# Person Reidentification via Ranking Aggregation of Similarity Pulling and Dissimilarity Pushing

Mang Ye, Chao Liang, Yi Yu, Zheng Wang, Qingming Leng, Chunxia Xiao, *Member, IEEE*, Jun Chen, and Ruimin Hu, *Senior Member, IEEE*

***Abstract*—Person reidentification is a key technique to match different persons observed in nonoverlapping camera views. Many researchers treat it as a special object-retrieval problem, where ranking optimization plays an important role. Existing ranking optimization methods mainly utilize the similarity relationship between the probe and gallery images to optimize the original ranking list, but seldom consider the important dissimilarity relationship. In this paper, we propose to use both similarity and dissimilarity cues in a ranking optimization framework for person reidentification. Its core idea is that the true match should not only be similar to those strongly similar galleries of the probe, but also be dissimilar to those strongly dissimilar galleries of the probe. Furthermore, motivated by the philosophy of multiview verification, a ranking aggregation algorithm is proposed to enhance the detection of similarity and dissimilarity based on the following assumption: the true match should be similar to the probe in different baseline methods. In other words, if a gallery blue image is strongly similar to the probe in one method, while simultaneously strongly dissimilar to the probe in another method, it will probably be a wrong match of the probe. Extensive experiments conducted on public benchmark datasets and comparisons with different baseline methods have shown the great superiority of the proposed ranking optimization method.**

***Index Terms*—Person reidentification, ranking aggregation, similarity and dissimilarity.**

Manuscript received January 30, 2016; revised June 29, 2016; accepted August 22, 2016. Date of publication August 31, 2016; date of current version November 15, 2016. This work was supported in part by the National Nature Science Foundation of China under Grant 61303114, Grant 61501413, Grant 61562048, and Grant 61472288, in part by the National High Technology Research and Development Program of China under Grant 2015AA016306, in part by the Internet of Things Development Funding Project of Ministry of Industry in 2013 (25), in part by the Technology Research Project of Ministry of Public Security under Grant 2014JSYJA016, in part by the Nature Science Foundation of Jiangsu Province under Grant BK20160386, in part by the Nature Science Foundation of Hubei Province under Grant 2014CFB712, in part by the Nature Science Foundation of Jiangxi Province under Grant 20151BAB217013, and in part by the Fundamental Research Funds for the Central Universities under Grant 2042014kf0250. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Winston Hsu. *(Corresponding author: Chao Liang.)*

M. Ye was with State Key Laboratory of Software Engineering, Collaborative Innovation Center of Geospatial Technology, and the National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430072, China. He is now with Department of Computer Science, Hong Kong Baptist University, Hong Kong, China (e-mail: yemang@whu.edu.cn).

C. Liang, Z. Wang, J. Chen, and R. Hu are with the State Key Laboratory of Software Engineering, Collaborative Innovation Center of Geospatial Technology, and the National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430072, China (e-mail: cliang@whu.edu.cn; wangzwhu@whu.edu.cn; chenj@whu.edu.cn; hurm1964@gmail.com).

Y. Yu is with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: yiyu@nii.ac.jp).

Q. Leng is with the School of Information Science and Technology, Jiujiang University, Jiujiang 332005, China (e-mail: lengqingming@126.com).

C. Xiao is with the State Key Lab of Software Engineering, and the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: cxxiao@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2016.2605058

## I. Introduction

In recent years, the person re-identification problem, namely matching people across disjoint camera views in a multi-camera system, has aroused an increasing interest in both computer vision and multimedia analysis communities [1], [2]. A direct application of person re-identification is that we can find out a common person target in various cameras, which is especially important in criminal investigation. Besides, it also underpins many advanced multimedia applications, such as person retrieval [3], [4], movement analysis [5]–[8], long term object tracking [9], [10] and personalized applications [11]–[13]. The main challenge of person re-identification can be attributed to the significant visual changes in various pose, illumination and viewpoint conditions, making intra-personal variations even larger than inter-personal ones. Moreover, background clutters and occlusions cause additional difficulties. As traditional biometrics, such as face and gait, are unreliable or even infeasible to be robustly extracted in the uncontrolled surveillance environment [14], body appearance is widely exploited for person re-identification task [15]–[18].

Generally speaking, person re-identification can be regarded as an image retrieval problem [21]. The paradigm usually consists of three stages: feature extraction, distance measure and final ranking. The feature extraction stage aims to construct discriminative and robust feature descriptions to separate different persons in various cameras [16], [18], [22]–[24]. However, designing a set of features satisfactorily is very difficult, especially in the presence of significant appearance variations caused by large view changes. The distance measure stage focuses on seeking a proper measure to reflect the identity consistency among person images [14], [19], [25]–[28]. Its main drawback is that the performance is heavily dependent on sufficient training samples, which is usually difficult to acquire in practical surveillance applications. The final ranking stage pays attention to optimize the original ranking lists by mining various similarity relationship in the initial ranking results [29]–[35]. Since ranking optimization is independent of concrete feature representation and distance measure methods, it owns great flexibility in practical applications [36], [37].

Based on the above analysis, we focus on the ranking optimization stage, especially on order-based automatic ranking

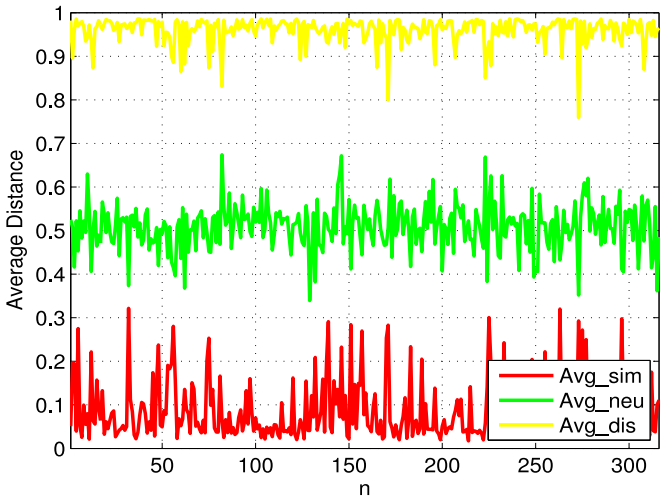



Fig. 1. Illustration of our *conductive similarity* and *insulative dissimilarity*. The preliminary experiment is conducted by KISSME [19]. A total of 316 image pairs are randomly selected from the VIPeR dataset [20] for testing. Note that the distances between true matches and the strongly similar galleries are significantly lower than those of strongly dissimilar galleries.

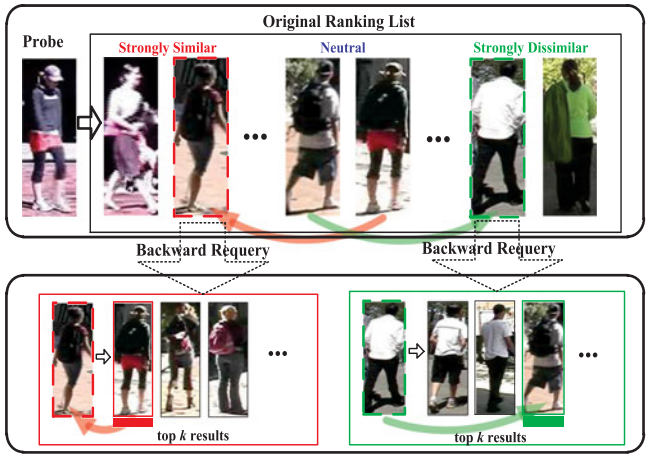



Fig. 2. Illustration of our main idea. The first row is the original rank list, while the second row denotes some backward requery top-k results of the new probes. The ones similar to (illustrated by arrow) the strongly similar galleries are much likely to be the correct match (illustrated by red arrow) and should be pulled, while the galleries similar to strongly dissimilar galleries are probably the wrong match (illustrated by green arrow) and should be pushed.

optimization method in this paper. Different from previous automatic ranking optimization methods, which mainly focus on exploring similarity relationships among galleries [30], [33], this paper considers both similarity and dissimilarity relationships for ranking optimization in an automatic manner to further improve person re-identification performance. The latent assumption is that the true match should not only be similar to those strongly similar galleries, but also dissimilar to those strongly dissimilar galleries of the probe person. This assumption can also be evidenced by the basic principles in social networks [30], [38]: Two persons who have many common friends would be much likely to be good friends. Correspondingly, if a gallery is strongly similar to the friends (strongly similar galleries) of the probe, it will be much likely to be a friend of the probe. We name this phenomenon as *conductive similarity*. On the other hand, if a gallery is very similar to the strangers (strongly dissimilar galleries) of the probe, it is much prone to differ from the probe. We name it as *insulative dissimilarity*.

To illustrate the above idea, a preliminary experiment is conducted to show the existence of *conductive similarity* and *insulative dissimilarity* shown in Fig. 1. Given an initial ranking result, the top-10 results are treated as strongly similar galleries while the bottom-10 as strongly dissimilar galleries. As a reference, the neutral galleries are denoted by the middle-10 galleries. For each probe, we compute the pair-wise distance between the true match (groundtruths) and other gallery images. The average distance between the true match and the strongly similar galleries (top-10) is reported as 'Avg_sim', and that between the match and neutral galleries (middle-10) and strongly dissimilar galleries (bottom-10) are denoted as 'Avg_neu' and 'Avg_dis', respectively. As can be seen from Fig. 1, true matches are much likely to be similar to strongly similar galleries and dissimilar to strongly dissimilar galleries. Therefore, it is reasonable to optimize the original ranking list based on the *conductive similarity* and *insulative dissimilarity*. An illustrative example is shown in Fig. 2. After the backward requery, the galleries similar to those strongly similar ones are pulled, while the galleries similar to strongly dissimilar galleries are pushed. In this way, the rank of the correct match is improved.

In addition, considering the unreliable unilateral similarity as illustrated in [39], [40], a more reasonable assumption is that the strongly similar galleries should be similar to the probe person in multiple different ranking results obtained by different baseline methods [41]. Therefore, we propose a ranking aggregation method combining two different person re-identification methods to enhance the *conductive similarity* and *insulative dissimilarity*. A depicted model is illustrated in Fig. 3. The strongly similar galleries are those which are very similar to the probe person, i.e. being verified by various baseline methods, e.g., $1_+$ and $3_+$ in Fig. 3(b). Thus the strongly similar galleries are achieved from the intersection set of the top-$k$ results of the baseline methods. On the contrary, the strongly dissimilar galleries, e.g., $1_-, 2_-, 3_-, 4_-$ and $5_-$ in Fig. 3(b), are achieved from the union set of the bottom-$k$ results of the baseline methods. After that, we pull the quasi-similar galleries (which are similar to strongly similar galleries) and push the quasi-dissimilar galleries (which are similar to strongly dissimilar galleries) to optimize the ranking orders. In this way, our proposed approach can be divided into two steps: similarity ranking aggregation (SRA) and dissimilarity ranking aggregation. The former focuses on improving the ranking orders of quasi-similar galleries, while the latter pays attention to penalizing the quasi-dissimilar galleries to improve the ranks indirectly.

The main contributions of this paper are summarized as follows:

1) We proposed a combination method exploring both similarity and dissimilarity relationship, which is seldom investigated in previous work, to optimize original ranking result for the person re-identification task.

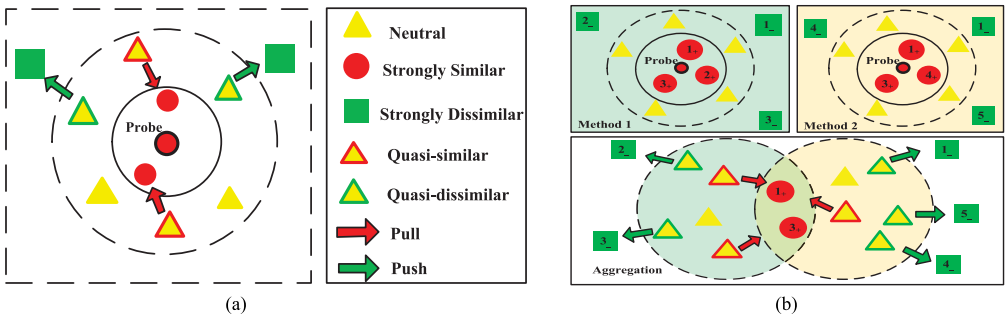



Fig. 3. Depicted model of our idea. (a) The quasi-similar galleries are pulled while the quasi-dissimilar galleries are pushed, which can lead to better results. (b) Our aggregation idea. The first row is two original ranking results achieved by two baseline methods, while the second row shows our aggregation strategy. Note that numbers denote different gallery images while subscripts "+" and "-" represent similar and dissimilar galleries, respectively. The strongly similar galleries are achieved from the intersection set of two methods, while strongly dissimilar galleries are obtained from the union set of dissimilar galleries.

2) We presented a ranking aggregation framework which is firstly introduced for the person re-identification task. According to the experiments, this framework is applicable to combine different baseline methods.
3) We validated our approach on three public datasets, VIPeR [20], CUHK01 [42] and PRID450S [43], and confirmed that the proposed method achieves more favorable performance than state-of-the-art methods.

The rest of this paper is organized as follows. In Section II, a brief review of related work is conducted for person re-identification. After that, we introduce our ranking aggregation method based on *conductive similarity* and *insulative dissimilarity* in Section III. Section IV presents the simplified version of our method for the optimization with a single input baseline method. Later, Section V shows the experimental results on three representative public datasets with some discussions. Finally, concluding remarks are given in Section VI.

## II. Related Work

According to above analysis, our focus in this paper is mainly about ranking optimization for person re-identification. In this section, some prior related works are introduced to illustrate the two contributions of this paper.

Ranking Optimization, locating in the final stage of person re-identification task [44], [45], aims to optimize original ranking lists by mining various similarity relationship in initial ranking results [29]–[34]. In the person re-identification field, relevant work can be generally categorized into two kinds: interactive relevance feedback methods and automatic re-ranking methods.

The interactive relevance feedback methods revised the initial re-identification results based on manual interaction to mine the optimization cues. Ali *et al.* [46] employed rank based constraints and convex optimization to efficiently learn the distance metric during a visual search process. Wang *et al.* [32] proposed a local-similarity based ranking optimization method in an interactive manner. They chose the local-part similar instead of global similar galleries. Liu *et al.* [31] presented a novel one-shot post-rank optimisation method (POP) at the user end, they manually selected some strongly and weakly similar samples to mine the negative cues to optimize the initial ranking results. Indeed, [31] considered the dissimilar cues, it generates some strong positive samples and weak negative samples to optimize the ranking results in an interactive manner by choosing strongly and weakly similar samples manually. However, this kind of interactive methods need lots of manpower which is unsuitable for large scale data scenarios [33], [47]. In contrast, our ranking optimization method focuses on optimizing the searching results in an automatic pattern.

The automatic re-ranking methods focused on improving the searching results automatically [30], which owns more flexibility. Li *et al.* [29] analyzed the commonness of the nearest neighbors to optimize the original ranking list. Le *et al.* [48] conducted a re-ranking method based on some soft biometrics(semantic attributes) cues. Andy *et al.* [33] presented a query-adaptive re-ranking method (QARR) based on locality preserving projections to model the query variations across cameras. Leng *et al.* [30] proposed an automatic bidirectional ranking method based on content and context similarity, the gallery images are treated as new probes to requery in the original gallery set, namely backward re-query, mining the similarity relationship between probe and gallery images. They all focused on revising the original ranking results with similarity cues among the galleries, while dissimilarity cues were always neglected.

Moreover, as a special image retrieval problem [49], the comparison with ranking optimization in general image retrieval is also conducted. Comparing to traditional large scale image retrieval [50], person re-identification is a fine-grained distinction problem [51], [52], ranking optimization in image retrieval is almost impossible to utilize the strongly dissimilar galleries for the reason that there are large amount of different type of objects [53], where a gallery image may be a person, a place or a distracter, which may make it impractical to mine the dissimilarity cues. In comparison, the gallery images in person re-identification task are all person images that share more common characteristics, while the differences among the gallery images are much smaller, which provides us an opportunity to use the insulative dissimilarity in a ranking optimization framework. Specifically, we exploited both similarity and dissimilarity cues in a ranking optimization framework to refine the searching results.

Apart from utilizing the dissimilarity relationship, another contribution of this paper is mainly about ranking aggregation. Ranking aggregation is a branch of ranking optimization, which

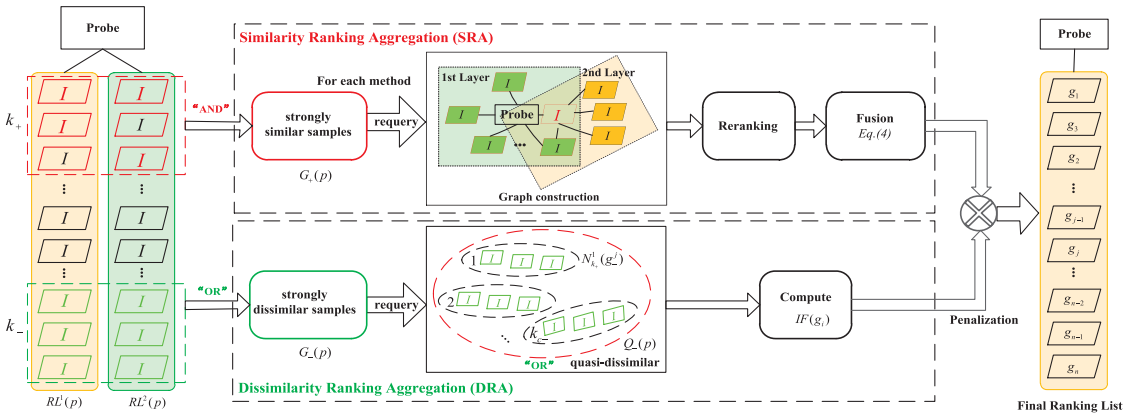



Fig. 4. Framework of our aggregation approach. It can be divided into two parts: similarity ranking aggregation and dissimilarity ranking aggregation.

is usually based on two or more ranking lists [34]. Most related aggregations in person re-identification are conducted by combining multiple similarity scores directly [18], [22]. However, ranking aggregation is never investigated for this task, all current ranking optimization works are conducted based on one SINGLE baseline input method except for our previous work in [34]. In general image retrieval, ranking aggregation is a straightforward solution to fusing different methods at the rank level. Zhang *et al.* [39] proposed a graph-based query specific fusion approach where multiple retrieval sets are merged and reranked by conducting a link analysis on a fused graph. In comparison, we propose a ranking aggregation method for person re-identification based on the *conductive similarity* and *insulative dissimilarity*. Specifically, our ranking aggregation method adopts a cross-view based backward requery strategy to enhance the complementarity of the two baseline input methods, and contains two parts: similarity ranking aggregation, which aims to improve the ranking orders of quasi-similar galleries that may be positive ones; dissimilarity ranking aggregation, which penalizes the quasi-dissimilar galleries that may be negative ones. They are further combined in a sequential way.

In addition, ranking aggregation can be divided into score-based re-ranking and order-based re-ranking [54]. The former associates the objects in input rankings with their original similarity scores while the latter focuses on the original ranking orders information. This paper considers order-based ranking aggregation due to two reasons [37]. First, this kind of methods are more stable and robust to outliers. Second, score-based methods can be readily converted to order-based ones.

## III. Proposed Approach

The framework of our aggregation approach is shown in Fig. 4, with two main parts: similarity and dissimilarity ranking aggregation. For a specific probe person, two original ranking lists are firstly obtained by two different baseline methods, and methods with more complementarities will work better. Note that any method that generates a ranking list for a probe can be used here. The similarity ranking aggregation is conducted for the strongly similar galleries, cross-view based backward requery is performed for the re-ranking of quasi-similar galleries, and the refined ranking lists are combined with different weights. The dissimilarity ranking aggregation is presented for the strongly dissimilar galleries, it targets to the quasi-dissimilar galleries whose ranks are penalized. Then, the two parts are combined to generate a final ranking list for the probe. The details are discussed in the following subsections.

### A. Similarity Ranking Aggregation

Similarity ranking aggregation (SRA) is divided into three steps: first, the strongly similar galleries are obtained from the intersection set of top-$k$ results achieved by two baseline methods. Second, the strongly similar galleries are treated as new probes to requery in the original gallery set, named backward requery [30]. More specifically, to enhance the complementarity of two methods, the cross-view based backward requery is conducted [39], i.e., the original ranking list is achieved by one method, and the other method is adopted for the requery to refine the original ranking list. And it's validated in later experiments as shown in Section V. Third, graph-based weighted reranking is introduced to generate a refined ranking list.

For the convenience of the following discussion, we denote the probe person image as $p$ and the gallery set as $G = \{g_i \mid i = 1, 2, \ldots N\}$, where $N$ is the number of images in the gallery set. We mark the first method with superscript "1", and the second method with "2". We firstly obtain two original ranking lists by two baseline methods for a probe $p$, denoted as $RL^1(p)$ and $RL^2(p)$. $N^1_{k_+}(p)$ and $N^2_{k_+}(p)$ denote the top-$k_+$ galleries in $RL^1(p)$ and $RL^2(p)$, respectively. The set of strongly similar galleries $G_+(p)$ is formulated as

$$G_+(p) = N^1_{k_+}(p) \cap N^2_{k_+}(p). \tag{1}$$

Next we use the $k_{c+} = |G_+(p)|$ strongly similar galleries in $G_+(p)$ for backward requery. In other words, we treat each $g^j_+ \in G_+(p)$ as a new probe to search in the original gallery set. Moreover, to enhance the complementarity of the original methods, we adopt the crossed-view backward requery. That is to say, we refine the first ranking list $RL^1(p)$ with the other method "2" for backward requery and vice versa. By backward requery, a new ranking list for each method is achieved for each $g^j_+$.

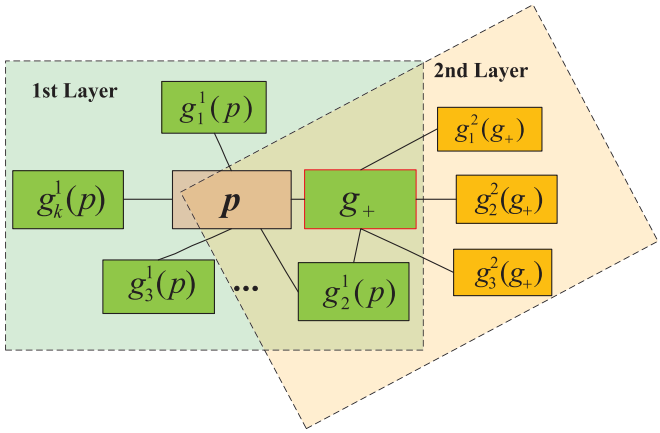



Fig. 5. Example of graph construction. $p$ is the probe and $g_+$ is a strongly similar sample. The superscripts denote different two methods. Note that the method of the second layer is differ to the first layer.

For a better presentation of our graph-based weighted reranking, we explain it via an example, where the original ranking list is generated by the first method and the second method is used for backward requery. We construct a weighted undirected graph $Graph^+ = \langle G, E, w \rangle$ for each probe and each strongly similar sample in $G_+(p)$, the center is the probe while the nodes are the neighbour images. We use $g_+$ as an example of $g_+^j$ for better expression as shown in Fig. 5. The green part indicates the neighbours of probe image $p$ in the original ranking list achieved by the first method, denoted as $N_{k_+}^1(p)$. The orange part expresses the backward requery neighbours of $g_+$, denoted as $N_{k_+}^2(g_+)$. They are linked by an edge $(p, g_+) \in E$. Note that in the second layer of the graph, a different method other than the one used for the first layer, is adopted to enhance the complementarity. The similarity between $p$ and $g_+$ is redefined as the weighted Jaccard similarity coefficient [55] between the neighborhoods of $p$ and $g_+$

$$Sim^1(p, g_+) = w(p, g_+) \frac{\left| N_{k_+}^1(p) \cap N_{k_+}^2(g_+) \right|}{\left| N_{k_+}^1(p) \cup N_{k_+}^2(g_+) \right|} \tag{2}$$

$$w(p, g_+) = {w_0}^{rank(g_+, G_+(p))} \tag{3}$$

where $|\cdot|$ denotes the cardinality and $w(p, g_+)$ is a weighting coefficient related to the original rank in $G_+(p)$. The decay factor is defined as $w_0$ and we set $w_0 = 0.8$ the same as [39] in all experiments. $rank(g_+, G_+(p))$ represents the rank of $g_+$ in $G_+(p)$, and expresses the original importance in the ranking list.

Similarly, $Sim^2(p, g_+)$ can be achieved by a crossed backward requery, i.e. the first method "1" is adopted for backward requery to refine the similarity score in $RL^2(p)$. Then, a late fusion [54] is introduced to combine the similarity scores, i.e. a weighted combination of the similarity defined in (2) is conducted

$$Sim'(p, g_+) = \alpha \cdot Sim^1(p, g_+) + (1 - \alpha) \cdot Sim^2(p, g_+), \tag{4}$$

**Algorithm 1:** The SRA Algorithm.

**Input:** A probe image $p$ and a gallery set $G = \{g_i \mid i = 1, 2, \ldots n\}$.

**Output:** A ranking list for the probe image.

**Offline:**

1: Querying every gallery image $g_i$ in the gallery $G$ with two different methods.
2: Achieve the top-$k$ galleries of each image $g_i$.

**Online:**

1: Probe $p$ in the gallery set $G$.
2: Obtain two original ranking lists $RL^1(p)$ and $RL^2(p)$ by two different methods for the probe $p$.
3: Get strongly similar galleries set $G_+(p) = N_{k_+}^1(p) \cap N_{k_+}^2(p)$
4: For method 1:
5: **for** $i = 1$ to $|G_+(p)|$ **do**
6: $g_+$ is the $i$-th item in $G_+(p)$
7: Cross-view based backward requery for $g_+$
8: Compute weighting coefficient $w(p, g_+)$ by Eq. (3)
9: Compute new score $Sim^1(p, g_+)$ by Eq. (2)
10: **end for**
11: Repeat step 4–10 to get $Sim^2(p, g_+)$
12: Late fusion of $Sim^1(p, g_+)$ and $Sim^2(p, g_+)$ by Eq. (4)
13: Use new scores to re-rank $RL(p)$.
14: The final ranking list $RL'(p)$ is achieved.

where $\alpha$ denotes the weighting parameter of the two baseline methods. It depends on the original performance of the two baseline methods, i.e. it's easy to assume that a higher weight is assigned to a better baseline method being modified. After that, the ranking orders of those strongly similar galleries are revised, other ranking orders are combined with $\alpha$ based on their original ranking scores in $RL^1(p)$ and $RL^2(p)$. Then, a refined ranking list $RL'(p)$ is achieved by SRA. Our SRA is summarized in Algorithm 1.

### *B. Dissimilarity Ranking Aggregation*

The above SRA does improve the quasi-similar galleries' ranking orders. Next, we introduce how to further penalize the quasi-dissimilar galleries via dissimilarity ranking aggregation (DRA). It can also be divided into three steps: first, the strongly dissimilar galleries are achieved from the union set of bottom-$k$ results obtained by two different baseline methods. Second, the strongly dissimilar galleries are treated as new probes to requery in the original gallery set with both the two original methods. Third, a quasi-dissimilar set is achieved by backward requery of the strongly dissimilar galleries, and the ranking orders of the quasi-dissimilar galleries are penalized according to their frequency in the quasi-dissimilar set.

The strongly dissimilar galleries are achieved from the union set of the bottom-$k$ neighbours. The bottom-$k_-$ farthest neighbours of each method are denoted as $N_{k_-}^1(p)$ and $N_{k_-}^2(p)$. To

get the strongly dissimilar galleries, we define their union set $G_-(p)$ as the strongly dissimilar galleries set of image $p$, which is formulated as

$$G_-(p) = N^1_{k_-}(p) \cup N^2_{k_-}(p). \tag{5}$$

Then we use the $k_{c-} = |G_-(p)|$ strongly dissimilar galleries $G_-(p)$ for backward requery. In other words, we treat each $g^j_- \in G_-(p)$ as a new probe to search in the original gallery set. After the backward requery, we'll get $2 \times k_{c-}$ backward ranking lists, for each of which top-$k$ can be expressed as $N^1_{k_+}(g^j_-)$ and $N^2_{k_+}(g^j_-)$. Then, the quasi-dissimilar union set of top-$k$ results $M^1_{k_+} = \{N^1_{k_+}(g^j_-) \mid j = 1, 2, \ldots k_{c-}\}$ and $M^2_{k_+} = \{N^2_{k_+}(g^j_-) \mid j = 1, 2, \ldots k_{c-}\}$ are achieved.

The image frequency of a quasi-dissimilar gallery $q_-$ appearing in $M^1_{k_+}$ and $M^2_{k_+}$ is denoted as $IF(q_-)$. Obviously, $\widehat{IF}(q_-) < 2 * k_{c-}$. Extending to a general case, the image frequency of all gallery images can be formalized as

$$\widehat{IF}(g_i) = \begin{cases} IF(g_i), & g_i \in \{M^1_{k_+}, M^2_{k_+}\} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

As the galleries in $Q_-(p)$ are much likely to be the wrong match of the probe, we penalize these quasi-dissimilar galleries based on the image frequency as follows:

$$Dis^*(g_i, p) = rank'(g_i) * \exp\left(\frac{1}{k_- + 1}\widehat{IF}(g_i)\right) \tag{7}$$

where $rank'(g_i)$ denotes the rank of $g_i$ in $RL'(p)$ achieved by SRA, i.e., the original rank orders are treated as the original distances between $g_i$ and $p$. Meanwhile, $RL^1(p)$ or $RL^2(p)$ can be treated as the input of the DRA to be optimized. After the penalization of the quasi-dissimilar galleries, the refined ranking list is achieved.

### C. Combination

The dissimilarity ranking aggregation is conducted together with the SRA as described in above subsection, in which it mainly focuses on revising the output ranking list achieved by SRA as shown in (7) and summarized in Algorithm 2. Specifically, after getting the ranking list output by SRA, we do the penalization operation to penalize the quasi-dissimilar galleries in a sequential way. In addition, as can be seen from the DSRA algorithm, Step 1–12 can be computed with the SRA simultaneously, which would help to decrease the computational cost significantly.

### D. Complexity Analysis

As can be seen from Algorithm 1 and Algorithm 2, the majority of the computation is spent on deriving the backward requery ranking lists, which consists of mass pair-wise similarity computation between gallery images. Note that both of the two algorithms adopted backward requery to generate new ranking lists, thus the time-consuming procedure needs to be done only once for all gallery images.

A traditional way of mass pair-wise similarity computation is calculating the distance of every image pairs directly and generating the ranking list based on these distances. Suppose that the gallery contains $n$ images, the computation complexity is $O(n^2)$ for distance measure and $O(n^2 \log n)$ for the ranking operations. In some practical applications, gallery images are obtained before querying the probe image, e.g. surveillance video investigation. In such cases, the computation of our method is divided into two separate phases. In offline phase, all gallery images are mutually compared with a computation complexity $O(\frac{n(n-1)}{2})$ and a ranking complexity $O(n^2 \log n)$. In online phase, it only needs to compute the distance between the probe and every gallery image with a computation complexity $O(n)$ and ranking complexity is $O(n \log n)$. With the above two-phrase implementation, the whole algorithm's complexity can be greatly reduced and its online part is only proportional to the size of the top-$k_+$ and bottom-$k_-$, which is especially suited to those real-time-requiring applications, e.g. video investigation. Specially, the computation complexity also relies on the original baseline methods.

**Algorithm 2:** The DSRA Algorithm.

**Input:** A probe image $p$, a gallery set $G$, and the ranking list $RL'(p)$ exported by SRA.
**Output:** A new ranking list for the probe image.
**Offline:**
1: Querying every gallery image $g_i$ in the gallery $G$ with two different methods.
2: Achieve the bottom-$k$ galleries of each image $g_i$.
**Online:**
1: Query $p$ in the gallery set $G = \{g_i \mid i = 1, 2, \ldots n\}$.
2: Obtain two original ranking lists $RL^1(p)$ and $RL^2(p)$ by two different methods for the probe $p$.
3: Get strongly dissimilar galleries $G_-(p) = N^1_{k_-}(p) \cup N^2_{k_-}(p)$
4: For method 1, 2:
5: **for** $i = 1$ to $|G_-(p)|$ **do**
6: $g_-$ is the $i$-th item in $G_-(p)$
7: Cross-view based backward requery for $g_-$
8: Compute $N^1_{k_+}(g_-)$ and $N^2_{k_+}(g_-)$
9: **end for**
10: **for** $i = 1$ to $|RL'(p)|$ **do**
11: $g_i$ is the $i$th image in $RL'(p)$
12: Compute $\widehat{IF}(g_i)$ by Eq. (6)
13: Penalization by Eq. (7)
14: **end for**
15: Use new scores to re-rank $RL'(p)$.

## IV. Re-ranking for a Single Method

The above Section III-A and III-B present our aggregation approach, which requires two baseline methods. When there is only one method available for optimization, our aggregation method is degenerated to a re-ranking method via *conductive similarity* and *insulative dissimilarity*. Specifically, the SRA is degenerated to a re-ranking method via context similarity similar


Fig. 6. Example image pairs captured from three public datasets. Each column shows two images of the same identity from two different cameras with significant changes on view point and illumination condition. (a) VIPeR dataset. (b) CUHK01 dataset. (c) PRID450S dataset.

to [30], [56]. In detail, the strongly similar galleries are reduced to the top-$k_+$ retrieval results, i.e. (1) is rewritten as

$$G_+(p) = N_{k_+}(p). \tag{8}$$

The cross-view based backward requery is reduced to a simple backward requery, and the revised similarity between $p$ and $g_+ \in G_+(p)$ shown in (2) is simplified to

$$Sim(p, g_+) = w(p, g_+) \frac{\left|N_{k_+}(p) \cap N_{k_+}(g_+)\right|}{\left|N_{k_+}(p) \cup N_{k_+}(g_+)\right|}. \tag{9}$$

Additionally, the DSRA can be deduced similarly. The strongly dissimilar galleries of the probe shown in (5) is simplified as

$$G_-(p) = N_{k_-}(p). \tag{10}$$

Moreover, the penalization procedure can be reproduced as shown in Section III-B. Thus, the re-ranking via insulative dissimilarity is realized.

Totally speaking, our aggregation method can also be simplified to conduct a re-ranking for a single method version. The re-ranking method can be described as an extended re-ranking method via context similarity [56], which also takes the dissimilarity cues into account. Furthermore, the reranking for a single method is a well illustration of our *conductive similarity* and *insulative dissimilarity*.

## V. Experimental Results

In this section, extensive experiments are conducted on three publicly available datasets: the VIPeR dataset [20], the CUHK01 dataset [42] and the PRID450S dataset [43]. We chose these datasets because they provide many challenges faced in practical surveillance, i.e., viewpoint, pose and illumination changes, different backgrounds, low image resolutions, occlusions, etc. Also, they provide two labeled image sets of persons captured by two cameras with non-overlapping fields of views, in which images of the same person have the same label, while images of the different persons have different labels. Fig. 6 shows some example pictures of these three datasets. All these person re-id datasets are released based on hand-drawn bounding boxes from practical surveillance videos, and these datasets are widely adopted in current person re-identification works [1], [28]. Alternatively, the bounding boxes can be achieved by an efficient pedestrian detection method in many practical applications.

### A. Datasets and Evaluation Protocols

The VIPeR is a dataset that mainly considers the influence of viewpoint change, and is most widely used for evaluating person re-identification methods. It contains 632 person image pairs captured from two different static camera views in outdoor academic environments. The dataset is challenging due to the intensive viewpoint changes with most of the matched image pairs containing a viewpoint change of $90^\circ$. Other variations are also considered, such as illumination conditions and the image qualities as shown in Fig. 6(a). All the images are normalized to $128 \times 48$ for experiments.

The CUHK01 dataset is also obtained from two disjoint camera views in an outdoor campus environment. It is the one that has the highest number of persons collected by a single camera pair, thus it is the most representative for a real scenario. It contains 971 persons with 3,884 images, and each person has two images in each camera. The person images in camera A are mostly captured by frontal view or back views while camera B captures the side views as shown in Fig. 6(b). All the images are normalized to $160 \times 60$ for experiments. As a single representative image per camera view for each person is considered in this paper, we randomly selected one image from two galleries per camera view for each people in the same way as done in [28].

The PRID450S is a new and more realistic dataset. It contains 450 singleshot image pairs captured over two spatially disjoint camera views. All images are normalized to $168 \times 80$ pixels. Different from the VIPeR dataset and CUHK01 dataset, this dataset has significant and consistent lighting changes and chromatic variation as shown in Fig. 6(c), which is more challenging. With this dataset, we mainly evaluate the extensibility of the proposed approach for different illumination conditions.

All the quantitative results are exhibited in standard Cumulated Matching Characteristics (CMC) curves [16]. The CMC curve is a plot of the recognition performance versus the rank score and represents the expectation of finding the correct match

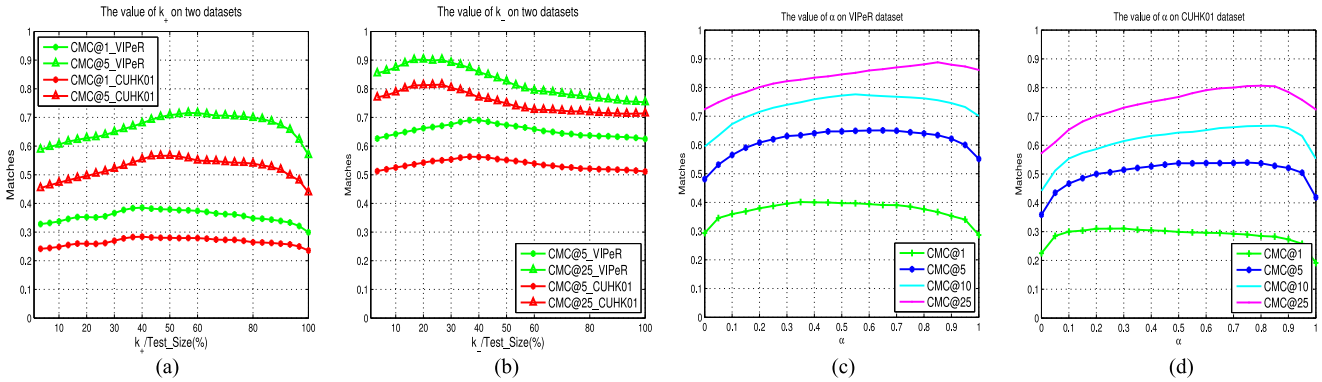



Fig. 7. Parameters analysis based on KISSME [19] and SDC [22] on the VIPeR and CUHK01 dataset. (a) $k_+$ is the parameter of the SRA. (b) $k_-$ is the parameter of the DRA. (c) and (d) $\alpha$ is the weighting parameter of two baseline methods.

inside top $k$ matches. On the other hand, nAUC describes how well a method performs irrespectively of the dataset size. Following the evaluation protocol described by many state-of-the-art works, we randomly partition the dataset into two even parts, $50\%$ for learning and $50\%$ for testing, without overlap on person identities. All images from Camera View A are treated as probes and those from Camera View B as gallery set. For each probe image, there is one person image matched in the gallery set. Rank-$k$ recognition rate is the expectation of finding the correct match within the first $k$ ranks, and the cumulated values of recognition rate at all ranks is recorded as one-trial CMC result. With two different methods, we use the same configuration for experiments at each trial to get the ranking lists. To achieve stable statistics, we repeated the evaluation procedure for 10 times.

### B. Implementation Details

To evaluate the effectiveness of our aggregation method, we adopted KISSME [19] and SDC [22] as the two main baseline methods. The reason of selecting these two methods is as follows: KISSME projects the concatenated feature histograms into subspace by PCA (Principal component analysis) to obtain their global information for metric learning, which can be treated as a global feature based method. In comparison, the SDC extracts distinctive features by finding the salience regions for constructing robust discriminative descriptions, which can be treated as local feature based method. The complementarity of global and local feature based methods can help to improve the effectiveness of our approach. To further evaluate the flexibility of the aggregation, other baseline methods, such as L2 distance, LOMO [57] and SCNCD [44] methods, are also used for experiment on the VIPeR and PRID450S dataset.

Moreover, to evaluate the effectiveness of dissimilarity, optimization for a single method is tested on the VIPeR dataset. To better compare with the re-ranking via context similarity [56], three baseline methods, L2 distance, KISSME and SCNCD are utilized for comparison. And the feature representation is a combination descriptor consisting of color and texture features as described in [21].

In addition, the following different terms represent different configurations. "*SRA*" denotes the similarity ranking aggregation, "*DRA*" denotes the dissimilarity ranking aggregation, "*DSRA*" denotes the DRA together with the SRA, and "*Fusion_Baseline*" denotes the direct ranking fusion used as a baseline to verify the superiority of our aggregation over plain fusion, *i.e.*, two baseline methods are fused based on their original similarity scores with $\beta$ as shown in the following equation:

$$Sim(p,g) = \beta \cdot S^1(p,g) + (1-\beta) \cdot S^2(p,g) \qquad (11)$$

where $S^1(p,g)$ and $S^2(p,g)$ denote the original similarity scores of two baseline methods, $\beta$ denotes the weighting parameter. Compared to $\alpha$ in SRA shown in Eq. (4), $\alpha$ means the different combining weights of two different baseline methods after the cross-view based backward requery, while $\beta$ is the weighting parameter of two baseline methods.

### C. Aggregation Parameters Analysis

The parameters of our method are analyzed in this section. The baselines methods are set to KISSME [19] and SDC [22], and both our SRA and DSRA are evaluated for different $k_+$s and $k_-$s on the VIPeR and CUHK01 datasets. $k_+$ is shown in Fig. 7(a) while $k_-$ is fixed as $k_- = N_t * 20\%$, where $N_t$ is the size of the test dataset, CMC@1 and 5 are reported for the reason that SRA has more influence on the top ranked results. For $k_+$, the peak of CMC@1 is somewhere around $k_+ = 40\% * N_t$. Note that rank-1 matching rate is much more important than others in real applications. It is evident from Section III-A that a smaller value of $k_+$ causes fewer variations in the ranking list, and leads to limited improvement, i.e. $|G_+(p)|$ is very small, only a few galleries are modified. On the other hand, when $k_+$ is too large, it may lead to unreliable strongly similar galleries, thus causing negative effects. Moreover, the curves drop earlier for CMC@1 than CMC@5. This further confirms that the *conductive similarity* mainly improves the ranking orders of similar galleries.

$k_-$ is shown in Fig. 7(b) while $k_+$ is fixed as $k_+ = N_t * 40\%$, CMC@5 and 25 are reported for the reason that DRA has more influence on the middle ranked galleries rather than the top ranked results. Using different $k_-$s has slight influence on

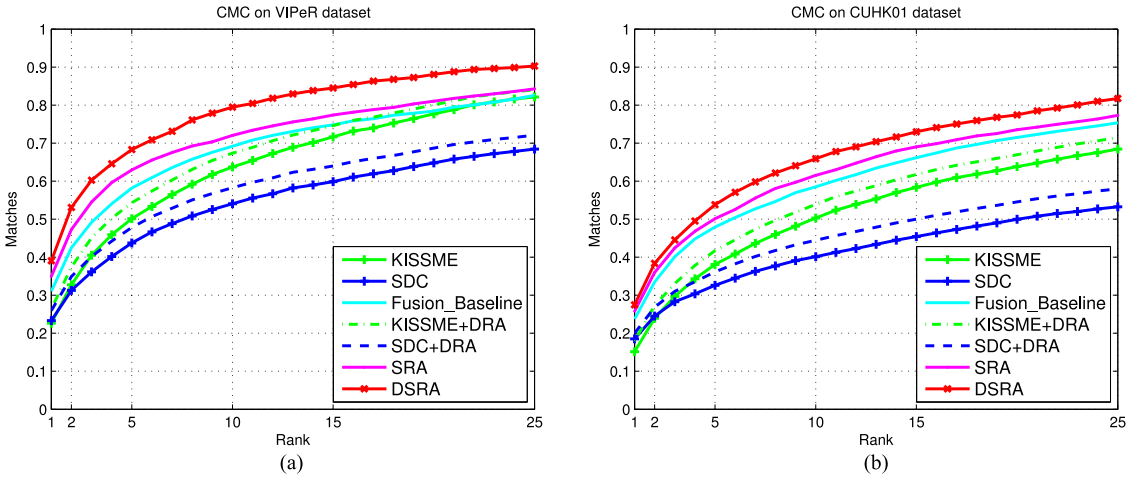



Fig. 8. Effectiveness of our aggregation method. "Fusion_Baseline" denotes the direct combination. Our approach: SRA and DSRA based on KISSME [19] and SDC [22]. (a) Performance on the VIPeR dataset. (b) Performance on the CUHK01 dataset.

CMC@5 for the reason that the *insulative dissimilarity* mainly improves the ranking orders of quasi-dissimilar galleries. And when $k_-$ is too large, it may treat lots of quasi-similar galleries as dissimilar galleries, thus lead to the performance degradation. The best choice for $k_-$ is around $20\% * N_t$ for CMC@25, because the CMC curves drop quickly when $k_-$ increases further.

The other parameter $\alpha$ is a weighting parameter of two baseline methods similarity scores after cross-view based backward requery and its impact is shown in Fig. 7(c) and (d). It's easy to assume that a higher weight is assigned to the better baseline method. Specially, if the two baseline method achieve the almost equivalent performance, $\alpha$ is close to 0.5. Meanwhile, when $\alpha = 0$ or $\alpha = 1$ the proposed method is degenerated to a ranking optimization for a single baseline method. As can be seen from the experiments, the performance always out performs the original baseline methods as $\alpha$ changes, which further which illustrates the effectiveness of the proposed method. In this paper, $\alpha$ is set to 0.6 for the reason that KISSME achieves a little higher performance than SDC.

Based on the above experiments, all the parameters are selected based on the test dataset size, while the $k_+$ is about 35%–45% of the test dataset size, correspondingly, $k_-$ is about half of $k_+$. Basically, $k_+$ should be relatively larger in order to achieve enough strongly similar galleries. On the contrary, if $k_-$ is too large, the correct match may be treated as false match. For the three datasets, the parameters are set based on their test dataset size ($N_t$) as follows: $k_+ = N_t * 40\%$ and $k_- = N_t * 20\%$ for the VIPeR and CUHK01 datasets. Additionally, to illustrate the practicability, we choose $k_+$ and $k_-$ based on this rule, while the test dataset sizes for three datasets are $N_t = 316$, 485 and 225, respectively.

### D. Aggregation Evaluation

*Effectiveness:* The results of the experiments conducted on VIPeR and CUHK01 datasets are shown in Fig. 8. As can be seen from Fig. 8(a), our approach yields consistent improvement compared with the baseline KISSME [19] and SDC [22]. Specifically, "*DSRA*" achieves nearly 18% improvement at rank 1, and 20% at rank 5 on the VIPeR dataset compared with the original baseline methods. In particular, rank 1 matching rate is around 35% for "*SRA*" with merely similarity ranking aggregation, and 39% for "*DSRA*" with further dissimilarity ranking aggregation. Comparing to the direct combination "*Fusion_Basline*", our method has significant improvement, both similarity ranking aggregation and dissimilarity ranking aggregation achieved promising results. In addition, we also evaluate the performance of the DRA conducted on the original baseline input ($RL^1(P)$ or $RL^2(P)$), the experimental results of "*KISSME + DRA*" and "*SDC + DRA*" are shown in Fig. 8, which verified the effectiveness of the dissimilarity cues further.

TABLE I
RUNTIME COMPARATIVE RESULTS OF OFFLINE AND ONLINE PHASES ON THE VIPER DATASET. NOTE THAT OFF-LINE PROCESSING CONTAINS THE TRAINING TIME AND THE BACKWARD REQUERY TIME. "+" MEANS THE ADDITIONAL BACKWARD REQUERY TIME. THE RESULTS ARE REPORTED IN SECONDS (S)

| | Offline | | Online | |
|---|---|---|---|---|
| | LOMO | SCNCD | Query | Aggregation |
| Time (s) | 5.09 + 1.02 | 0.64 + 0.06 | 1.22 | 2.04 |

The results of another experiment conducted on the CUHK01 dataset are shown in Fig. 8(b). The improvement achieved by "*DSRA*" is about 10% and 18% at rank 1 and rank 10, respectively, compared to the better original baseline methods. This further validates the effectiveness of our aggregation methods, both SRA and DSRA. In a word, both the conductive similarity and insulative dissimilarity can help to improve the performance of person re-identification.

*Flexibility:* To verify the flexibility of our proposed method, another three experiments are conducted to answer the following three questions:

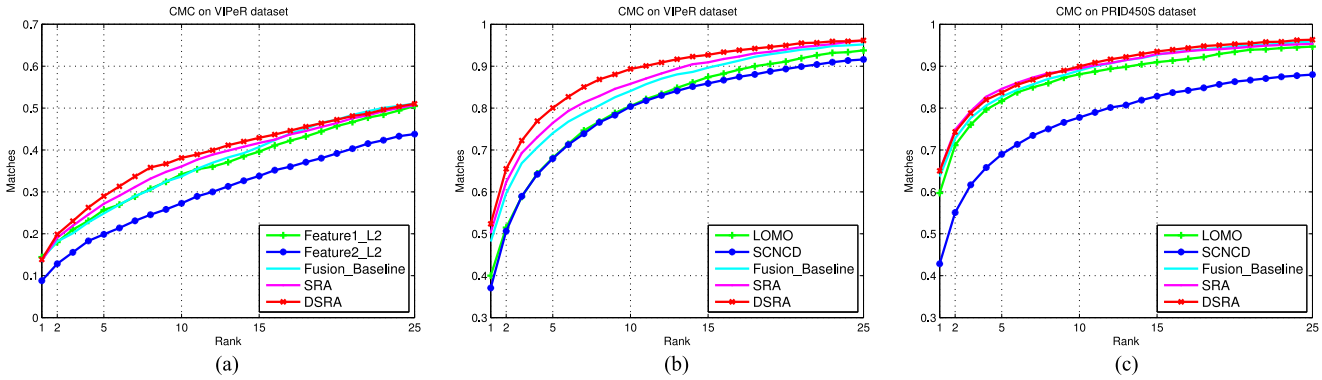



Fig. 9. Flexibility of our proposed method. (a) Experiment on the VIPeR dataset based L2 distance with two visual features on [17] and [57]. (b) Experiment on the VIPeR dataset based on LOMO [57] and SCNCD [44]. (c) Experiment on the PRID 450S dataset based on LOMO [57] and SCNCD [44].

TABLE II
DIFFERENT BASELINE METHODS AGGREGATED IN PAIRS ON THE VIPER DATASET. THE RESULTS ARE CONDUCTED BY "*DSRA*". "↑ @1" INDICATES THE IMPROVEMENT RATIO COMPARED TO THE BETTER ORIGINAL BASELINE METHOD AT RANK 1. THE MATCHING RATES ARE REPORTED IN PERCENTAGE (%)

| $Rank \rightarrow$ | $r = 1$ | $r = 5$ | $r = 15$ | ↑ @1 |
|---|---|---|---|---|
| L2 | 10.70 | 19.78 | 33.13 | - |
| SDALF [18] | 19.87 | 38.89 | 58.22 | - |
| KISSME [19] | 22.63 | 50.13 | 71.65 | - |
| SDC [22] | 23.32 | 43.73 | 59.87 | - |
| SCNCD [44] | 37.09 | 64.21 | 85.89 | - |
| LOMO [57] | 40.00 | 64.40 | 87.37 | - |
| L2+SDALF | 19.92 | 39.03 | 58.12 | 0.3% |
| L2+KISSME | 22.98 | 50.32 | 71.78 | 1.5% |
| SDALF+KISSME | 25.42 | 53.02 | 71.99 | 12.3% |
| SDALF+SDC | 27.56 | 59.52 | 68.54 | 18.2% |
| SDALF+SCNCD | 37.32 | 64.29 | 85.92 | 0.6% |
| SDALF+LOMO | 42.37 | 68.86 | 87.34 | 5.9% |
| KISSME+SDC | 39.07 | 68.29 | 79.82 | 67.5% |
| KISSME+SCNCD | 37.85 | 68.80 | 86.77 | 2.0% |
| KISSME+LOMO | 45.60 | 72.78 | 89.40 | 14.0% |
| SDC+SCNCD | 43.58 | 70.29 | 88.06 | 17.5% |
| SDC+LOMO | 43.32 | 70.34 | 89.06 | 8.3% |
| SCNCD+LOMO | 52.37 | 80.03 | 92.69 | 30.9% |

*Does it work with a naive L2 distance?* The first experiment is implemented with a classical L2 distance, two unsupervised visual descriptors from [17] and [57] are adopted as two original input baseline methods, the descriptors are measured with a naive L2 distance. As can be seen from Fig. 9(a), the improvement is also quite satisfying, although not as large as before. The main reason is that the methods used as baselines do not work well with the L2 distance, which limits the overall performance.

*Does it work with other baseline methods?* The second experiment is conducted with the two baseline methods, LOMO [57] and SCNCD [44] on the VIPeR dataset, verifying the effectiveness of our aggregation method on other baseline methods. Note that these two baseline methods are both newly proposed with high performance and low complexity, and their parameters are set as described in the previous section. The experimental results are shown in Fig. 9(b). As can be seen from the figure, our aggregation method can be well applied to other baseline methods. Moreover, by aggregating these two methods, our aggregation method achieved the best performance on the VIPeR dataset.

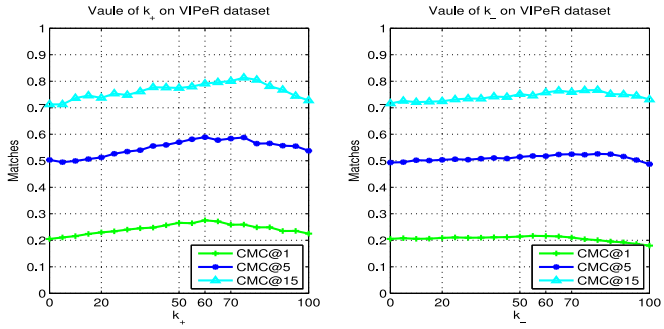



Fig. 10. Different $k_+$ s and $k_-$s based on KISSME [19] on VIPeR. $k_+$ and $k_-$ are the parameters of the similarity ranking optimization and the dissimilarity ranking optimization, respectively.

*Does it work on other datasets?* The third experiment is presented to illustrate the availability of our method to other datasets. The LOMO [57] and SCNCD [44] are adopted as the baseline methods and our aggregation method is evaluated on the PRID 450S dataset. As shown in Fig. 9(c), the matching rate at rank 1 rises form 59.2% to 64.4%, while the relatively improvement is about 9%. In a conclusion, our aggregation method and the proposed parameter selection can be easily applied to other datasets.

*Efficiency:* In this part, we test the runtime of off-line and on-line processes of our aggregation method to illustrates the efficiency. The experimental environment is a desktop PC with an Intel i7-5500U @2.40 GHz CPU. All algorithms are implemented in MATLAB. And the time reported mainly focuses on off-line backward requery and on-line ranking aggregation. Specially, the off-line processing contains LOMO training and SCNCD training, and backward requery. And the on-line processing mainly contains the probe and the ranking aggregation. All the reported time is averaged over 10 random trials on the VIPeR dataset. To speed up our algorithm, we implemented that similarity and dissimilarity part in a simultaneous way, i.e., we compute the step 1–13 in Algorithm 2 parallelly with

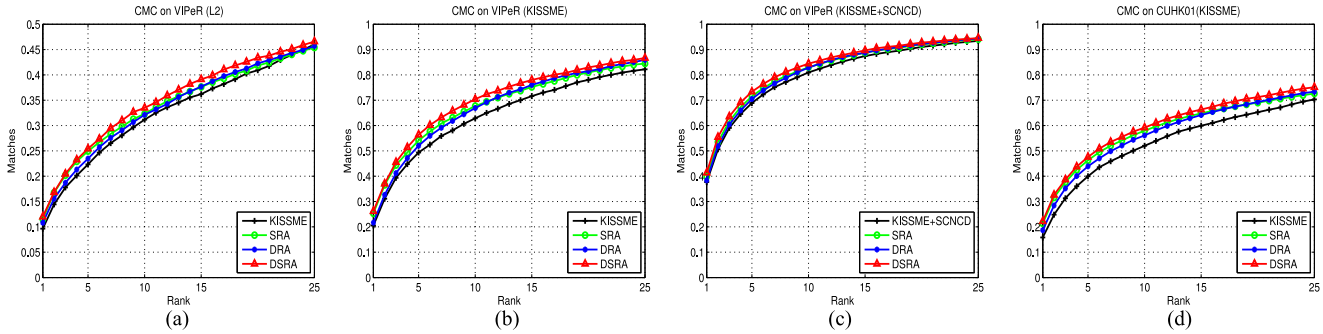



Fig. 11. Effectiveness of our method for a single baseline method. (a) L2 distance based method on VIPeR. (b) Original KISSME [19] on VIPeR. (c) KISSME + SCNCD [44] on VIPeR. (d) KISSME on CUHK01.

the SRA. As can be seen in Table I, online processing is less time-consuming than off-line processing. Although the ranking aggregation leads to a little increase in computation time, it still satisfies some real-time-requirement applications. In addition, the computation time can be further reduced by implementing the algorithm via other efficient languages instead of MATLAB.

### *E. Aggregation of other Baseline Methods*

Table II shows the performance achieved by aggregating other baseline methods to show the universality of our DSRA. Totally, six baseline methods (including a naive L2 distance based method, three classical person re-id methods, i.e., SDALF [18], KISSME [19] and SDC [22], two state-of-the-art baseline methods, i.e., LOMO [57] and SCNCD [44]) are implemented to evaluate the superiority of our ranking aggregation further. They are aggregated in pairs as shown in Table II. As can be seen from this table, the proposed DSRA has achieved consistent improvements compared with the original baseline methods, though at different degrees. This result shows the effectiveness of the proposed algorithm. Particularly, if the two baseline methods have strong complementarity, the improvement would be quite obvious. For example, LOMO is a local descriptor generated by local maximum processing caring more about local information, and SCNCD is an integrated descriptor extracted by PCA focusing on the global miniature. The improvement of their aggregation is about 30% over the better LOMO algorithm due to the strong complementarity of global and local features. On the contrary, the aggregation of SDALF and SCNCD does not perform as well as others. This is because SDALF is also an integrated descriptor which can also be treated as a global-feature based method. Their weak complementarity leads to the limited improvements. On the other hand, if the original baseline methods do not work well, i.e., a naive L2 distance based method, the overall performance will also be limited. It can be seen from the aggregation results between L2 and SDALF, L2 and KISSME. It is concluded that the proposed method can be well applied to combine other baseline methods.

### *F. Optimization for a Single Method*

In this section, we'll present our optimization for a single baseline method. Especially, multi-feature matching based on a classic metric method L2 distance and KISSME [19] are adopted for comparison because their higher speeds facilitate fast testing.

*Parameter analysis:* Two important parameters $k_+$ and $k_-$ are analyzed for the optimization of a single baseline method. Note that the baseline method is the original KISSME [19]. And CMC@1, CMC@5 and CMC@15 on the VIPeR dataset are reported for different $k_+s$ and $k_-s$ as shown in Fig 10. Similar conclusions to previous sections can be made here. The *conductive similarity* mainly improves the ranking orders of quasi-similar galleries, while the *insulative dissimilarity* mainly penalizes the ranking orders of quasi-dissimilar galleries.

*Effectiveness:* To verify the effectiveness of our proposed method, experiments with other three baseline methods are conducted on the VIPeR dataset and the results are shown in Fig. 11(a)–(c) and the results of another experiment on the CUHK01 dataset are shown in Fig. 11(d). The parameters are set as $k_+ = 60$ and $k_- = 60$ in all the experiments. Several conclusions can be drawn from these figures: (1) *Conductive similarity* can truly improve matching results by pulling the quasi-similar galleries. As shown in Fig. 11, our "*SRA*" has 5–10% improvements compared to the baseline methods. (2) *Insulative dissimilarity* ameliorates the results by pushing the quasi-dissimilar galleries. Specially, according to the growth rate of the CMC curves, the "*SRA*" mainly improves the top part of the ranking list which corresponds to the quasi-similar galleries, while the "*DRA*" improves the middle part with a higher increasing speed which corresponds to the quasi-dissimilar galleries. (3) Combination (DSRA) of the two cues truly improves the matching results further.

### *G. Comparison With Other Ranking Optimization Methods*

Firstly, the comparison of our DSRA with other ranking optimization methods is shown in Table III, including some ad hoc re-ranking methods in person re-identification and some general re-ranking algorithms in general object retrieval. It expresses the improvement magnitude. Our ranking aggregation method can categorized into two versions: ranking aggregation for two input baseline methods (Section III, termed as $DSRA_2$) and ranking optimization for a single input baseline method (Section IV, termed as $DSRA_1$). Therefore, the comparison are conducted in two folds: 1) Two ranking fusion methods in general image retrieval, graph fusion [39] and random walk [58] are

TABLE III
COMPARISON ON IMPROVEMENT MAGNITUDE OF DIFFERENT RANKING OPTIMIZATION METHODS (%). "RW_PR" REPRESENTS RANDOM WALK WITH PARTIAL RANKING, WHILE "RW_PRA" IS PARTIAL RANKING WITH ADDITIONAL INFORMATION [58]. "$\text{DSRA}_2$" REPRESENTS OUR DSRA FOR TWO INPUT BASELINE METHODS WHILE "$\text{DSRA}_1$" IS FOR A SINGLE INPUT BASELINE METHOD VERSION. THE MATCHING RATES ARE REPORTED IN PERCENTAGE (%)

| Method \ Rank | $r=1$ | $r=5$ | $r=10$ | $r=25$ |
|---|---|---|---|---|
| KISSME [19] | 22.63 | 50.13 | 63.40 | 82.12 |
| SDC [22] | 23.32 | 43.73 | 54.32 | 68.45 |
| Fusion_Baseline | 31.23 | 58.54 | 69.03 | 82.34 |
| Graph_fusion [39] | 29.13 | 58.51 | 74.15 | 84.52 |
| RW_PRA [58] | 34.18 | 62.71 | 74.02 | 85.13 |
| **$\text{DSRA}_2$** | **39.07** | **68.29** | **79.82** | **90.82** |
| KISSME + RW [58] | 23.56 | 53.24 | 67.05 | 84.23 |
| KISSME + Bi-ranking [30] | 24.56 | 54.46 | 68.67 | 85.34 |
| KISSME + SB [48] | 23.75 | 52.47 | 66.73 | 83.84 |
| **KISSME + $\text{DSRA}_1$** | **25.32** | **56.78** | **70.03** | **86.40** |

compared. For a fair comparison, we implemented all methods under the same baseline (KISSME [19] and SDC [22] are treated as two input baseline methods) and report the matching rates after re-ranking directly. For the *"RW_PRA"*, one of the input baseline method is utilized as prior information which is similar to [58]; 2) We also compared our results with the other two ad hoc ranking optimization methods, which are SB [48] and Bi-ranking [30]. Considering [48] and [30] conduct ranking optimization based on a single input baseline method rather than two methods, we adopt the single input version as described in Section IV for a fair comparison. As their common input, KISSME [19] is used in all these methods as shown in Table III.

As can be seen from Table III, the re-ranking methods in general information retrieval usually do not work well in the person re-identification task. The reason can be ascribed as follows. In general image retrieval, a core assumption is that the expected groundtruths of the query image always appear in the top $k$, but it differs for current person re-identification task, where the top $k$ results contain too many false matchings because the number of ground-truth is much less than general image retrieval [50], making the general algorithm difficult to converge and easily to generate false matching due to the introduction of sample noises. In addition, by introducing the dissimilarity cues, our DSRA has consistent improvements for the re-identification task.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we investigated ranking optimization approaches for the person re-identification problem, and suggested a novel and efficient ranking aggregation method considering both similarity and dissimilarity. Specifically, a combination of similarity and dissimilarity relationships is innovatively utilized for person re-identification task, although dissimilarity relationship is seldom investigated in pervious works. The main idea is that the correct match should be similar to the probe's strongly similar galleries and dissimilar to the strongly dissimilar galleries. In addition, ranking aggregation is firstly introduced in person re-identification task, which is applicable to combine different baseline methods to achieve better results. Extensive experiments on three public datasets with multiple different baseline methods input have validated the effectiveness of our proposed method.

In the list below we outline a number of ideas for future work,

1) Our proposed aggregation method chooses $k_+$ and $k_-$ based on the dataset size, it may produce two problems when conducted in a practical dataset with millions of items, the computational cost will be increased and many not-so-similar items will be regarded as similar. To this end, we will further refine the parameter adjustment in the future. We may choose a small proportion of the original results to conduct the ranking aggregation for the reason that most of the wrong results can be filtered by the original baseline methods. Only a small subset is filtered out for optimization, this strategy will reduce the computational cost significantly.
2) Currently, the proposed method is evaluated on public datasets which are achieved by hand-drawn bounding boxes. We will further investigate this problem under a more open environment (in real applications) to test the effectiveness of the proposed method.
3) We infer that this principle (*conductive similarity* and *insulative dissimilarity*) can also be a generalized to other fine-grained problems, for instance, face recognition. We will further investigate the idea in other applications in the future.

## REFERENCES


[1] S. Sunderrajan and B. S. Manjunath, "Context-aware hypergraph modeling for re-identification and summarization," *IEEE Trans. Multimedia*, vol. 18, no. 1, pp. 51–63, Jan. 2016.
[2] N. Fox, R. Gross, J. F. Cohn, and R. B. Reilly, "Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts," *IEEE Trans. Multimedia.*, vol. 9, no. 4, pp. 701–714, Jun. 2007.
[3] L. L. Presti, S. Sclaroff, and M. L. Cascia, "Path modeling and retrieval in distributed video surveillance databases," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 346–360, Apr. 2012.
[4] M. Ye *et al.*, "Specific person retrieval via incomplete text description," in *Proc. 5th ACM Int. Conf. Multimedia Retrieval*, pp. 547–550, 2015.
[5] J.-W. Hsieh, Y.-T. Hsu, H.-Y. M Liao, and C.-C. Chen, "Video-based human movement analysis and its application to surveillance systems," *IEEE Trans. Multimedia*, vol. 10, no. 3, pp. 372–384, Apr. 2008.
[6] P. Wang *et al.*, "Action recognition from depth maps using deep convolutional neural networks," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 4, pp. 498–509, Aug. 2016.
[7] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sensors J.*, vol. 16, no. 3, pp. 773–781, Feb. 2016.
[8] C. Chen, R. Jafari, and N. Kehtarnavaz, "Improving human action recognition using fusion of depth camera and inertial sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 1, pp. 51–61, Feb. 2015.
[9] K. Hariharakrishnan and D. Schonfeld, "Fast object tracking using adaptive block matching," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 853–859, Oct. 2005.
[10] K.-W. Chen, C.-C. Lai, P.-J. Lee, C.-S. Chen, and Y.-P. Hung, "Adaptive learning for target tracking and true linking discovering across multiple non-overlapping cameras," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 625–638, Aug. 2011.
[11] N. O'Hare and A. F. Smeaton, "Context-aware person identification in personal photo collections," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 220–228, Feb. 2009.

[12] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.
[13] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, Sep. 2016.
[14] W. -S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 649–656.
[15] N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2006, vol. 2, pp. 1528–1535.
[16] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[17] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 262–275.
[18] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 2360–2367.
[19] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, 2288–2295.
[20] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. 10th IEEE Int. Workshop Perform. Eval. Tracking Surveillance*, 2007, pp. 1–7.
[21] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing kiss metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.
[22] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3586–3593.
[23] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2528–2535.
[24] R. Lan, Y. Zhou, and Y. Y. Tang, "Quaternionic local ranking binary pattern: A local descriptor of color images," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 566–579, Feb. 2016.
[25] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 780–793.
[26] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: what features are important?" in *Proc. Eur. Conf. Comput. Vis., Workshops Demonstrations*, 2012, pp. 391–401.
[27] Z. Li *et al.*, "Learning locally-adaptive decision functions for person verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3610–3617.
[28] Y. Wang, R. Hu, C. Liang, C. Zhang, and Q. Leng, "Camera compensation using feature projection matrix for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1350–1361, Aug. 2014.
[29] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analysis for person re-identification," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep.–Oct. 2012, pp. 1621–1624.
[30] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, "Bidirectional ranking for person re-identification," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2013, pp. 1–6.
[31] C. Liu, C. C. Loy, S. Gong, and G. Wang, "POP: Person re-identification post-rank optimisation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 441–448.
[32] Z. Wang, R. Hu, C. Liang, Q. Leng, and K. Sun, "Region-based interactive ranking optimization for person re-identification," in *Proc. 15th Annu. Pacific-Rim Conf. Multimedia*, 2014, pp. 1–10.
[33] A. J. Ma and P. Li, "Query based adaptive re-ranking for person re-identification," in *Proc. 12th Asian Conf. Comput. Vis.*, 2014, pp. 397–412.
[34] M. Ye *et al.*, "Copuled-view based ranking optimization for person re-identification," in *Proc. 21st IEEE Int. Conf. Multimedia Model.*, Jun. 2015, pp. 105–117.
[35] M. Ye, C. Liang, Z. Wang, Q. Leng, and J. Chen, "Ranking optimization for person re-identification via similarity and dissimilarity," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1239–1242.
[36] J. Yu, Y. Rui, and B. Chen, "Exploiting click constraints and multi-view features for image re-ranking," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 159–168, Jan. 2014.
[37] T. Qin, X. Geng, and T. Y. Liu, "A new probabilistic model for rank aggregation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1948–1956.
[38] J. Qin, L. Liu, Z. Zhang, Y. Wang, and L. Shao, "Compressive sequential learning for action similarity labeling," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 756–769, Feb. 2016.
[39] S. Zhang, M. Yang, T. Cour, K. Yu, D. N. Metaxas, "Query specific fusion for image retrieval," in *Proc. 12th Eur. Conf. Comput. Vis., Workshops Demonstrations*, 2012, pp. 660–673.
[40] S. C. H. Hoi and M. R. Lyu, "A multimodal and multilevel ranking scheme for large-scale video retrieval," *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 607–619, Jun. 2008.
[41] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific rank fusion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 803–815, Apr. 2014.
[42] L. Wei and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 3594–3601.
[43] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof, "Mahalanobis distance learning for person re-identification," in *Person Re-Identification*. New York, NY, USA: Springer, 2014.
[44] J. Yan *et al.*, "Salient color names for person re-identification," in *Proc. 13th Eur. Conf. Comput. Vis.*, 2014, pp. 536–551.
[45] W. S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.
[46] S. Ali, O. Javed, N. Haering, and T. Kanade, "Interactive retrieval of targets for wide area surveillance," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, 895–898.
[47] R. Nabiei, M. Najafian, M. Parekh, P. Jancovic, and M. Russell, "Delay reduction in real-time recognition of human activity for stroke rehabilitation," in *Proc. 1st Int. Workshop Sens. Process. Learning Intell. Mach.*, 2016, pp. 1–5.
[48] L. An, C. Chen, M. Kafai, S. Yang, and B. Bhanu, "Improving person re-identification by soft biometrics based reranking," in *Proc. 7th Int. Conf. Distrib. Smart Cameras*, 2013, pp. 1–6.
[49] C. Liang *et al.*, "A unsupervised person re-identification method using model based representation and ranking," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 771–774.
[50] L. Zheng *et al.*, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1116–1124.
[51] W. Zheng *et al.*, "Zero-shot person re-identification via cross-view consistency," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 260–272, Feb. 2016.
[52] C. Zhang *et al.*, "Fine-grained image classification via low-rank sparse coding with general and class-specific codebooks," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
[53] L. Zheng, S. Wang, Z. Liu, and Q. Tian, "Fast image retrieval: Query pruning and early termination," *IEEE Trans. Multimedia*, vol. 17, no. 5, pp. 648–659, May 2015.
[54] L. Zheng *et al.*, "Query-adaptive late fusion for image search and person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 1741–1750.
[55] C. Zhang, Q. Huang, and Q. Tian, "Contextual exemplar classifier based image representation for classification," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
[56] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, "Person re-identification with content and context re-ranking," *Multimedia Tools Appl.*, vol. 74, pp. 6989–7014, 2015.
[57] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 2197–2206.
[58] W. H. Hsu, L. S. Kennedy, and S. -F. Chang, "Video search reranking through random walk over document-level context graph," in *Proc. 15th ACM Int. Conf. Multimedia*, 2007, pp. 971–980.

**Mang Ye** received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2013 and 2016, respectively, and is currently working toward the Ph.D degree in computer science at the Hong Kong Baptist University, Hong Kong, China.

His research interests include multimedia content analysis and retrieval, and computer vision and pattern recognition.

**Chao Liang** received the Ph.D degree from the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2012.

He is currently an Associate Professor with the National Engineering Research Center for Multimedia Software, Computer School, Wuhan University, Wuhan, China. His research interests include multimedia content analysis and retrieval, computer vision and pattern recognition. He has authored or coauthored more than 30 papers, including for conferences such as Computer Vision and Pattern Recognition and ACM Multimedia, and journals such as the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Prof. Liang was the recipient of the Best Paper Award of PCM 2014.

**Yi Yu** received the Ph.D. degree in information and computer science from Nara Women's University, Nara, Japan.

She is currently an Assistant Professor with the National Institute of Informatics (NII), Tokyo, Japan. Before joining NII, she was a Senior Research Fellow with the School of Computing, National University of Singapore, Singapore. Her research interests include large-scale multimedia data mining and pattern analysis, location-based mobile media service, and social media analysis.

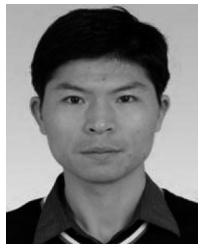
**Zheng Wang** received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2006 and in 2008, respectively, and is currently working toward the Ph.D degree at the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China.

His research interests focus on multimedia content analysis and retrieval, and computer vision and pattern recognition.

Mr. Wang was the recipient of the Best Paper Award at the 15th Pacific-Rim Conference on Multimedia.

**Qingming Leng** received the B.S. degree in life science from Nanchang University, Nanchang, China, in 2007, the M.S. degree from the International School of Software, Wuhan University, Wuhan, China, in 2009, and the Ph.D. degree from the National Engineering Research Center for Multimedia Software, Wuhan University, in 2014.

He is currently a Lecturer with the School of Information Science and Technology, Jiujiang University, Jiujiang, China. His research interests include computer vision, machine learning, and person reidentification.



**Chunxia Xiao** (M'11) received the B.Sc. and M.Sc. degrees in mathematics from the Hunan Normal University, Hunan, China, in 1999 and 2002, respectively, and the Ph.D. degree from the State Key Lab of Computer Aided Design and Computer Graphics (CAD&CG), Zhejiang University, Hangzhou, China, in 2006.

He is currently a Professor with the School of Computer, Wuhan University, Wuhan, China. From October 2006 to April 2007, he was a Postdoc with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China, and from February 2012 to February 2013, he visited the University of California at Davis, Davis, CA, USA. He has authored or coauthored more than 50 papers in journals and conferences. His research interests include computer graphics, computer vision, and image and video processing.

**Jun Chen** received the M.S. degree in instrumentation from the Huazhong University of Science and Technology, Wuhan, China, in 1997, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008.

He is the Deputy Director of the National Engineering Research Center for Multimedia Software, Wuhan University. His research interests include multimedia communications and security emergency information processing.

**Ruimin Hu** (M'09–SM'09) received the B.S. and M.S. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1984 and 1990, respectively, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1994.

He is the Dean of the School of Computer, Wuhan University. He has authored or coauthored two books and more than 100 scientific papers. His research interests include audio/video coding and decoding, and video surveillance and multimedia data processing.