Query Based Adaptive Re-Ranking for Person Re-Identification

Andy J Ma and Ping Li

Department of Statistics, Department of Computer Science,
Rutgers University, Piscataway, NJ 08854, USA

Abstract. Existing algorithms for person re-identification hardly model query variations across non-overlapping cameras. In this paper, we propose a query based adaptive re-ranking method to address this important issue. In our work, negative image pairs can be easily generated for each query under non-overlapping cameras. To infer query variations across cameras, nearest neighbors of the query positive match under two camera views are approximated and selected from positive matches in the training set. Locality preserving projections (LPP) are employed to ensure that each feature vector under one camera shares similar neighborhood structure with the corresponding positive match. Using existing re-identification algorithms as base score function, the optimal adaptive model is learnt by least-square regression with manifold regularization. Experimental results show that the proposed method can improve the ranking performance and outperforms other adaptive methods.

1 Introduction

The task of person re-identification is to re-identify a person when she/he disappears from the field of view of a camera and appears in another. The problem is very challenging due to non-trivial variations of viewpoint, illumination condition, human pose, etc. Existing methods solve these challenges by extracting features robust to these variations [1–10] or using label information to train discriminative models [11–17]. Since there are limited labeled images for each person and the query person who needs to be re-identified is usually not contained in the training set, most discriminative methods [11–16] assume all individuals share an unified model for identification. Based on this assumption, these methods generate matched (positive) and unmatched (negative) image pairs by limited person labels to train a score function. However, the learnt generic model may not be optimal for each query image, and consequently the re-identification performance is often not satisfactory.

To learn a matching function specific to each query image, Liu *et al.* [18] proposed an unsupervised approach to on-the-fly feature importance mining by person appearance attributes for re-identification. Based on manifold ranking, the score of the probe image is propagated to the gallery for performance improvement in [19]. Given the assumption that the transition time across cameras is available to prune the candidate set, Li *et al.* [20] proposed to learn an adaptive

metric by selecting and re-weighting the training data according to the query and pruned candidates. Using a different approach, in [21], feature vectors of a query-gallery image pair were first projected to a locally aligned space and then matched by specific local metrics. For post-rank optimization, weak negatives and strong negatives are manually selected from the gallery set to train an adaptive classification function for each query image in [22]. It was shown that the selected weak negatives and strong negatives can help improve the re-identification performance remarkably, but the process of manual labeling is still costly, especially when the number of query images becomes large.

In this paper, we propose a novel re-ranking method without using manually labeled information for the query data. Although negative image pairs can be easily generated for each query under non-overlapping cameras, it is more difficult and important to infer the query based positive information across cameras. To model query variations, positive image pairs in the training set are selected as nearest neighbors of the query positive match under two cameras. Since the query match is unknown, such nearest neighbors cannot be computed directly. Thus, we propose to approximate the neighborhood of the query match by employing Locality Preserving Projections (LPP) [23] to ensure that each image under the camera of the query shares the same neighborhood with the corresponding positive image pair across two cameras. Based on the available negative and estimated positive information for the query, the optimal adaptive re-ranking model is learnt by least-square regression with manifold regularization for the smoothness of the decision function.

Our contributions are summarized as follows:

1. We propose a novel query variation inference method to select nearest neighbors of the query match under two cameras based on images under one camera. Positive image pairs under two cameras in the training set are used to construct the adjacency graph for the training images under the camera of the query. A locality preserving mapping is learnt to preserve the neighborhood structure, so that nearest neighbors of the query match can be determined by the query image. Thus, query variations across cameras can be modeled by the positive image pairs in the training set corresponding to nearest neighbors of the query image.

2. We develop a new **Q**uery based **A**daptive **R**e-**R**anking (QARR) algorithm to improve the ranking performance of existing re-identification methods. Given a (generic) base score function, we learn a regression based adaptive re-ranking model by negative image pairs generated under non-overlapping cameras and the query match estimated by modeling query variations. To ensure the smoothness of the adaptive score function over the query-gallery image pairs, a manifold regularization term is incorporated in the objective function to learn the optimal QARR model.

We will first briefly review related work in Sec. 2. Then we elaborate on our proposed method in Sec. 3. Experimental results are given in Sec. 4. Finally, Sec. 5 concludes the paper.

2 Related Work

[24] proposed a descriptive and discriminative classification model for person re-identification. Given a specific query, all the images are ranked by appearance features of region covariance descriptors. After that, a human operator is assigned to check whether the searched person has a high rank. If not, a discriminative model is learnt for re-ranking under the assumption that there are multiple frames for the query under a camera. While it was shown that this query based re-ranking model could achieve better results than the generic models, the method did not model query variations across non-overlapping cameras.

Besides person re-identification, many adaptive re-ranking algorithms [25–29] have been developed for image retrieval. Most of these methods [25–28] first rank a query image by key word features and then re-rank the text-based search results by adaptive visual similarity measure. For content-based image retrieval [29], initial ranked lists are first determined by comparing the similarities between visual features and then images are re-ranked based on the similarities of their ranked lists as contextual information. While these methods are designed for image retrieval, they do not take advantage of the special characteristics in person re-identification under non-overlapping cameras.

Domain adaptation [30] is one of the research areas related to this paper. If we consider the training data as the source domain, the query image and the gallery set as the target domain, domain adaptation techniques can be employed to learn an adaptive classification model. Without any label information in the target domain, the unsupervised domain adaptation methods [31, 32] aim at aligning the marginal distributions under the assumption that the conditional probabilities are equal with each other in the source and target domain. Since the equal conditional probability assumption may not be valid, adaptive learning methods [33, 34] make use of a small amount of labeled data in the target domain to improve the recognition performance. Nevertheless, such label information is not available in query based learning for person re-identification, hence existing adaptive learning methods cannot be applied directly.

3 Proposed Method

We consider the re-identification task for images from a pair of cameras a and b. Denote feature vectors of images under cameras a and b as $\boldsymbol{x}_i^a$ and $\boldsymbol{x}_j^b$, respectively. As indicated in [15], the absolute difference space exhibits certain advantages over the common difference space. Hence we use the Absolute Difference Vector (ADV) $\boldsymbol{z}_{ij}$ as feature representation for each image pair:

$$\boldsymbol{z}_{ij} = (|\boldsymbol{x}_i^a(1) - \boldsymbol{x}_j^b(1)|, \cdots, |\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)|, \cdots, |\boldsymbol{x}_i^a(D) - \boldsymbol{x}_j^b(D)|)^T \quad (1)$$

where $\boldsymbol{x}(d)$ is the d-th element of feature vector $\boldsymbol{x}$ and D is the dimension of $\boldsymbol{x}$. Let the training data be $\boldsymbol{x}_i^a$ under camera a and $\boldsymbol{x}_j^b$ under camera b. With corresponding person labels y_i^a and y_j^b for training, positive and negative

ADVs can be generated and denoted by $\boldsymbol{z}_{ij}^{+}$ for $y_i^a = y_j^b$ and $\boldsymbol{z}_{mn}^{-}$ for $y_m^a \neq y_n^b$, respectively. Without loss of generality, suppose the query image come from camera a with feature vector $\boldsymbol{x}_q^a$. Let feature vector for gallery image g under camera b be $\boldsymbol{x}_g^b$. Since the same person cannot be presented at the same instant under different non-overlapping cameras a and b, negative ADVs $\boldsymbol{z}_{qg^-}$ can be obtained for each $\boldsymbol{x}_q^a$. Therefore, the key problem is to infer information about the positive ADV $\boldsymbol{z}_{qg^+}$ for query variation modeling across cameras. In Sec. 3.1, we present a query variations inference method. Based on the inferred information for query variations, an adaptive re-ranking model is reported in Sec. 3.2.

3.1 Cross-Cameras Query Variation Inference

Let the person image in the gallery set sharing the same identity with $\boldsymbol{x}_q^a$ be $\boldsymbol{x}_{g^+}^b$. The corresponding positive ADV for the query is $\boldsymbol{z}_{qg^+}$. With the positive ADVs $\boldsymbol{z}_{ij}^{+}$ in the training set, we propose to select some $\boldsymbol{z}_{ij}^{+}$ such that the distance between $\boldsymbol{z}_{ij}^{+}$ and $\boldsymbol{z}_{qg^+}$ is small. According to (1), the l_1 distance between two ADVs $\boldsymbol{z}_{ij}^{+}$ and $\boldsymbol{z}_{qg^+}$ is given by the following equation:

$$\|\boldsymbol{z}_{qg^+} - \boldsymbol{z}_{ij}^{+}\| = \sum_{d=1}^{D} \Big| |\boldsymbol{x}_q^a(d) - \boldsymbol{x}_{g^+}^b(d)| - |\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)| \Big| \tag{2}$$

If $(\boldsymbol{x}_q^a(d) - \boldsymbol{x}_{g^+}^b(d))(\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)) < 0$, then the element-wise difference on the right hand side of (2) becomes

$$\begin{aligned} &\Big| |\boldsymbol{x}_q^a(d) - \boldsymbol{x}_{g^+}^b(d)| - |\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)| \Big| \\ =&|(\boldsymbol{x}_q^a(d) + \boldsymbol{x}_i^a(d)) - (\boldsymbol{x}_{g^+}^b(d) + \boldsymbol{x}_j^b(d))| \end{aligned} \tag{3}$$

Since the variations between non-overlapping cameras a and b can be large, the right hand side of (3) is a large number. In this case, we cannot obtain positive ADVs $\boldsymbol{z}_{ij}^{+}$ from the training data, which are close to the positive ADV $\boldsymbol{z}_{qg^+}$ for the query. This implies that, in order to have small distance between $\boldsymbol{z}_{ij}^{+}$ and $\boldsymbol{z}_{qg^+}$, it is necessary to have $(\boldsymbol{x}_q^a(d) - \boldsymbol{x}_{g^+}^b(d))(\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)) \geq 0$, which means

$$\begin{aligned} &\Big| |\boldsymbol{x}_q^a(d) - \boldsymbol{x}_{g^+}^b(d)| - |\boldsymbol{x}_i^a(d) - \boldsymbol{x}_j^b(d)| \Big| \\ =&|(\boldsymbol{x}_q^a(d) - \boldsymbol{x}_i^a(d)) + (\boldsymbol{x}_j^b(d) - \boldsymbol{x}_{g^+}^b(d))| \\ \leq&|\boldsymbol{x}_q^a(d) - \boldsymbol{x}_i^a(d)| + |\boldsymbol{x}_j^b(d) - \boldsymbol{x}_{g^+}^b(d)| \end{aligned} \tag{4}$$

According to (4), we have an upper bound for $\|\boldsymbol{z}_{qg^+} - \boldsymbol{z}_{ij}^{+}\|$, i.e.

$$\|\boldsymbol{z}_{qg^+} - \boldsymbol{z}_{ij}^{+}\| \leq \|\boldsymbol{x}_q^a - \boldsymbol{x}_i^a\| + \|\boldsymbol{x}_j^b - \boldsymbol{x}_{g^+}^b\| \tag{5}$$

By (5), it is reasonable to see that if $\boldsymbol{x}_i^a$ is close to $\boldsymbol{x}_q^a$ and $\boldsymbol{x}_j^b$ close to $\boldsymbol{x}_{g^+}^b$ for $y_i^a = y_j^b$, the distance between $\boldsymbol{z}_{ij}^{+}$ in the training set and $\boldsymbol{z}_{qg^+}$ for the query is

small. Therefore, we can obtain the information about the positive ADV $\boldsymbol{z}_{qg^+}$ by the intersection of the neighborhood of $\boldsymbol{x}_q^a$ and the one of $\boldsymbol{x}_{g^+}^b$.

Although the feature vector $\boldsymbol{x}_{g^+}^b$ under camera b corresponding to the query under camera a is unknown, we can select the corresponding positive ADVs from the training data by the feature vector of the query $\boldsymbol{x}_q^a$. Since $\boldsymbol{x}_q^a$ and $\boldsymbol{x}_{g^+}^b$ are feature vectors for the same person under different camera views, they must be related and it is reasonable to assume that $\boldsymbol{x}_{g^+}^b$ can be obtained by a mapping Φ on $\boldsymbol{x}_q^a$, i.e., $\boldsymbol{x}_{g^+}^b = \Phi(\boldsymbol{x}_q^a)$. Although we may not be able to determine such Φ due to limited size of available training data, we make use of this assumption as follows. Applying Φ on $\boldsymbol{x}_i^a$ for $y_i^a = y_j^b$, we get $\boldsymbol{x}_j^b = \Phi(\boldsymbol{x}_i^a)$. Therefore, the second term on the right hand side of (5) becomes

$$\|\boldsymbol{x}_j^b - \boldsymbol{x}_{g^+}^b\| = \|\Phi(\boldsymbol{x}_i^a) - \Phi(\boldsymbol{x}_q^a)\| \tag{6}$$

If Φ is a continuously differentiable function, the right hand side of (6) is bounded by the following equation according to mean value theorem [35],

$$\|\Phi(\boldsymbol{x}_i^a) - \Phi(\boldsymbol{x}_q^a)\| \leq \|\boldsymbol{J}(\Phi)\|\|\boldsymbol{x}_i^a - \boldsymbol{x}_q^a\| \tag{7}$$

where $\boldsymbol{J}$ denotes the Jacobian matrix of all first-order partial derivatives of mapping function Φ. With (6) (7), the inequality (5) becomes

$$\|\boldsymbol{z}_{qg^+} - \boldsymbol{z}_{ij}^+\| \leq (1 + \|\boldsymbol{J}(\Phi)\|)\|\boldsymbol{x}_q^a - \boldsymbol{x}_i^a\| \tag{8}$$

Therefore, if $\boldsymbol{x}_i^a$ is a neighbor of $\boldsymbol{x}_q^a$, the positive ADV $\boldsymbol{z}_{ij}^+$ in the training set is a neighbor of $\boldsymbol{z}_{qg^+}$ for the query. This means the neighborhood of the positive ADV $\boldsymbol{z}_{qg^+}$ can be determined by the neighborhood of the feature vector $\boldsymbol{x}_q^a$.

According to (8), the upper bound of the distance between two ADVs can be determined by the distance between two feature vectors $\boldsymbol{x}_q^a$ and $\boldsymbol{x}_i^a$ under the same camera. Thus, we propose to construct the neighborhood of the query image pair by the neighborhood of the query image. Although it is plausible to directly select the nearest neighbors of the query image, we need to consider that the neighborhood structures are different in the image pair and image spaces. Thus, we propose to employ Locality Preserving Projections (LPP) [23] to align such differences by learning a projection matrix P.

For each feature vector $\boldsymbol{x}_i^a$ under camera a in the training set, we compute the corresponding positive ADV as

$$\boldsymbol{z}_i^+ = \frac{1}{N_i^+} \sum_{y_j^b = y_i^a} \boldsymbol{z}_{ij}^+ \tag{9}$$

where N_i^+ is the number of positive matches for $\boldsymbol{x}_i^a$. To construct the weight matrix A, k nearest neighbors are selected for each positive ADV $\boldsymbol{z}_i^+$. Then, the simple-minded weighting scheme is employed to determine the weight between $\boldsymbol{z}_i^+$ and $\boldsymbol{z}_{i'}^+$. In other words, if $\boldsymbol{z}_i^+$ is in the neighborhood of $\boldsymbol{z}_{i'}^+$, or $\boldsymbol{z}_{i'}^+$ is in the neighborhood of $\boldsymbol{z}_i^+$, $A_{ii'} = 1$; otherwise, $A_{ii'} = 0$. Thus, the neighborhood

Algorithm 1 Cross-Camera Query Variation Inference

Input: Feature vectors $\boldsymbol{x}_i^a$ under camera a and positive ADVs $\boldsymbol{z}_{ij}^+$ in the training set, query feature vector $\boldsymbol{x}_q^a$, projection dimension p, neighborhood parameters k for LPP and k_q for query based positive ADV;

1: Compute positive ADV $\boldsymbol{z}_i^+$ by (9) for each $\boldsymbol{x}_i^a$;

2: Construct k nearest neighbors for each $\boldsymbol{z}_i^+$ to obtain weight matrix A;

3: Solve optimization problem (10) to obtain projection P with dimension p;

4: Compute the distances between the projected feature vector $P^T\boldsymbol{x}_q^a$ for the query and the projected feature vectors $P^T\boldsymbol{x}_i^a$ for the training data;

5: Select k_q nearest neighbors $P^T\boldsymbol{x}_{i_1}^+, \cdots, P^T\boldsymbol{x}_{i_{k_q}}^+$ of $P^T\boldsymbol{x}_q^a$;

6: Calculate the estimation $\tilde{\boldsymbol{z}}_{qg^+}$ for the query positive ADV by (11);

Output: Estimated query positive ADV $\tilde{\boldsymbol{z}}_{qg^+}$.

information for the positive ADVs $\boldsymbol{z}_i^+$ is enclosed in the weight matrix A. We would like to learn a projection matrix such that the indexes of the neighbors of $\boldsymbol{x}_i^a$ are nearly the same as those of $\boldsymbol{z}_i^+$. We use A to define the objective function for feature vectors $\boldsymbol{x}_i^a$ as follows,

$$\min_{\boldsymbol{e},\ \text{s.t. } \sum_{i,i'} A_{ii'}(\boldsymbol{e}^T\boldsymbol{x}_i^a)^2=1} \sum_{i,i'} A_{ii'}(\boldsymbol{e}^T\boldsymbol{x}_i^a - \boldsymbol{e}^T\boldsymbol{x}_{i'}^a)^2 \tag{10}$$

where $\boldsymbol{e}$ denotes the column vector in P. The optimization problem (10) can be solved by calculating the eigenvectors and eigenvalues for the generalized eigenvalue problem. The projection matrix P is obtained by the eigenvectors corresponding to the first p eigenvalues (details can be referred to [23]).

Based on the above analysis, we infer the cross-camera query variations by selecting k_q nearest neighbors $P^T\boldsymbol{x}_{i_1}^+, \cdots, P^T\boldsymbol{x}_{i_{k_q}}^+$ of $P^T\boldsymbol{x}_q^a$ from the training data. The corresponding positive ADVs $\boldsymbol{z}_{i_1}^+, \cdots, \boldsymbol{z}_{i_{k_q}}^+$ in the training set are used to represent $\boldsymbol{z}_{qg^+}$ for the query. Since the assumption that $\boldsymbol{x}_j^b = \Phi(\boldsymbol{x}_i^a)$ may not be satisfied for all the selected positive ADVs $\boldsymbol{z}_{i_1}^+, \cdots, \boldsymbol{z}_{i_{k_q}}^+$, we compute the mean of them for the estimation of $\boldsymbol{z}_{qg^+}$, i.e.,

$$\tilde{\boldsymbol{z}}_{qg^+} = \frac{1}{k_q}(\boldsymbol{z}_{i_1}^+ + \cdots + \boldsymbol{z}_{i_{k_q}}^+) \tag{11}$$

Algorithm 1 lists the procedure for cross-camera query variation inference.

3.2 Adaptive Regression with Graph Propagation for Re-Ranking

Given a (generic) base score function f for feature vectors $\boldsymbol{x}_q^a$ of the query and $\boldsymbol{x}_g^b$ of the gallery image, we learn an adaptive function f_q specific to the query.

Inspired by adaptive learning methods [33, 34] for domain adaptation, we define

$$f_q(\boldsymbol{x}_q^a, \boldsymbol{x}_g^b) = \theta f(\boldsymbol{x}_q^a, \boldsymbol{x}_g^b) + \boldsymbol{w}^T \boldsymbol{z}_{qg} \tag{12}$$

where $\boldsymbol{z}_{qg}$ denote the ADV between feature vectors $\boldsymbol{x}_q^a$ of the query and $\boldsymbol{x}_g^b$ of a gallery image as defined in (1), θ is a positive parameter to measure the importance of the base score function f and $\boldsymbol{w}$ is the perturbation weight vector adapted for the query.

With the estimated query positive ADV $\tilde{\boldsymbol{z}}_{qg^+}$ and negative ADVs $\boldsymbol{z}_{qg^-}$ generated under non-overlapping cameras, we formulate the objective function in a least-square regression framework. Since the score of the positive image pair must be larger than the negative ones, we set $f_q(\boldsymbol{x}_q^a, \boldsymbol{x}_{g^+}^b) - f_q(\boldsymbol{x}_q^a, \boldsymbol{x}_{g^-}^b) \approx 1$. This way, we can formulate the following optimization problem:

$$\min_{\theta, \boldsymbol{w}} \frac{1}{N_q^-} \sum_{g^-} [\boldsymbol{w}^T (\tilde{\boldsymbol{z}}_{qg^+} - \boldsymbol{z}_{qg^-}) + \theta(\tilde{s}_{qg^+} - s_{qg^-}) - 1]^2 + \lambda \boldsymbol{w}^T \boldsymbol{w} + \mu\theta^2 \tag{13}$$

where $\tilde{s}_{qg^+} = \frac{1}{k_q} \sum_{t=1}^{k_q} \frac{1}{N_i^+} \sum_{y_j^b = y_{i_t}^a} f(\boldsymbol{x}_{i_t}^a, \boldsymbol{x}_j^b)$, $s_{qg^-} = f(\boldsymbol{x}_q^a, \boldsymbol{x}_{g^-}^b)$, N_q^- is the number of negative image pairs for the query, λ and μ are positive parameters for the regularization terms of $\boldsymbol{w}$ and θ, respectively. To solve the optimization problem (13), we convert it to a matrix form as,

$$\begin{aligned} &\min_{\bar{\boldsymbol{w}}} \bar{\boldsymbol{w}}^T M \bar{\boldsymbol{w}} - 2\bar{\boldsymbol{w}}^T \boldsymbol{m} + \bar{\boldsymbol{w}}^T M_r \bar{\boldsymbol{w}} \\ &\bar{\boldsymbol{w}} = \begin{pmatrix} \boldsymbol{w} \\ \theta \end{pmatrix}, M = \frac{1}{N_{q-}} \sum_{g^-} \begin{pmatrix} \tilde{\boldsymbol{z}}_{qg^+} - \boldsymbol{z}_{qg^-} \\ \tilde{s}_{qg^+} - s_{qg^-} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{z}}_{qg^+} - \boldsymbol{z}_{qg^-} \\ \tilde{s}_{qg^+} - s_{qg^-} \end{pmatrix}^T, \\ &\boldsymbol{m} = \frac{1}{N_{q-}} \sum_{g^-} \begin{pmatrix} \tilde{\boldsymbol{z}}_{qg^+} - \boldsymbol{z}_{qg^-} \\ \tilde{s}_{qg^+} - s_{qg^-} \end{pmatrix}, M_r = \begin{pmatrix} \lambda I & \boldsymbol{0} \\ \boldsymbol{0} & \mu \end{pmatrix} \end{aligned} \tag{14}$$

where I is the unit matrix with the same dimension as $\boldsymbol{w}$.

Note that, if two ADVs $\boldsymbol{z}_{qg}$ and $\boldsymbol{z}_{qg'}$ are close to each other, they must have similar matching scores. Thus, we employ manifold regularization [36] in our adaptive re-ranking method. A weight matrix A_q is constructed for the ADVs $\boldsymbol{z}_{q1}, \cdots, \boldsymbol{z}_{qN_G}, \boldsymbol{z}_{q(N_G+1)}$, where N_G is the number of images in the gallery set and $\boldsymbol{z}_{q(N_G+1)} = \tilde{\boldsymbol{z}}_{qg^+}$. For each query-gallery ADV $\boldsymbol{z}_{qg}$, k_m nearest neighbors are selected and the weights are determined by the simple-minded weighting scheme to avoid parameter selection in heat kernel as described in the previous section. Then, the manifold based regularization term for the continuity of the query score function f_q is given as follows,

$$\frac{1}{(N_G+1)^2} \sum_{g=1}^{N_G+1} \sum_{g'=1}^{N_G+1} A_{qgg'} [(\boldsymbol{w}^T \boldsymbol{z}_{qg} + \theta s_{qg}) - (\boldsymbol{w}^T \boldsymbol{z}_{qg'} + \theta s_{qg'})]^2 \tag{15}$$

where $s_{q(N_G+1)} = \tilde{s}_{qg^+}$. Denote the column concatenation of $\boldsymbol{z}_{qg}$ and s_{qg} by $\bar{\boldsymbol{z}}_{qg}$, and $Z_q = (\bar{\boldsymbol{z}}_{q1}, \cdots, \bar{\boldsymbol{z}}_{q(N_G+1)})$, respectively. Adding the regularization term (15)

Algorithm 2 Training Query Score Function

Input: ADVs $\boldsymbol{z}_{qg}$ for query-gallery image pairs, estimated query positive ADV $\tilde{\boldsymbol{z}}_{qg^+}$, negative ADVs $\boldsymbol{z}_{qg^-}$ under non-overlapping cameras, base scores $\tilde{s}_{qg^+}, s_{q1}, \cdots, s_{qN_G}$, parameters λ, μ, η, k_m;

1: Compute $M, M_r, \boldsymbol{m}$ by (14);

2: Construct k_m nearest neighbors for each $\boldsymbol{z}_{qg}$ to obtain weight matrix A_q;

3: Calculate the normalized Laplacian matrix L_q by (16);

4: Obtained the optimal augmented weight vector $\bar{\boldsymbol{w}}^*$ by (17);

Output: Optimal weights $\boldsymbol{w}^*$ and θ^* for the query score function f_q.

(in matrix form) into (14), the optimization problem becomes

$$\begin{aligned} &\min_{\bar{\boldsymbol{w}}} \bar{\boldsymbol{w}}^T M \bar{\boldsymbol{w}} - 2\bar{\boldsymbol{w}}^T \boldsymbol{m} + \bar{\boldsymbol{w}}^T M_r \bar{\boldsymbol{w}} + \bar{\boldsymbol{w}}^T Z_q L_q Z_q^T \bar{\boldsymbol{w}}, \\ &\text{s.t. } L_q = \frac{\eta(D_q - A_q)}{(N_G+1)^2} \end{aligned} \tag{16}$$

where η is a positive parameter for the manifold based regularization term and D_q is a diagonal matrix with diagonal element $D_{qgg} = \sum_{g'} A_{qgg'}$. The optimization problem (16) can be solved by setting the first derivative of the objective function to zero. The solution is given by

$$\bar{\boldsymbol{w}}^* = (M + M_r + Z_q L_q Z_q^T)^{-1} \boldsymbol{m} \tag{17}$$

According to the definition of $\bar{\boldsymbol{w}}$ in (14), the optimal $\boldsymbol{w}^*$ and θ^* can be obtained to determine the query score function defined in (12).

Algorithm 2 summarizes the algorithmic procedure for training the Query based Adaptive Re-Ranking (QARR) model.

4 Experiments

We first introduce the datasets and settings for experiments. Then we present the results on query variation inference across cameras in Sec. 4.2. Based on the inferred query variations, we demonstrate that our method can improve the ranking performance for person re-identification in Sec. 4.3. Finally, we compare our method with existing adaptive re-identification algorithms in Sec. 4.4.

4.1 Datasets and Settings

Two publicly available datasets, namely VIPeR[1] [37] and CUHK[2] [21], are used for experiments. Example images in these two datasets are shown in Fig. 1 and

[1] `http://soe.ucsc.edu/~dgray/VIPeR.v1.0.zip`

[2] `http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html`

Fig. 1. Examples of $k_q = 5$ nearest neighbors obtained by Algorithm 1 on VIPeR [37] dataset (better viewed in color).

Fig. 2, respectively. VIPeR is a re-identification dataset containing 632 person image pairs captured by two cameras outdoor. In this dataset, 632 image pairs are randomly separated into half for training and the other half for testing. CUHK dataset contains five pairs of camera views. Under each camera view, there are two images for each person. Following the single shot setting in [21], images from camera pair one with 971 persons are used for experiments. For this dataset, 971 persons are randomly split into 485 for training and 486 for testing. For the testing data in VIPeR or CUHK, the evaluation is performed by searching the 316 or 486 persons in one camera view from another view. Ten negative image pairs are randomly generated for each query image. These experiments were performed ten times and the average results are reported. For feature representation, we follow [11, 12, 15] and divide a person image into 6

Fig. 2. Examples of $k_q = 5$ nearest neighbors obtained by Algorithm 1 on CUHK [21] dataset (better viewed in color).

horizontal stripes and compute the RGB, YCbCr, HSV color features and two types of texture features extracted by Schmid and Gabor filters on each stripe.

In our experiments, we implemented three state-of-the-art algorithms, namely Ranking Support Vector Machines (RSVM) [12], Relaxed Pairwise Metric Learning (RPML) [14] and Relative Distance Comparison (RDC) [15], and use each as the base score function. The parameter C in RSVM is empirically set as 1, while the PCA dimension in RPML is set as 80 for robust performance. To avoid singular matrix problem, we perform PCA with dimension 80 before learning the projection matrix P in our method. The parameters for neighborhood construction are set as $k = k_q = k_m = 5$. For the regularization parameters, if λ is too large, the norm of the adaptive weight $\boldsymbol{w}$ will be very small. In this case, the query score function f_q will be very close to the base score function f. On the other hand, if λ is too small, the norm of $\boldsymbol{w}$ will be very large, which

implies the model could be over-fitted and the base score function hardly affects the decision for the query. Similar analysis can be applied to parameter μ for the weight of the base score function. Thus, we empirically set $\lambda = 10^{-2}$ and $\mu = 10^{-3}$. Since η is the parameter to measure the importance of the manifold based regularization term, we set it with a larger number as $\eta = 10^{-1}$.

4.2 Results on Cross-Camera Query Variation Inference

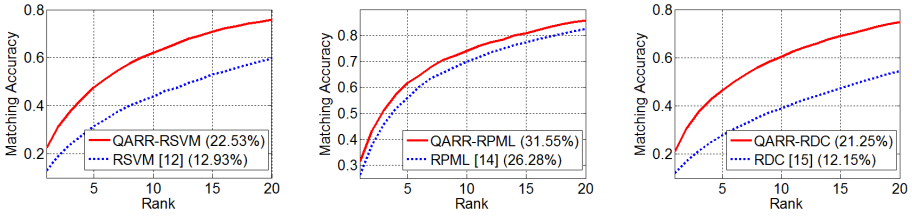
In this section, we first evaluate whether the proposed cross-camera query variation inference method can discover the true neighborhood of the positive ADV $\boldsymbol{z}_{qg^+}$. For evaluation, $k_q = 5$ nearest neighbors of $\boldsymbol{z}_{qg^+}$ are selected from the positive ADVs in the training set as ground truth. We calculate the intersection ratio given by the number of elements in the intersection set of the true neighborhood and the one constructed by Algorithm 1 divided by $k_q = 5$. The intersection ratios averaged over all the query images are 59.18% on VIPeR dataset and 58.89% on CUHK dataset, respectively. This means that on average nearly 3 out of 5 nearest neighbors are correctly detected by Algorithm 1 for the query positive ADV $\boldsymbol{z}_{qg^+}$. Since the majority of the detected nearest neighbors are in the true neighborhood of $\boldsymbol{z}_{qg^+}$, the inferred query variations can help improve the re-identification performance across cameras.

To visualize the query based positive inference results, we show the true nearest neighbors and the ones selected by Algorithm 1 for three query images under camera a in Fig. 1 for VIPeR and Fig. 2 for CUHK dataset. The image pairs in the intersection of the nearest neighbor sets are marked in the same color. From the first to the fourth rows in Fig. 1 and Fig. 2, we can see that 3 nearest neighbors selected by Algorithm 1 are in the true neighborhood of the query match, which is approximately equal to the average intersection ratios. Since we do not know which nearest neighbors are correctly selected, we compute the mean of them by (11) to reduce the error caused by incorrect selection. It is also possible that the intersection of the selected nearest neighbors and the true ones is an empty set as illustrated in the last two rows in Fig. 1 and Fig. 2. Although the order of the positive image pairs in the training set computed by Algorithm 1 may not be the same as the true one, the selected image pairs still look similar to the query ones, e.g., similar jackets, pants, and/or pose under the same camera. Therefore, the selected positive ADVs can still help improve the ranking performance, which will be shown in the following subsection.

4.3 Results on Query Based Adaptive Re-Ranking

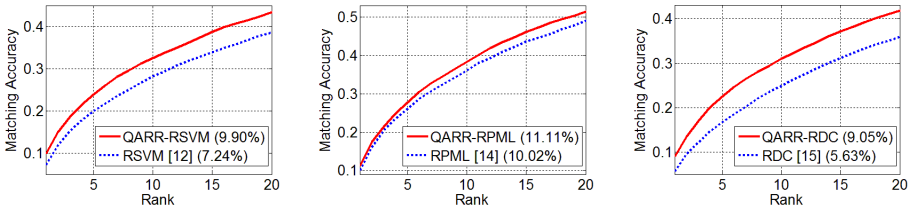
The CMC curves of the proposed Query based Adaptive Re-Ranking (QARR) method are compared with those of RSVM, RPML and RDC in Fig. 3(a)-3(c) on VIPeR and Fig. 4(a)-4(c) on CUHK dataset. From these figures, we can see that our method outperform RSVM, RPML and RDC on both datasets by learning a score function specific to the query. Results in Fig. 3 show that the rank one accuracy of our method using RSVM, RPML or RDC as the base score function is over 5% higher than that without adaptive learning on VIPeR

dataset. Interestingly, although the RPML should be a better score function compared with RSVM and RDC on these two datasets, our method can still achieve higher matching accuracies with different numbers of top ranks based on it. In other words, regardless of the base score function, our adaptive learning method can improve the re-identification performance robustly.



(a) (b) (c)

Fig. 3. CMC curves of our method using (a) RSVM [12], (b) RPML [14] or (c) RDC [15] as base score function on VIPeR [37] dataset with 316 image pairs for training.



(a) (b) (c)

Fig. 4. CMC curves of our method using (a) RSVM [12], (b) RPML [14] or (c) RDC [15] as base score function on CUHK [21] dataset with 485 image pairs for training.

Note that the proposed method implicitly assumes that there are positive image pairs in the training set which are similar to the query positive match. When the number of training image pairs increases, this assumption will more easily satisfied and we should observe better improvement of the ranking performance. To verify this argument, we increase the number of persons for training from 316 to 500 on VIPeR and 485 to 700 on CUHK dataset. The CMC curves on VIPeR dataset in Fig. 5 show that the rank one accuracy improvement by our method is increased from 9.60% to 14.70% using base score function RSVM, from 5.27% to 6.06% using RPML and from 9.10% to 15.94% using RDC. Similar statistics can be observed on CUHK dataset in Fig. 6. These results confirm that our method can achieve better improvement with more training data.

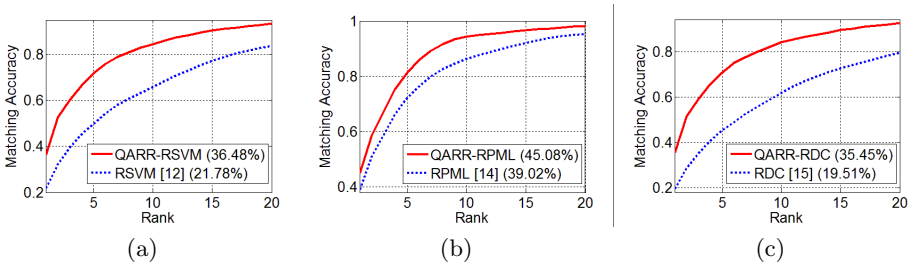


Fig. 5. CMC curves of our method using (a) RSVM [12], (b) RPML [14] or (c) RDC [15] as base score function on VIPeR [37] dataset with 500 image pairs for training.

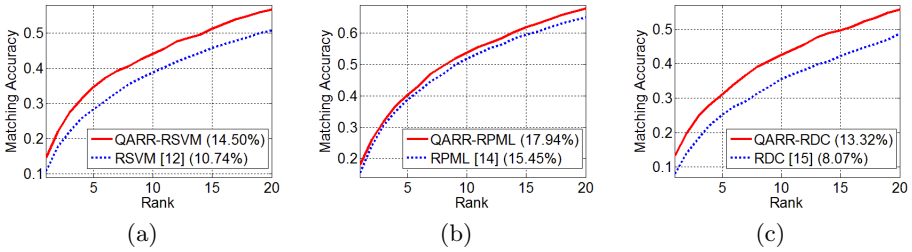


Fig. 6. CMC curves of our method using (a) RSVM [12], (b) RPML [14] or (c) RDC [15] as base score function on CUHK [21] dataset with 700 image pairs for training.

4.4 Comparison with Existing Adaptive Re-Ranking Methods

In this section, we compare our method with other query based re-identification algorithms namely, Prototype-Specific Feature Importance (PSFI) [18], Individual-Specific Feature Importance (ISFI) [18], Manifold Ranking with Normalised graph Laplacian (MRNL) [19] and Manifold Ranking with Unnormalised iterated graph Laplacian (MRUL) [19]. The results are copied from their papers and recorded in Table 1 for comparison. It is shown in Table 1 that all the query based re-ranking methods can achieve higher matching accuracy by learning a score function specific to the query. Comparing our method with PSFI and ISFI, we can see that our method remarkably outperforms them using either RSVM or RDC as the base score function. The rank one accuracy of our method is over 6% higher than those of the PSFI and ISFI based on RSVM and over 4% higher than them based on RDC. Furthermore, our method can also outperforms the manifold ranking algorithms, MRNL and MRUL, by modeling the inter-camera variations specific to the probe image.

Table 1. Top rank matching accuracy (%) on VIPeR

Method \ Rank	1	5	10	15	20
QARR-RSVM	**22.53**	**47.59**	**62.20**	**70.85**	**75.82**
MRNL-RSVM [19]	19.27	42.41	55.00	63.86	70.06
MRUL-RSVM [19]	19.34	42.47	55.51	64.11	70.44
PSFI-RSVM [18]	15.76	38.70	51.36	n/a	66.84
ISFI-RSVM [18]	16.46	38.76	51.36	n/a	67.18
RSVM [12]	12.93	31.46	43.91	53.05	59.64
QARR-RDC	**21.15**	**46.46**	**60.47**	**68.94**	**74.84**
MRNL-RSVM [19]	19.37	42.78	54.78	63.77	69.62
MRUL-RSVM [19]	18.45	41.74	53.67	62.72	69.27
PSFI-RDC [18]	16.99	38.10	52.37	n/a	66.84
ISFI-RDC [18]	17.12	38.96	52.94	n/a	67.34
RDC [15]	12.15	27.78	38.94	47.36	54.46

5 Conclusions

In this paper, we have developed a Query based Adaptive Re-Ranking (QARR) method to learn a discriminative model specific to the query data. Negative image pairs can be generated for the query under non-overlapping cameras, while positive information about the query across cameras is inferred by approximating the neighborhood of the query positive match. By analyzing the distance between two positive ADVs, we show that such neighborhood can be determined by the nearest neighbors of the query feature vector. Locality Preserving Projection (LPP) [23] is employed to ensure the similarity of the neighborhood structures between the ADV space and feature vector space under the camera of the query. Given a base score function, a regression based adaptive re-ranking model is learnt by propagating the negative and estimated positive information about the query match to all the query-gallery image pairs.

Experimental results show that the majority of the nearest neighbors selected by our method are in the true neighborhood of the query positive match. Thus, the QARR method can improve the ranking performance of existing re-identification methods by using the positive matching information of the query across cameras. Compared with other adaptive methods for person re-identification, our method achieves the best results on VIPeR dataset.

Acknowledgement

The work is supported in part by ONR-N00014-13-1-0764, NSF-III-1360971, AFOSR-FA9550-13-1-0137, and NSF-Bigdata-1419210.

References

1. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR. (2010)
2. Bauml, M., Stiefelhagen, R.: Evaluation of local features for person re-identification in image sequences. In: AVSS. (2011)
3. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC. (2011)
4. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: Appearance-based person reidentification in camera networks: problem overview and current approaches. JAIHC **2** (2011) 127–151
5. Jungling, K., Arens, M.: View-invariant person re-identification with an implicit shape model. In: AVSS. (2011)
6. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recognition Letters **33** (2012) 898–903
7. Bąk, S., Charpiat, G., Corvée, E., Brémond, F., Thonnat, M.: Learning to match appearances by correlations in a covariance metric space. In: ECCV. (2012)
8. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: ECCV Workshop. (2012)
9. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. TPAMI **35** (2013) 1622–1634
10. Xu, Y., Lin, L., Zheng, W.S., Liu, X.: Human re-identification by matching compositional template with cluster sampling. In: ICCV. (2013)
11. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. (2008)
12. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVC. (2010)
13. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: ECCV Workshop. (2012)
14. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: ECCV. (2012)
15. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. TPAMI **35** (2013) 653–668
16. Ma, A.J., Yuen, P.C., Li, J.: Domain transfer support vector ranking for person re-identification without target camera label information. In: ICCV. (2013)
17. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching. In: ICCV. (2013)
18. Liu, C., Gong, S., Loy, C.C.: On-the-fly feature importance mining for person re-identification. Pattern Recognition **47** (2014) 1602–1615
19. Loy, C.C., Liu, C., Gong, S.: Person re-identification by manifold ranking. In: ICIP. (2013)
20. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: ACCV. (2012)
21. Li, W., Wang, X.: Locally aligned feature transforms across views. In: CVPR. (2013)
22. Liu, C., Loy, C.C., Gong, S., Wang, G.: POP: Person re-identification post-rank optimisation. In: ICCV. (2013)
23. He, X., Niyogi, P.: Locality preserving projections. In: NIPS. (2003)

24. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: SCIA. (2011)
25. Cui, J., Wen, F., Tang, X.: Real time google and live image search re-ranking. In: ACM MM. (2008)
26. Zitouni, H., Sevil, S., Ozkan, D., Duygulu, P.: Re-ranking of web image search results using a graph algorithm. In: ICPR. (2008)
27. Jain, V., Varma, M.: Learning to re-rank: query-dependent image re-ranking using click data. In: ACM WWW. (2011)
28. Wang, X., Liu, K., Tang, X.: Query-specific visual semantic spaces for web image re-ranking. In: CVPR. (2011)
29. Pedronette, D.C.G., da S Torres, R.: Image re-ranking and rank aggregation based on similarity of ranked lists. Pattern Recognition **46** (2013) 2350–2360
30. Pan, S.J., Yang, Q.: A survey on transfer learning. TKDE **22** (2010) 1345–1359
31. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: ICCV. (2011)
32. Pan, S.J., Ivor W. Tsang, J.T.K., Yang, Q.: Domain adaptation via transfer component analysis. TNN **22** (2011) 199–210
33. Yang, J., Yan, R., Hauptmann, A.G.: Cross-domain video concept detection using adaptive svms. In: ACM MM. (2007)
34. Duan, L., Xu, D., Tsang, I.H., Luo, J.: Visual event recognition in videos by learning from web data. TPAMI **34** (2012) 1667–1680
35. Rudin, W.: Principles of mathematical analysis. McGraw-Hill (1976)
36. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. JMLR **7** (2006) 2399–2434
37. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: PETS. (2007)