifier diversity. The ensemble model also enjoys the advantage of discriminatively selecting random forests for each individual. In this paper, we introduce this concept into the human re-id problem and validate it on various challenging datasets from the surveillance community.

In the following sections, we first review the basic concepts of random forests and random projections. Then we present the proposed algorithm using ensemble random forests and describe how to further build a personally-discriminative model based on multi-shot information. In Section 4, we discuss the choice of critical parameters of the proposed algorithm, and then conduct experiments to compare it with other multi-shot re-id algorithms.

3. The Proposed Method

In many distance metric learning re-id algorithms, pairwise constraints generated from data samples play the fundamental role in model construction [34, 22, 13, 24, 19, 20]. The goal is to minimize the distance between data points from the same class while maximizing the distance between data points from different classes. Formally, suppose $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ represents a set of image feature vectors, where $x_i \in \mathbb{R}^n, i = 1, 2, \ldots, N$ is the feature vector for an image. Divide all sample data point pairs into two sets, $\mathcal{S} = \{(x_i, x_j) \mid x_i, \ x_j \text{ are from the same class}\}$ and $\mathcal{D} = \{(x_i, x_j) \mid x_i, \ x_j \text{ are from different classes}\}$. In general distance metric learning methods, the goal is to find a metric function $Dist(x_i, x_j)$ to minimize the objective function

$$\frac{1}{|\mathcal{S}|} \sum_{(x_i, x_j) \in \mathcal{S}} Dist(x_i, x_j) - \frac{1}{|\mathcal{D}|} \sum_{(x_i, x_j) \in \mathcal{D}} Dist(x_i, x_j)$$

where (x_i, x_j) are training sample feature vector pairs.

The re-id problem tries to recognize a target person among a number of candidates from different camera views; that is, to compare the target person with each candidate and determine whether they are the same person (in the same class). With this interpretation, the re-id problem can be reformulated into a classification problem. Given a sample pair (u, v), we decide whether it belongs to $\mathcal{S}$ or $\mathcal{D}$.

3.1. Random Forest Classifier

A random forest is an ensemble decision tree trained with bootstrap samples on the feature space. We denote a random forest $\mathbf{T} = \{t_1, t_2, \ldots, t_K\}$, where each $t_k, k = 1, 2, \ldots, K$ is a decision tree. Let y represent the label of training sample pair (x_i, x_j), and define

$$y_{ij} = \begin{cases} 1 & \text{if } (x_i, x_j) \in \mathcal{S} \\ 0 & \text{if } (x_i, x_j) \in \mathcal{D} \end{cases} \tag{1}$$

Now given any feature vector pair (x_i, x_j), to determine whether they belong to the same class/person, the classification output of each decision tree can be expressed as

$$y_k = t_k\left(\theta(x_i, x_j)\right), k = 1, \ldots, K$$

where $\theta(x_i, x_j)$ is a function transferring feature vectors to the decision tree input vector. Here we employ the approach Xiong et al. proposed in [32],

$$\theta(x_i, x_j) = \left[\begin{array}{c} |x_i - x_j| \\ \frac{1}{2}(x_i + x_j) \end{array} \right] \qquad (2)$$

which takes both the relative and absolute location of paired vectors into considerations.

Using a voting scheme we compute a similarity score defined as

$$Y = \frac{1}{K} \sum_{k=1}^{K} t_k\left(\theta(x_i, x_j)\right) \qquad (3)$$

The higher the score is, the more confident we are that the pair $(x_i, x_j) \in \mathcal{S}$. Generally, for classification problems, a threshold is used to make the final decision. In the re-id problem, we usually aim to identify a target person in a gallery of several different people, so instead of making a hard decision on classification, we use the similarity score to describe our confidence about whether two sample images belong to the same person.

3.2. Random Projection

In order to avoid training using high-dimensional feature vectors, while preserving the original information, we apply dimensionality reduction using the multiple random projection technique [8]. It has been shown [6] that for data points lying in a low-dimensional manifold in an ambient high-dimensional space $\mathbb{R}^n$, their pairwise ambient distances are well-preserved under random projections from $\mathbb{R}^n$ to $\mathbb{R}^m$, where $m \ll n$.

Again, denote $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ as the set of image feature vectors, where $x_i \in \mathbb{R}^n, i = 1, 2, \ldots, N$. Let p be an n-dimensional random vector in which each coordinate is independently sampled from the normal distribution $\mathcal{N}(0, 1)$. Independently sampling m such vectors, $p_1, p_2, \ldots, p_m$, and normalizing each to a unit vector, the matrix $P = [p_1 \ p_2 \ \cdots \ p_m]$ is an $n \times m$ random projection matrix that maps data vectors from $\mathbb{R}^n$ to $\mathbb{R}^m$, where $m \ll n$. In the proposed method, we orthogonalized the matrix P based on the theory in [6]. Then, for each feature vector in $\mathbf{X}$, we project it onto $\mathbb{R}^m$ by $x_i' = P^\top x_i$, $x_i' \in \mathbb{R}^m$, so that the dimensionally reduced feature vector set is $\mathbf{X}' = \{x_1', x_2', \ldots, x_N'\}$.

3.3. Multiple Random-Projection-Based Random Forest Algorithm

Since random projection has no bias on dimension selection, different sets of projection vectors can lead to un-

even classification results. The ensemble nature of the random forest algorithm leads us to improve the original single random forest classifier by aggregating multiple random forests, each with a different random projection matrix. The overall algorithm for learning the random forest classifier is shown in Algorithm 1.

Algorithm 1: Learning the random forest classifier with multiple random projections

---

Input: R: number of random projections
$\{x_1, x_2, \ldots, x_N\}, x_i \in \mathbb{R}^n$: feature vectors extracted from each training person's image
$\{l_1, l_2, \ldots, l_N\}, l_i \in \mathbf{C}$: class label for each image, $\mathbf{C}$ is the set of all classes
m: projected dimension number
Output: A set of random forests $\{\mathbf{T}_r\}, r = 1, \ldots, R$
for $r = 1$ **to** R **do**
    Generate an $n \times m$ orthoprojector matrix P_r;
    Compute dimensional reduced feature vectors
    $x_i' = P_r^\top x_i, i = 1, \ldots, N$;
    for *each data pair* $(x_i', x_j'), i > j$ **do**
        if $l_i \neq l_j$ **then**
            put (x_i', x_j') in set $\mathcal{D}$;
        else
            put (x_i', x_j') in set $\mathcal{S}$;
        end
    end
    Randomly sample the same number of elements from $\mathcal{S}$ and $\mathcal{D}$ without replacement, such that $|\mathcal{S}| = |\mathcal{D}|$;
    Form the constraint set $\mathcal{C} = \{\theta(x_i', x_j'), y\}$ with (1), where $(x_i', x_j') \in \mathcal{S} \cup \mathcal{D}$;
    Train random forest $\mathbf{T}_r$ with constraint set $\mathcal{C}$;
end

We should point out that the class label set $\{l_1, l_2, \ldots, l_N\}$ is taken as input to Algorithm 1 for the simplicity of the formulation. It is not actually necessary to have the class information for each image. Only the pairwise similar or dissimilar relationship is required to form the data pair sets $\mathcal{S}, \mathcal{D}$ and the constraint set $\mathcal{C}$.

An important issue is the dimension of the random projection subspace m. In [6], the authors estimated the projected subspace dimension that preserves the pairwise points' ambient distances, but this requires the parameters of the underlying manifold such as its dimension and volume, which are difficult to determine. Hegde et al. [17] proposed a manifold learning algorithm to estimate the projection subspace dimension. In fact, most dimension reduction algorithms for manifold-modeled data involve manifold learning [29, 26, 11]. However, when the training points are not sufficiently representative, they may fail to describe the whole structure, which can mislead the manifold learning result.

In our algorithm, since we aggregate results from multiple random projections, which is equivalent to increasing the projection dimension, it is unnecessary to learn the underlying structure. Thus, there exists a trade-off between the projected dimension number m and the number of random projections R. We will show in the experiments that

different choices of (m, R) can result in similar classification accuracies.

Algorithm 1 only employs a subset of the similarity set $\mathcal{S}$ and dissimilarity set $\mathcal{D}$ to reduce the computational complexity. With N feature vectors, there will be $\mathcal{O}(N^2)$ vector pairs in total. Depending on the number of training samples and number of classes, the number of pairs is usually very large. Also, the sizes of $\mathcal{S}$ and $\mathcal{D}$ tend to be uneven; i.e., positive pairs are much less common than negative pairs. It is unrealistic to train the random forest with all these constraints, which requires a huge computational effort. We thus randomly sample the same number of elements from both sets to build balanced sets $\mathcal{S}$ and $\mathcal{D}$. Moreover, multiple runs based on different random projections will result in diverse, complementary random selections of constraints. In our experiments, we fixed the number of sampled elements from the sets $\mathcal{S}$ and $\mathcal{D}$ as 1000.

It is natural to question why Principal Component Analysis (PCA) could not be used to perform dimension reduction in this case. One important reason is that PCA is not a suitable feature selection method for classification problems, since it discards the discriminative information. Some literature [28, 30] also proposes ensemble methods based on PCA. They either keep all the components generated by PCA but use different subsets of components to train classifiers, or use different subsets of PCA-extracted features to train classifiers. The classifier diversity is not as good as in our approach, where we employ multiple independent random projections, and each classifier is trained with randomly extracted features. Another advantage of random projection is that the computational complexity is much lower compared to PCA. Comprehensive comparisons [8] between PCA and random projection suggest that random projection outperforms PCA, on both accuracy and computational efficiency, especially with higher dimensional data.

After the set of multiple random-projection-based random forests $\{\mathbf{T}_r\}, r = 1, \ldots, R$ has been trained, the similarity score defined in Equation (3) can be updated as

$$Y = \frac{1}{RK} \sum_{r=1}^{R} \sum_{k=1}^{K} t_k^r \left(\theta(x_i, x_j) \right) \tag{4}$$

where (x_i, x_j) is the unclassified data point pair, and t_k^r is the kth decision tree of random forest $\mathbf{T}_r$.

3.4. Personally Discriminative Forests

Current re-id metric learning algorithms focus on developing a metric model applied universally to any input feature vector pair; thus they are inherently unable to adapt to multi-shot information. Since our approach aggregates the results of multiple trained random forests, it is possible to "personalize" these forests based on multiple descriptors. Let $\tilde{\mathbf{X}} = \{\tilde{x}_1, \ldots, \tilde{x}_M\}$ where $\tilde{x}_i \in \mathbb{R}^n, i = 1, \ldots, M$
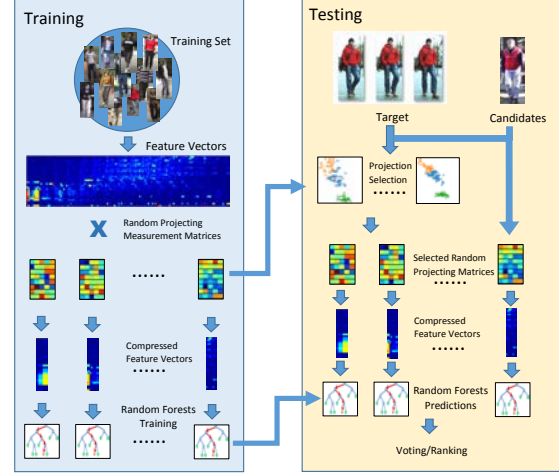

Figure 1. Flowchart of the proposed algorithm

be the set of all available feature vectors for a single person. With multiple image signatures, we can form same-class pairs $(\tilde{x}_i, \tilde{x}_j) \in \mathcal{S}, i \neq j$, and put these constraints into each trained random forest, expecting high similarity scores. Similarly, we can pair $\tilde{x}_i$ with feature vectors in the training set to form the different-class set $\mathcal{D}$ and expect low similarity scores. Then the ratio of the two scores can be used to measure the personal discrimination for a specific random forest. Only those random forests above the average score ratio for a specific person are selected as personally discriminative classifiers. The detailed algorithm is shown in Algorithm 2.

Algorithm 2: Personally Discriminative Random Forest Selection

---

Input: $\{P_1, P_2, \ldots, P_R\}$: random projection matrices
$\{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_R\}$: trained random forests
$\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_M\}, \tilde{x}_i \in \mathbb{R}^n$: feature vectors extracted from multi-shot images of a person
$\{x_1, x_2, \ldots, x_N\}, x_i \in \mathbb{R}^n$: feature vectors in the training set
Output: random forest set $\{\mathbf{T}_{r'}\} \subset \{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_R\}$
for $r = 1$ to R **do**
    Compute dimensionally reduced feature vectors
    $\tilde{x}_i' = P_r^\top \tilde{x}_i, i = 1, \ldots, M$, and $x_i' = P_r^\top x_i, i = 1, \ldots, N$;
    for *each data pair* $(\tilde{x}_i', \tilde{x}_j'), i > j$ **do**
        | Calculate similarity score $\tilde{y}_{ij}$ with (3);
    end
    Average all $\tilde{y}_{ij}$ to get positive-pair similarity score Y_r^+;
    for *each data pair* $(\tilde{x}_i', x_j'), i = 1, \ldots, M, j = 1, \ldots, N$ **do**
        | Calculate similarity score y_{ij} with (3);
    end
    Average all y_{ij} to get negative-pair similarity score Y_r^-;
    Score ratio $\sigma_r = \frac{Y_r^+}{Y_r^-}$
end
Find the random projection trees $\mathbf{T}_{r'}$, whose score ratio
$\sigma_{r'} > mean\{\sigma_r\}, r = 1, \ldots, R$;
Return $\{\mathbf{T}_{r'}\}$;

Define the set $\mathcal{R} = \{r'\}$, where r' collects the indices of the selected random forests from Algorithm 2. Then,

Equation (4) can be updated as

$$Y = \frac{1}{|\mathcal{R}|K} \sum_{r' \in \mathcal{R}} \sum_{k=1}^{K} t_k^{r'} \left(\theta(u, v) \right) \qquad (5)$$

where $u \in \tilde{\mathbf{X}}$ which is used to select the appropriate person-discriminative random forests $\{\mathbf{T}_{r'}\}$. Figure 1 illustrates the overall method.

4. Experimental Results

The proposed approach is designed for the multi-shot human re-identification problem, so we chose the multi-shot benchmark datasets ETHZ [27], CAVIAR4REID [12], and 3DPeS [5] to evaluate the performance. All the results are shown as ranked matching rates in average cumulative match characteristic (CMC) curves.

In our experiments, we extracted image feature vectors as described in [16]. The image is evenly divided into 6 horizontal strips, and the features of each image strip are extracted in the form of 8 color channels and 19 texture channels, where each channel contains a histogram with 16 bins. By concatenating all the histograms we get a feature vector with a dimension of 2592. The experimental settings follow the multi-shot re-id case in [7]. Specifically, for each testing person, we randomly take J images to form a gallery set and J additional random images to form a probe set. When there are not enough shots available to fill J images in both sets, we evenly assign the images into two sets. As opposed to [7], which uses multi-shot images from the same person to build a single signature, we use multi-shot information to select random forest models as described in Section 3.4. To estimate the similarity, for instance, between Person A in the probe set and person B in the gallery set, we calculate the similarity score of every image pair consisting of one image from Person A and one image from Person B, and the averaged score is the final similarity score between the two people. This score is then used to rank all the candidates. All of our results are averaged over 100 repeated experiments.

Since most multi-shot re-id results are reported in the context of feature (or descriptor) learning algorithms, they usually do not require an extra training set. To make a fair comparison, we formed the training set using as few people as possible: 14 for each ETHZ sequence and 16 for CAVIAR4REID. 3DPeS has only been reported in [23], where the training size is 95.

A remaining question is how many decision trees should be learned for a random forest. Increasing forest size will increase the computational and storage cost, and there will be little or no gain after passing a certain point. While a detailed answer to this question is beyond the scope of this paper, we observed that the number of trees should scale with the input vector dimension.

4.1. Experimental Datasets

ETHZ [27] This dataset consists of three video sequences captured by moving cameras in a street scene. Schwartz et al. [27] extracted an image set of pedestrians to perform appearance-based model learning, which essentially converts it into a re-identification dataset. In the dataset, sequence 1 contains 4857 images from 83 pedestrians, sequence 2 contains 1936 images from 35 pedestrians, and sequence 3 contains 1762 images from 28 pedestrians. The images have a range of appearance variation resulting from the moving cameras. The dataset also suffers from illumination changes and occlusion. All the images are normalized to 64×128 pixels.

CAVIAR4REID [12] This dataset is extracted from videos captured in a shopping center. It is suitable for the re-id problem since the videos come from two camera views. It contains large variance in image resolution, large illumination variation within and between cameras, and severe occlusion and pose changes. The dataset contains 72 pedestrians with 1220 images in total. All the images are normalized to 64×128 pixels.

3DPeS [5] This dataset is specifically designed for the human re-id problem. The images are captured from 8 different surveillance cameras with non-overlapping fields of view in a university campus. The main challenge in this dataset is the severe lighting condition and viewpoint variation. The dataset includes 193 individuals with a total of 1012 images. All the images are normalized to 64×128 pixels.

4.2. Trade-off Between Projected Subspace Dimension and Number of Random Projections

In Section 3.3, we argued that, because of the voting nature of random forests, the aggregate results from multiple random projections should converge independent of the projected subspace dimension. We demonstrate this idea with following experiment.

In ETHZ sequence 1, we randomly chose 14 people to form the training set, and put the others in the testing set. For each person in the testing set, 2 images were randomly selected to form the gallery set, and another 2 images selected to form the probe set ($J = 2$). We varied the projected dimension number m in $\{10, 20, 50, 100, 200\}$ and the number of random projections R in $\{1, 5, 10, 20, 50, 100\}$. The purpose is to examine, for each projected subspace dimension, how many random projections are required for convergence. To isolate the convergence process, in this experiment we didn't apply the random forest selection discussed in Section 3.4. The results are shown in Figure 2.

We can see that all the reduced-dimension cases converged after 100 random projections, and that the higher the projected dimension is, the faster the convergence. For
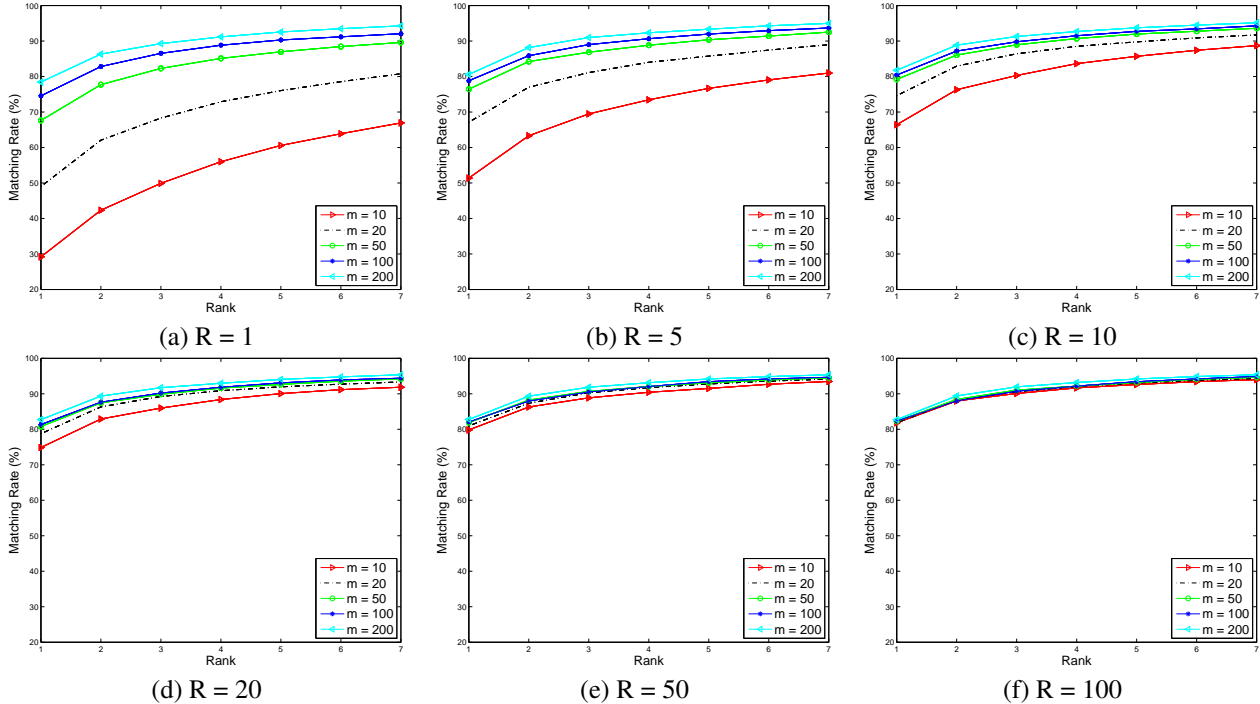
Figure 2. Ranking performance on ETHZ sequence 1 using different projected dimension numbers m and numbers of random projections R.

example, the results for $m = 100$ change little after 5 random projections. It is interesting to note the similarity of the $m = 20$ curve in Figure 2b with the $m = 10$ curve in Figure 2c, which suggests that the random forests are statistically aggregating. This supports our intuition.

Projection Dimension	Running Time (seconds)
10	0.11
20	0.25
50	1.32
100	4.99
200	19.59

Table 1. Training time for a single projection with different projected subspace dimensions. The MATLAB program was run on a 4-core 3.6GHz PC.

The running time for a single projection under each projected subspace dimension is reported in Table 1. When the projected subspace dimension decreases, the training time drops tremendously. With a smaller subspace dimension, the storage and computational requirements are much lower. Also, more projection iterations provide a wider range of feature subspaces and constraint sets, which can contribute to more reliable results. We note that, without random projection, the running time for a single training instances is 500–600 seconds. In the remaining experiments, we chose $m = 20, R = 50$.

4.3. Performance Evaluation

We compared our results for the ETHZ dataset against the competing algorithms SDALF [7], eBiCov [21], and eSDC [33]. The results are shown in Figure 3. We abbreviate the proposed method as **RPRF**. The experiments were separately conducted using 2 and 5 shots. The proposed method outperforms current state-of-the-art algorithms on all sequences when the shot number is increased to 5, supporting the intrinsic design of our method for multi-shot cases. When $J = 2$, only one pair of similar samples can be formed to select preferred random forests, so it might be biased. In sequence 2, when $J = 2$ the ranking performance is even slightly worse than other approaches. Because the people in sequence 2 have fewer images on average, there are fewer training constraint samples, which may compromise random forest training.

The results for the CAVIAR4REID dataset are compared against the competing algorithms PRDC [34], PCCA [22], CPS [12], SDALF [7], and LF [23], shown in Figure 4(a). Our method demonstrates a compelling advantage. We note that [23] used 36 people to train, while in our experiments we only use 16 people in the training set.

In Figure 4(b), we present the performance comparison with PRDC [34], PCCA [22], LF [23] and LMNN-R [14] on the 3DPeS dataset. We note that around 75% of the people in this dataset have fewer than 5 images. With inadequate
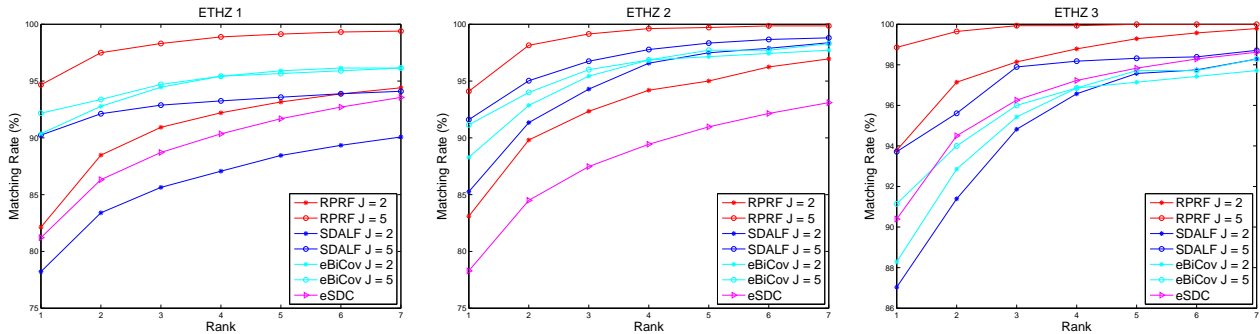
Figure 3. Ranking performance on the ETHZ dataset compared with state-of-the-art algorithms.

multi-shot information, the advantage of our method is not as significant as the previous two datasets.

Note that some of the above comparison ranking data is not directly available for the experiments on the CAVIAR4REID and 3DPeS datasets; we roughly extracted these from the plots in the original papers. For PRDC [34] and PCCA [22] we implemented the algorithms and ran the same experiments to generate the results. We adapted the same strategy described in Section 4 for these algorithms to obtain a single similarity score between any two people when multiple images are available.
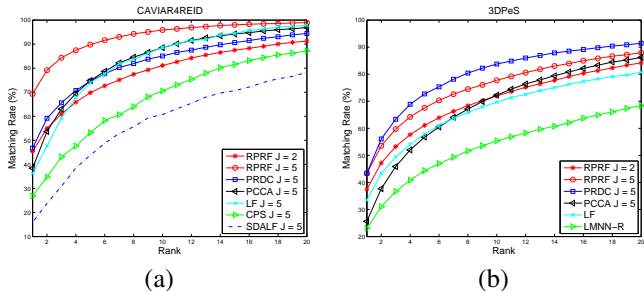

Figure 4. (a) Ranking performance on the CAVIAR4REID dataset. (b) Ranking performance on the 3DPeS dataset.

All of the above experiments demonstrate significant improvement when more images are available for the same person. The voting scheme allows results to be directly summed across multiple random forests, as well as across different vector pairs. On the other hand, the running time of the proposed algorithm is not affected by the sample size while other approaches suffer higher computational cost in either training or testing processes.

4.4. Personalized Random Forests

The final question is how the "personalized" random forests introduced in Section 3.4 improve the overall results. In this experiment, we randomly chose 5 images per person to form the probe set and another single image to form the gallery set. The multi-shot information in the probe set is used to select personalized random forests fol-

lowing Algorithm 2. When estimating the overall similarity between two people, we randomly chose 1 shot from each person in the probe set to match with the ones in the gallery set. In this way, we eliminate other factors that possibly contribute to the ranking performance. For each pair we computed the similarity score using both the personalized random forests and the same number of randomly selected random forests. The experiment was conducted on the CAVIAR4REID dataset, and the results are reported in Figure 5.

The results show that the ranking performance with personalized random forests is steadily better than that of using randomly selected random forests. Another advantage we noticed during the experiment is that the performance has smaller variance with this technique, since it filters out "less suitable" random forests.
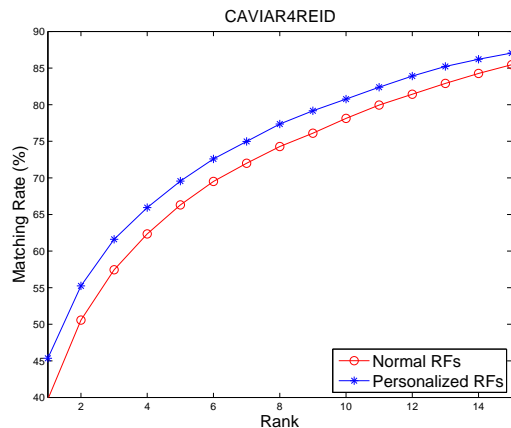

Figure 5. Improvement evaluation using personally discriminative random forests on CAVIAR4REID. 5% improvement can been seen at rank 1.

5. Conclusions

We proposed a novel approach to address the multi-shot human re-identification problem. An ensemble of random forests is built based on multiple random projections. The method substantially reduces the dimension of the data, is

time- and space-efficient, performs stably and robustly, and is simple to implement and tune. The proposed method demonstrates superior ability in the realistic multi-shot case, since the trained model can be further customized by multiple views of the target and the voting-based measurement is easy to aggregate. Experiments conducted on benchmarking datasets show that the proposed method outperforms current state-of-the-art multi-shot solutions for the re-id problem. In future work, we hope to gain a deeper understanding of the convergence behavior of aggregating multiple trained random forests. Also, we plan to further investigate the personally discriminative random forests.

References

[1] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7):1545–1588, 1997.

[2] S. Bak, G. Charpiat, E. Corvée, F. Brémond, and M. Thonnat. Learning to match appearances by correlations in a covariance metric space. In *ECCV*, 2012.

[3] S. Bak, E. Corvée, F. Brémond, and M. Thonnat. Person re-identification using Haar-based and DCD-based signature. In *AVSS*, 2010.

[4] S. Bak, E. Corvée, F. Bremond, and M. Thonnat. Boosted human re-identification using Riemannian manifolds. *Image and Vision Computing*, 30(6):443–452, 2012.

[5] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *J-HGBU*, 2011.

[6] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009.

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

[8] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In *ACM SIGKDD*, 2001.

[9] J. Blitzer, K. Q. Weinberger, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2005.

[10] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[11] D. Broomhead and M. Kirby. A new approach to dimensionality reduction: Theory and algorithms. *SIAM Journal on Applied Mathematics*, 60(6):2114–2142, 2000.

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

[13] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, 2007.

[14] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *ACCV*, 2011.

[15] X. Z. Fern and C. E. Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *ICML*, 2003.

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

[17] C. Hegde, M. Wakin, and R. Baraniuk. Random projections for manifold learning. In *NIPS*, 2008.

[18] M. Hirzer, P. M. Roth, and H. Bischof. Person re-identification by efficient impostor-based metric learning. In *AVSS*, 2012.

[19] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, 2012.

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

[21] B. Ma, Y. Su, F. Jurie, et al. Bicov: a novel image representation for person re-identification and face verification. In *BMVC*, 2012.

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

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

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

[25] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1619–1630, 2006.

[26] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[27] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *SIBGRAPI*, 2009.

[28] M. Skurichina and R. P. Duin. Combining feature subsets in feature selection. In *MCS*, 2005.

[29] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[30] K. Tumer and N. C. Oza. Input decimated ensembles. *Pattern Analysis & Applications*, 6(1):65–77, 2003.

[31] Z. Wu, Y. Li, and R. Radke. Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. In press.

[32] C. Xiong, D. Johnson, R. Xu, and J. J. Corso. Random forests for metric learning with implicit pairwise position dependence. In *ACM SIGKDD*, 2012.

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

[34] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *CVPR*, 2011.

[35] W.-S. Zheng, S. Gong, and T. Xiang. Transfer re-identification: From person to set-based verification. In *CVPR*, 2012.