Group Re-Identification: Leveraging and Integrating Multi-Grain Information

Hao Xiao[1], Weiyao Lin[1*], Bin Sheng[1], Ke Lu[2], Junchi Yan[1], Jingdong Wang[3],
Errui Ding[4], Yihao Zhang[5], Hongkai Xiong[1]

[1] Shanghai Jiao Tong University, China (*Corresponding author: wylin@sjtu.edu.cn)
[2] University of Chinese Academy of Sciences, China
[3] Microsoft Research, Beijing, China
[4] Department of Computer Vision Technology (VIS), Baidu Inc., China
[5] Tencent YouTu Lab, China

ABSTRACT

This paper addresses an important yet less-studied problem: re-identifying groups of people in different camera views. Group re-identification (Re-ID) is very challenging since it is not only interfered by view-point and human pose variations in the traditional single-object Re-ID tasks, but also suffers from group layout and group member variations. To handle these issues, we propose to leverage the information of multi-grain objects: individual person and subgroups of two and three people inside a group image. We compute multi-grain representations to characterize the appearance and spatial features of multi-grain objects and evaluate the importance weight of each object for group Re-ID, so as to handle the interferences from group dynamics. We compute the optimal group-wise matching by using a multi-order matching process based on the multi-grain representation and importance weights. Furthermore, we dynamically update the importance weights according to the current matching results and then compute a new optimal group-wise matching. The two steps are iteratively conducted, yielding the final matching results. Experimental results on various datasets demonstrate the effectiveness of our approach.

KEYWORDS

Re-identification; Multi-grain representation; Group Re-ID

Figure 1: (a)-(b) Left: Probe groups in cam A; Right: The correct match groups (up) and confusing groups (down) in cam B. (c): Illustration of deriving multi-grain information for group Re-ID. The colored lines and rectangles in (c) indicate the importance weights for people/people subgroups. (Best viewed in color)

1 INTRODUCTION

Person re-identification (Re-ID) aims at matching and identifying pedestrians across non-overlapping camera views. It is of increasing importance in visual surveillance and has attracted many researches [7, 11, 14, 23, 37–40]. However, most researches focus on individual person Re-ID, while the Re-ID of groups of people are seldom studied. In practice, since most events (e.g., fighting or violent actions) are performed by groups instead of an individual, when people analyze events across cameras, they are more interested in identifying groups rather than individuals [5, 44]. Therefore, it is non-trivial to obtain reliable group matching across cameras.

There are two more basic challenges for group Re-ID besides viewpoint changes or human pose variations [14, 41] for each individual person. (i) Group layout change: The layout of people in a group often show large and uncontrolled changes in different camera views. For example, in Fig. 1a, due to the dynamic movements of people, the relative positions of people in a group have large differences in two camera views. (ii) Group membership change: people may often join or leave a group (cf. Fig. 1b).
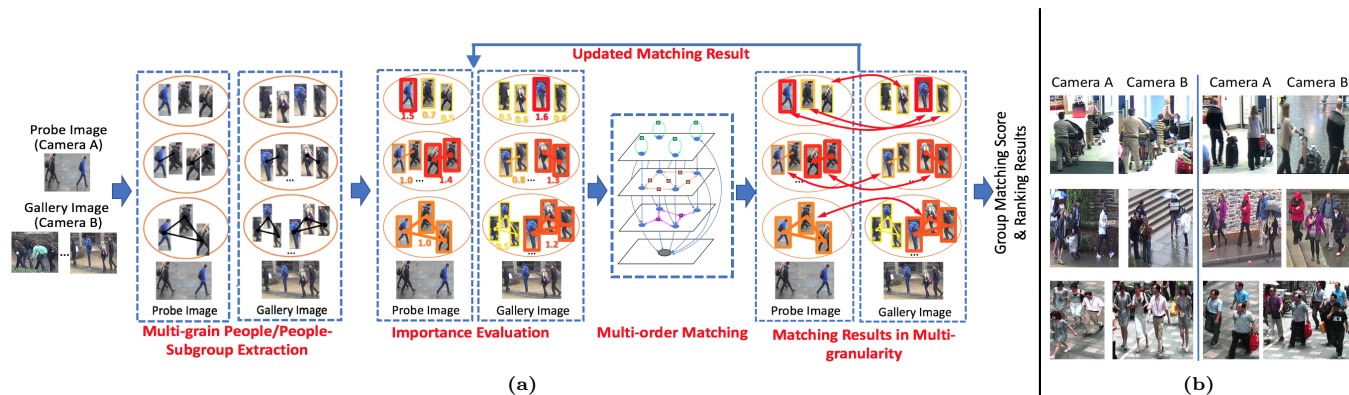
Figure 2: (a) Framework of the proposed approach. (b) Example groups for the three datasets used in our experiment. First row: the public *i-LID MCTS* dataset; Second and third rows: our constructed *DukeMTMC Group* and *Road Group* datasets. (Best viewed in color)

Most existing methods, e.g., [5, 15, 44], view the input group image as an entire unit and extract global/semi-global features without explicitly doing individual people matching and considering layout changes to perform group-wise matching [5, 44]. A recent study [46] attempts to use descriptors of local patches to partially handle layout change and membership change.

In this paper, we introduce the idea of *group granularity* and characterize a group image by multi-grain objects: fine-grain objects each formed by a single person, medium-grain objects each formed by a group of two people, coarse-grain objects each formed by a group of three people, and global-grain object formed by the group of all people. We argue that the multi-grain object characterisation is helpful to handle layout and membership changes.

Let's look at the example shown in Fig. 1a. Due to the large layout variation and camera viewpoint change, the same group shows large global appearance differences in two camera views. The re-id performance is poor if merely adopting global features for the entire group for re-identification (cf. the down-right image in Fig. 1a). This issue can be resolved if we include information of finer group granularity (e.g., individual people). On the other hand, merely using the information of individual people is also not always reliable. An example is shown in Fig. 1b: the two groups in camera B include visually similar group members to the probe group in camera A. In this case, the information of medium-level granularities (e.g., subgroups of two people) is helpful.

Our approach leverages the representations of multi-grain objects (cf. Fig. 1c) for group re-identification. In addition, motivated by the observations that groups in different cameras may be interfered by group member variation, occlusion, and mismatching, and that multi-grain objects have different reliabilities on Re-ID performances, we propose to introduce the importance weights for objects in each granularity to improve the re-identification performance.

In summary, our contributions lie in three folds: (1) We introduce the multi-grain representations for a group image

to better handle the layout change and membership change issues, and the dynamic weighting scheme for better person matching. (2) We solve the group-wise matching problem by using the multi-order matching algorithm, which integrates multi-grain representation and combines the information of both matched and unmatched objects to achieve a more reliable matching result. (3) We create two challenging group Re-ID datasets with large group member and layout variations as the existing group Re-ID datasets contain relatively small variations in group member and group layout.

2 RELATED WORKS AND OVERVIEW

Person Re-ID has been studied for years. Most of them focus on developing reliable features [7, 11, 17, 28, 29, 33], deriving accurate feature-wise distance metric [2, 6, 8, 23, 35, 37, 38], and handling local spatial misalignment between people [14, 26, 30, 31, 40, 41]. Some recent research works extend Re-ID algorithms to more object types (e.g., cars [19, 24]) or more complex scenarios (e.g., larger camera numbers [10], long-term videos [16, 42, 45], untrimmed images [32, 34, 43]).

Most existing works focus on the Re-ID of individual person and the group-level Re-ID problem is seldom considered. Since group Re-ID contains significant group layout change and group member variation, it introduces new challenges and meanwhile more information to explore compared with single/few-person view matching scenario as addressed by single person Re-ID methods. Although some works [1, 3, 12, 27] introduce people interaction into Re-ID process, they are only targeted at improving the Re-ID performance of individual person. The characteristics of groups are still less considered and not fully modeled.

Only a few works are developed to address group Re-ID tasks [5, 15, 44, 46]. Most of them develop global or semi-global features to perform group-wise matching. For example, Cai et al. [5] propose a discriminative covariance descriptor to capture the global appearance & statistic properties of group images. Zheng et al. [44] segment a group image into multiple ring regions and derive semi-global descriptors

for each region. Lisanti et al. [15] combine sparsity-driven descriptions of all patches into a global group representation. Since global/semi-global features cannot capture the local interaction information in groups, they have limitations in handling complex scenarios with significant global appearance variations.

Recently, Zhu et al. [46] developed a local-based method which performs group Re-ID by selecting proper patch-pairs and conducting patch matching between cross-view group images. However, in order to reduce patch mismatches, this method includes prior restrictions on vertical misalignments. This limits their capability in handling significant group layout changes or group member variations.

Our approach differs from the existing group Re-ID works in two aspects: (1) The existing works perform Re-ID with single-grain information (i.e., either global or patch level information). Comparatively, our approach leverages multi-grain information to fully capture the characteristics of a group. (2) Our approach does not include any prior restrictions on spatial misalignments, which is able to handle arbitrary group layout changes or group member variations.

Overview of our approach. Given the probe group image captured from one camera, our goal is to find the matched group images from a set of gallery group images captured from another camera. We represent each group image by a set of multi-grain objects, and extract the features for the multi-grain objects. The matching process is an iterative process. We compute the static and dynamic importance weights of multi-grain objects for the probe and gallery images according to the intermediate matching results. Then, we use a multi-order matching algorithm to compute intermediate matching results, which are used to update the dynamic importance weights. We perform the two stages iteratively, and obtain the final matching results. The entire framework is shown in Fig. 2a.

3   MULTI-GRAIN REPRESENTATION

A group image I contains a set of people: $\mathcal{G} = \{b_1, b_2, \ldots, b_N\}$, where N is the number of people and b_i (or simply denoted by i for presentation clarity) corresponds to the person bounding box. The representation is computed by building multi-grain objects (people/subgroups): 1) Fine granularity, including objects of individual person, $\mathcal{O}_1 = \{i | i = 1, \ldots, N\}$; 2) Medium granularity, including objects of two-people subgroups, $\mathcal{O}_2 = \{(i_1, i_2) | i_1, i_2 = 1, \ldots, N, i_1 \neq i_2\}$; 3) Coarse granularity, including objects of three-people subgroups, $\mathcal{O}_3 = \{(i_1, i_2, i_3) | i_1, i_2, i_3 = 1, \ldots, N, i_1 \neq i_2 \neq i_3\}$; and 4) Global granularity, referring to the entire group, $\mathcal{O}_g = \{(1, 2, \ldots, N)\}$. In the extreme case there are only two people in the group image, we simply let $\mathcal{O}_2$ be the coarse granularity.

The fine granularity helps reduce the confusion in the global appearance under large layout or group member changes, while the medium and coarse granularities, capable of capturing the local layout structure in a group, help resolve the ambiguous subgroup/person matches in the fine granularity

by incorporating local layout or co-occurrence information in a group.

The feature of an object o in the fine granularity, denoted by $\mathbf{f}_o$, is about the appearance. The feature of an object o for the medium and coarse granularities, denoted by $\mathbf{f}_o$, consists of two parts: appearance, which is an aggregation of the appearances of all people, and spatial relation. We use the color and texture feature [8] as the *appearance* part of a person. We use the relative distance & angle histograms among individual people in an object [9] as the spatial relation part. The representation for the global granularity is the color+texture features of the entire group image.

4   IMPORTANCE WEIGHTING

We introduce an importance weight α_o for each object o (except the global-grain object) to indicate the object's discriminativity and reliability inside the probe group image for group person matching. The importance weighting scheme is partially inspired by but different from the saliency-learning methods [40, 41, 46] for differentiating *patch* reliabilities in person re-identification: (i) Our scheme aims to weight each granularity object rather than patches; (ii) Our scheme dynamically adjusts the importance weights in an iterative manner, by using the intermediate matching results at each iteration. (cf. Fig. 2a).

4.1   Fine-grain Object

The importance weight (α_i) for each individual person (o_i) in the probe image I consists of two components: static weight, which is only dependent on the group image, and dynamic weight, which is dynamically updated according to the intermediate matching results with the gallery group images, from another camera in our approach. The formulation is given as follows,

$$\alpha_i = t_1(i, \mathcal{G}_{\setminus i}) + s(i, \mathcal{M}_i) + p(\mathcal{M}_i, \mathcal{M}_{\mathcal{G}_{\setminus i}}), \qquad (1)$$

where the first term is the static weight, and the second and third terms form the dynamic weight.

Static weight. The static weight $t_1(i, \mathcal{G}_{\setminus i})$, where $\mathcal{G}_{\setminus i} = \mathcal{G} - \{i\}$ denotes the set of other individual people in $\mathcal{G}$, is used to describe the stability. It is computed as follows,

$$t_1(i, \mathcal{G}_{\setminus i}) = \lambda_{t_1} \sum_{i' \in \mathcal{G}_{\setminus i}} \frac{\rho_i}{\rho_{i'}}, \qquad (2)$$

where ρ_i is the local density around person i in group $\mathcal{G}$. It reflects the density of people in a neighborhood around i, which is computed by following [4].

By Eq. 2, the static weight t_1 is mainly obtained by evaluating the relative local density ratios between person i and his/her peer group members i' in $\mathcal{G}$. If the local density around i is larger than the density around his/her peer group members i', the stability of i is increased, indicating that i is located in the *center* region of group $\mathcal{G}$ and should be a more reliable member in group Re-ID (cf. person 1 in Fig. 3a). On the contrary, when i's local density is smaller than his/her peer group members, a small stability value will be assigned,

indicating that i is located in the outlier region of the group and is less reliable (cf. person 2 in Fig. 3a).

Dynamic weight. The dynamic weight $s(i, \mathcal{M}_i) + p(\mathcal{M}_i, \mathcal{M}_{\mathcal{G} \setminus i})$ consists of two terms: the saliency term $s(i, \mathcal{M}_i)$ and the purity term $p(\mathcal{M}_i, \mathcal{M}_{\mathcal{G} \setminus i})$, where $\mathcal{M}_i$ is the set of matches from the gallery group images, and $\mathcal{M}_{\mathcal{G} \setminus i}$ is the set of the matches for all people except i in the probe image, $\mathcal{M}_{\mathcal{G} \setminus i} = \{\mathcal{M}_{i'} | i' \notin \mathcal{G}\}$. The sets of matches are illustrated in Figure 3a.

The saliency term is computed as

$$s(i, \mathcal{M}_i) = \lambda_s \frac{d_f(\mathbf{f}_i, \mathbf{f}_{\mathcal{M}_i})}{|\mathcal{M}_i|}. \tag{3}$$

Here $d_f(\cdot)$ is the Euclidean distance between features. $|\mathcal{M}_i|$ is the cardinality of $\mathcal{M}_i$. $\mathbf{f}_{\mathcal{M}_i}$ is the feature describing the set of matches $\mathcal{M}_i$, and we use the feature of the $\frac{1}{2}|\mathcal{M}_i|$th nearest neighbor of i in $\mathcal{M}_i$ as done in [40, 41]. λ_s is a constant which normalizes the range of s to be within 0 to 1.

According to Eq. 3, if the appearance of an individual person i is discriminative, a large portion of individuals in i's matched set $\mathcal{M}_i$ are visually dissimilar to i. This leads to a large $d_f(\mathbf{f}_i, \mathbf{f}_{\mathcal{M}_i})$ and a large saliency value [40, 41] (cf. person 1 in Fig. 3a). Moreover, due to the variation of group members in group Re-ID, each individual person may have different number of matched people in his/her $\mathcal{M}_i$. Therefore, we further introduce $|\mathcal{M}_i|$ in Eq. 3, such that person with fewer matched people can indicate more discriminative appearance.

The purity term is computed as

$$p(\mathcal{M}_i, \mathcal{M}_{\mathcal{G} \setminus i}) = \sum_{i' \in \mathcal{G} \setminus i} \lambda_p d_m(\mathcal{M}_i, \mathcal{M}_{i'}), \tag{4}$$

where $d_m(\cdot)$ is the Wasserstein-1 distance [22], a measure to evaluate the dissimilarity between two feature sets. λ_p is calculated in the same way as λ_s in Eq. 3.

According to Eq. 4 and Fig. 3a, the purity measurement reflects the relative appearance uniqueness of person i inside group $\mathcal{G}$. If i has similar appearance features as other group members in $\mathcal{G}$, their matched people in camera B should also be visually similar and located close to each other in the feature space (cf. $\mathcal{M}_3$ and $\mathcal{M}_4$ in Fig. 3a), resulting in a small purity value. On the other hand, if a person includes *unique* appearance features in $\mathcal{G}$, his/her matched people in camera B should have larger feature distances to those of the other members in $\mathcal{G}$, and lead to a large purity value (cf. $\mathcal{M}_1$ in Fig. 3a).

4.2 Medium and Coarse Grain Objects

The importance weight $\alpha_{i_1 i_2}$ of a medium-grain object (i_1, i_2) is computed as:

$$\alpha_{i_1 i_2} = \alpha_{i_1} + \alpha_{i_2} + t_2(i_1, i_2). \tag{5}$$

Here, $t_2(i_1, i_2)$ is the stability measure of the sub-group (i_1, i_2). A two-people sub-group is thought more stable if its members are spatially closer to each other. Thus, we simply compute t_2 by the inverse of spatial distance between i_1 and i_2.

(a)   **(b)**
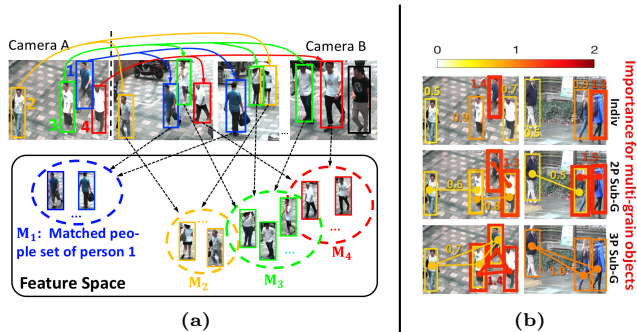
Figure 3: (a) Illustration of matched-people sets and their distributions in the feature space (The color solid arrows indicate the one-to-one mapping results between individuals. People circled by the same-color rectangles in camera B are matched to the same person in A, and belong to the same matched-people set). (b) The derived importance weights for multi-grain objects (individuals, 2-people subgroups, 3-people subgroups) in two group images. Note: the importance weights for some 2-people/3-people subgroups are not displayed in order for a cleaner illustration. (Best viewed in color)

The importance weight $\alpha_{i_1 i_2 i_3}$ of a coarse-grain object (i_1, i_2, i_3) is computed as:

$$\alpha_{i_1 i_2 i_3} = \alpha_{i_1 i_2} + \alpha_{i_2 i_3} + \alpha_{i_1 i_3} + t_3(i_1, i_2, i_3). \tag{6}$$

Here, $\alpha_{i_1 i_2}$ is the importance of a two-people pair in (i_1, i_2, i_3) (cf. Eq. 5). t_3 is the stability of a three-people subgroup. We assume equilateral triangle as the most stable structure for three-people subgroups and model t_3 by evaluating its similarity to equilateral triangle.

Fig. 3b shows the importance weights of some groups. From Fig. 3b, we can see that our process can effectively set larger weights on more discriminative & reliable objects.

4.3 Iterative Update

We utilize an iterative process which updates the importance weights and group-wise matching results iteratively. We initialize the dynamic weights for all objects by 1 and compute the optimal matching through multi-order matching (cf. Sec. 5) to obtain an initial matching result: $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_N$. This matching result is used to update the dynamic importance weights. This procedure is repeated until the importance weights become converged or the maximum iteration is reached. Although the exact convergence of our iterative process is difficult to analyze due to the inclusion of multi-order matching, our experiments show that most importance weights become stable within 5 iterations, which implies the reliability of our approach.

5 MULTI-ORDER MATCHING

Given a probe image I_p and a gallery image I_g, our goal is to compute the matching score between the two groups of

people. Suppose that there are N_p people in the probe image I_p and N_g people in the gallery image I_g. The goal of the multi-order matching process aims to find: (1) an optimal one-to-one mapping, $\mathcal{C} = \{(i,j)|\ \forall (i,j), (i',j'), i \neq i', j \neq j'\}$, where $(i,j)\ (= c_{ij})$ denotes a match between the ith person from the probe image and the jth person from the gallery image, and (2) the matching score.

The objective function is formulated with multi-order potentials:

$$\mathcal{Q}(\mathcal{C}) = \mathcal{P}_1(\mathcal{C}) + \mathcal{P}_2(\mathcal{C}) + \mathcal{P}_3(\mathcal{C}) + \mathcal{P}_g(\mathcal{C})$$
$$+ \sum_{k_1 \neq k_2, k_1, k_2 = 1,2,3,g} \mathcal{P}_{k_1 k_2}(\mathcal{C}), \qquad (7)$$

where $\mathcal{P}_1(\mathcal{C})$, $\mathcal{P}_2(\mathcal{C})$, $\mathcal{P}_3(\mathcal{C})$, and $\mathcal{P}_g(\mathcal{C})$ are the first-order, second-order, third-order, and global potentials, evaluating the matching quality over each subgroup of people, and $\mathcal{P}_{k_1 k_2}(\mathcal{C})$ is the inter-order potential.

5.1 Multi-order Potentials

First order potential. $\mathcal{P}_1(\mathcal{C})$ is used to model the matching scores over individual people. It is calculated by the sum of the matching scores of all the individual matches in $\mathcal{C}$:

$$\mathcal{P}_1(\mathcal{C}) = \sum_{c_{ij} \in \mathcal{C}} m_1(c_{ij}) = \sum_{c_{ij} \in \mathcal{C}} w_1(\mathbf{f}_i, \alpha_i, \mathbf{f}_j, \alpha_j) \qquad (8)$$

where $\mathbf{f}_i$, α_i and $\mathbf{f}_j$, α_j are the feature vector and importance weight for probe-image person i and gallery-image person j, respectively (cf. Eq. 1). $m_1(c_{ij}) = w_1(\mathbf{f}_i, \alpha_i, \mathbf{f}_j, \alpha_j)$ is the matching score for match $c_{ij} = (i,j)$, calculated by:

$$m_1(c_{ij}) = w_1(\mathbf{f}_i, \alpha_i, \mathbf{f}_j, \alpha_j) = \lambda_{w_1} \frac{\psi(\alpha_i, \alpha_j)}{d_{\mathbf{f}}(\mathbf{f}_i, \mathbf{f}_j)} \qquad (9)$$

where $\psi(\alpha_i, \alpha_j) = \frac{\alpha_i + \alpha_j}{1 + |\alpha_i - \alpha_j|}$ is the fused importance weight, which will have large value if the importance weights of α_i and α_j are both large and close to each other. $D_{\mathbf{f}}(\cdot)$ is the Euclidean distance and λ_{w_1} is the normalization constant for the first-order potential.

By Eq. 9, the matching score $m_1(c_{ij})$ is computed by the importance-weighted feature similarity $w_1(\mathbf{f}_i, \alpha_i, \mathbf{f}_j, \alpha_j)$ between the matched individuals i and j.

Second order potential. $\mathcal{P}_2(\mathcal{C})$ is used to model the matching scores over two-people subgroups:

$$\mathcal{P}_2(\mathcal{C}) = \sum_{c_{i_1 j_1}, c_{i_2 j_2} \in \mathcal{C}} m_2(c_{i_1 j_1}, c_{i_2 j_2})$$
$$= \sum_{c_{i_1 j_1}, c_{i_2 j_2} \in \mathcal{C}} w_2(\mathbf{f}_{i_1 i_2}, \alpha_{i_1 i_2}, \mathbf{f}_{j_1 j_2}, \alpha_{j_1 j_2}). \qquad (10)$$

where $\mathbf{f}_{i_1 i_2}$, $\alpha_{i_1 i_2}$ and $\mathbf{f}_{j_1 j_2}$, $\alpha_{j_1 j_2}$ are the feature vector and importance weight for probe-image subgroup (i_1, i_2) and gallery-image subgroup (j_1, j_2), respectively (cf. Eq. 5). $m_2(c_{i_1 j_1}, c_{i_2 j_2}) = w_2(\mathbf{f}_{i_1 i_2}, \alpha_{i_1 i_2}, \mathbf{f}_{j_1 j_2}, \alpha_{j_1 j_2})$ is the second order match score between two-people subgroups (i_1, i_2) and (j_1, j_2), which is calculated in a similar way as Eq. 9:

$$w_2(\mathbf{f}_{i_1 i_2}, \alpha_{i_1 i_2}, \mathbf{f}_{j_1 j_2}, \alpha_{j_1 j_2}) = \lambda_{w_2} \frac{\psi(\alpha_{i_1 i_2}, \alpha_{j_1 j_2})}{d_{\mathbf{f}}(\mathbf{f}_{i_1 i_2}, \mathbf{f}_{j_1 j_2})} \qquad (11)$$
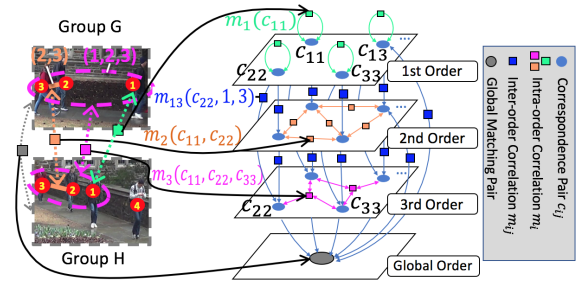
Figure 4: Illustration of multi-order association graph. Left: A cross-view group pair being matched; Right: The multi-order association graph constructed for the group pair. (Best viewed in color)

Third order potential. $\mathcal{P}_3(\mathcal{C})$ is used to model the matching scores over three-people subgroups:

$$\mathcal{P}_3(\mathcal{C}) = \sum_{c_{i_1 j_1}, c_{i_2 j_2}, c_{i_3 j_3} \in \mathcal{C}} m_3(c_{i_1 j_1}, c_{i_2 j_2}, c_{i_3 j_3})$$
$$= \sum_{c_{i_1 j_1}, c_{i_2 j_2}, c_{i_3 j_3} \in \mathcal{C}} w_3(\mathbf{f}_{i_1 i_2 i_3}, \alpha_{i_1 i_2 i_3}, \mathbf{f}_{j_1 j_2 j_3}, \alpha_{j_1 j_2 j_3}). \qquad (12)$$

where $w_3(\mathbf{f}_{i_1 i_2 i_3}, \alpha_{i_1 i_2 i_3}, \mathbf{f}_{j_1 j_2 j_3}, \alpha_{j_1 j_2 j_3}) = m_3(c_{i_1 j_1}, c_{i_2 j_2}, c_{i_3 j_3})$ is the third order match score between three-people subgroups (i_1, i_2, i_3) and (j_1, j_2, j_3). It is calculated in the same way as Eq. 11.

Global potential. The global potential is calculated by the global matching score between probe and gallery images I_p and I_g:

$$\mathcal{P}_g(\mathcal{C}) = \sum_{\mathcal{C}} m_g(c_{i_1 j_1}, c_{i_2 j_2}, \ldots, c_{i_{N_p} j_{N_q}})$$
$$= w_g(\mathbf{f}_p, \alpha_p, \mathbf{f}_g, \alpha_g) \qquad (13)$$

where $\mathbf{f}_p$ and $\mathbf{f}_g$ are the global feature vectors for the entire group images I_p and I_g. $\alpha_p = \alpha_g = 1$ are the importance weights for global objects. In this paper, we simply use the global feature similarity as the global matching score, as: $w_g(\mathbf{f}_p, \alpha_p, \mathbf{f}_g, \alpha_g) = \frac{1}{D_{\mathbf{f}}(\mathbf{f}_p, \mathbf{f}_g)}$.

Inter-order potential. Since each match c_{ij} is described by potentials in multiple orders (cf. Eqs. 8-13), we also introduce inter-order potentials to properly combine these multi-order potential information. Specifically, the inter-order potential between orders k_1 and k_2 is calculated by:

$$\mathcal{P}_{k_1 k_2}(\mathcal{C}) = \sum_{c_{ij} \in \mathcal{C}} m_{k_1 k_2}(c_{ij}, k_1, k_2) \qquad (14)$$

where $m_{k_1 k_2}(c_{ij}, k_1, k_2)$ is the inter-order correlation for match c_{ij}. It is calculated by:

$$m_{k_1 k_2}(c_{ij}, k_1, k_2) = \frac{\overline{m}_{k_1}(c_{ij}, k_1) + \overline{m}_{k_2}(c_{ij}, k_2)}{1 + |\overline{m}_{k_1}(c_{ij}, k_1) - \overline{m}_{k_2}(c_{ij}, k_2)|}$$
$$\text{for } \overline{m}_k(c_{ij}, k) = \lambda_k \sum_{c_{i_1' j_1'} = c_{ij}} m_k(c_{i_1' j_1'} \ldots c_{i_k' j_k'}) \qquad (15)$$
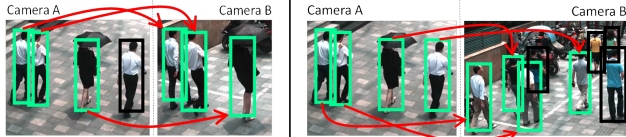
Figure 5: Illustration of the unmatched term in E-q. 16. (a) is a true match pair and (b) is a false match pair. Green and black rectangles are matched and unmatched individuals, respectively. Since the right group in (b) includes more individuals, we can find more matched pairs. This may misleadingly create a high similarity score. However, if considering the large number of unmatched people in (b), the matching score of (b) can be properly reduced.

where λ_k is the normalization constant for order k, $m_k(c_{i'_1 j'_1}...c_{i'_k j'_k})$ is the intra-order match score in order k (as in Eqs. 9 and 11). From Eq. 15, if a match c_{ij} creates large and similar intra-order match scores in both k_1 and k_2 orders, it will be considered as more valuable and reliable, and thus will have larger inter-level potentials.

5.2 Optimization

The objective function in Eq. 7 properly integrates the information of multi-grain objects. Thus, by maximizing Eq. 7, we are able to obtain the optimal one-to-one mapping result among individuals in probe and gallery groups.

To solve the multi-order matching problem in Eq. 7, we construct a multi-order association graph to incorporate all candidate matches & multi-order potentials in the objective function, as in Fig. 4. In Fig. 4, each layer includes all candidate matches c_{ij} (the circular nodes) and their corresponding intra-order matching scores m_k (the rectangular nodes in green, orange, or pink), which models the intra-order potentials in a specific order. Besides, the blue rectangular nodes linking circular nodes in different layers represent the inter-order correlations $m_{k_1 k_2}(c_{ij}, k_1, k_2)$. They model the inter-order potentials between different orders.

With this association graph, we are able to solve Eq. 7 by adapting general-purpose hyper-graph matching solvers [13, 20, 36]. Specifically, we first initialize a mapping probability for each candidate match in the associate graph, and then apply reweighted random walk [13] to update these mapping probabilities via the inter/intra-order links and potential weights in the association graph. Finally, the mapping probabilities in all layers in the association graph are combined to obtain the optimal one-to-one mapping result from the candidate matches [36].

5.3 Fused Matching Score

After obtaining one-to-one mapping between individual people in two groups, we are able to calculate matching scores accordingly. In order to obtain a more reliable matching score, we introduce a fused scheme by integrating the information of both matched and unmatched objects:

$$S(I_p, I_g) = \sum_k \sum_{(i_1..i_k) \in \mathcal{R}_p} \frac{w_k(\mathbf{f}_{i_1..i_k}, \alpha_{i_1..i_k}, \mathbf{f}_{M(i_1..i_k)}, \alpha_{M(i_1..i_k)})}{|\mathcal{R}_p|}$$

$$- \lambda_s \cdot \sum_k \left(\sum_{(i_1..i_k) \in \overline{\mathcal{R}}_p} \frac{a_{i_1..i_k}}{|\overline{\mathcal{R}}_p|} + \sum_{(j_1..j_k) \in \overline{\mathcal{R}}_g} \frac{a_{j_1..j_k}}{|\overline{\mathcal{R}}_g|} \right) \quad (16)$$

where $(i_1,..,i_k)$ is a person/subgroup in probe group image I_p, $M(i_1..i_k)$ is its one-to-one matched person/subgroup in gallery image I_g. $w_k(\cdot)$ is the similarity matching score between $(i_1,..,i_k)$ and $M(i_1..i_k)$, as in Eqs. 9 and 11. α is the importance weight. λ_s=0.5 is a balancing factor. $\mathcal{R}_p$ and $\mathcal{R}_g$ are the sets of reliably matched objects in groups I_p and I_g, and $\overline{\mathcal{R}}_p$ and $\overline{\mathcal{R}}_g$ are the unmatched object sets. We select matched object pairs with large similarity values as the reliably matched objects, and put them into $\mathcal{R}_p$ and $\mathcal{R}_g$. The remaining unmatched or less similar objects are put into $\overline{\mathcal{R}}_p$ and $\overline{\mathcal{R}}_g$.

From Eq. 16, our fused scheme integrates four granularities (i.e., $k = 1, 2, 3, g$) to compute the group-wise matching score. Inside each granularity, we select matched object pairs with high similarities to compute the similarity (the first term in Eq. 16), so as to reduce the interference of confusing or mismatched people/people subgroups. Meanwhile, we introduce an unmatched term evaluating the importance of unmatched objects (the second term in Eq. 16). As such, we can properly avoid misleadingly high matching scores in false group pairs (as in Fig. 5) and obtain a more reliable result.

6 EXPERIMENTAL RESULTS

We perform experiments on three datasets: (1) the public *i-LID MCTS* dataset [44] which contains 274 group images for 64 groups; (2) our own constructed *DukeMTMC Group* dataset which includes 177 group image pairs extracted from a 8-camera-view DukeMTMC dataset [21]; (3) our own constructed *Road Group* dataset which includes 162 group pairs taken from a two-camera crowd road scene[1].

When constructing our own datasets, we use [25] to automatically detect groups, and randomly select groups with different sizes & variations as the target groups in our dataset. Moreover, we define two cross-view groups as the same group when they have more than 60% members in common.

Some example groups for the three datasets are shown in Fig. 2b. Note that the *i-LID MCTS* dataset suffers from low image quality & large illumination, while the *DukeMTMC Group* and *Road Group* datasets include severe object occlusion and large layout & group member variation. To have a fair comparison with other methods, we follow the evaluation protocol in [46] and evaluate the Cumulated Matching Characteristic (CMC) results [14] on half of each dataset.

6.1 Results for Multi-Grain Re-ID Framework

In order to evaluate the effectiveness of our multi-grain group Re-ID framework, we compare eight methods: (1) Only using global features [8] of the entire group for Re-ID (*Global*);

---

[1]Available at http://min.sjtu.edu.cn/lwydemo/GroupReID.
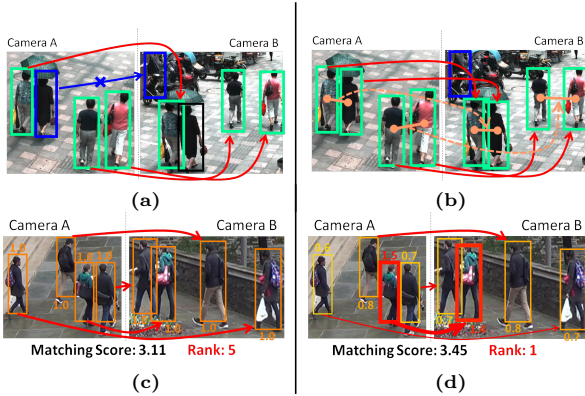
Figure 6: Matching results by: (a) only using information of individuals; (b) using multi-grain information; (c) setting equal importance weights for all individuals/subgroups; (d) using our importance evaluation process to obtain importance weights. The red and blue links indicate correct and wrong matches. (Note: to make the illustration not too busy, we only display the matching results between individuals.)

Table 1: Group Re-ID results on Road Group dataset

Rank	1	5	10	15	20	30
Global	15.8	31.6	43.0	48.6	54.8	61.7
Fine	62.0	82.2	89.6	95.1	96.5	97.3
Fine+Medium	66.7	87.2	93.3	96.0	96.8	97.3
Fine+Medium+Coarse	71.1	89.4	94.1	97.0	97.3	97.5
Equal-weight	55.8	78.0	88.1	92.1	93.6	97.8
Proposed-no spatial	69.6	88.6	94.0	96.2	96.5	97.4
Proposed-auto	72.3	90.6	94.1	97.1	97.5	98.0
Proposed-GT	**76.0**	**91.8**	**95.3**	**97.2**	**98.0**	**98.0**

(2) Only using features of individual people for group Re-ID (*Fine*); (3) Using features of individual people and two-people subgroups for Re-ID (*Fine+Medium*); (4) Using features of individual, two-people, three-people subgroups for Re-ID (*Fine+Medium+Coarse*); (5) Using our multi-grain group Re-ID framework, but setting the importance weights for all people/people subgroups as 1 (*Equal-weight*); (6) Using our multi-grain framework, but deleting the spatial relation features in multi-grain representation (cf. Sec. 3, *Proposed-no spatial*); (7) Using our multi-grain framework, but using the groundtruth pedestrian detection results (*Proposed-GT*); (8) Using our multi-grain framework and using automatic pedestrian detection method with 85% $F1$ accuracy [18] to identify individual people in groups (*Proposed-auto*).

Table 1 shows the CMC results of group Re-ID on Road Group dataset, which measures the correct match rates in different Re-ID rank ranges. Fig. 6 shows some group-wise matching results under different methods. We observe that:

(1) The *Global* method achieves poor results. This implies that simply using global features cannot effectively handle the challenges in group Re-ID. Comparatively, the *Fine* method has obviously better performance by extracting and matching individual people to handle the challenges from group

Table 2: Matching and Re-ID results on Road Group dataset (*MA*: matching accuracy between individuals; *R*-1*(Gr)*: Rank-1 CMC for group Re-ID; *R*-1*(In)*: Rank-1 CMC for individual Re-ID)

Method	Single	No inter	No dis	Hyp-E	**Proposed-Auto**
MA(%)	83.1	87.0	88.2	86.4	**88.2**
R-1(Gr)	62.0	70.1	65.8	55.1	**72.3**
R-1(In)	60.1	68.9	63.4	53.3	**71.4**

dynamics. However, its performance is still restrained by the interferences of pedestrian misdetection or mismatching (cf. Fig. 6a). These interferences are properly reduced by the *Fine+Medium* and *Fine+Medium+Coarse* methods, which include sub-group level information to achieve better results. Finally, our proposed framework (*Proposed-GT* and *Proposed-auto*), which includes four-grain information, can achieve the best performance.

(2) The *Equal-weight* method has obviously lower results than the methods with importance weights (*Proposed-GT* and *Proposed-auto*). This clearly indicates that: a) assigning importance weights to different people/people subgroups is significant in guaranteeing group Re-ID performances; b) Our proposed importance evaluation process is effective in finding proper importance weights, such that reliable and discriminative people/people subgroups are highlighted to obtain satisfactory results (cf. Fig. 6c-6d).

(3) The *Proposed-no spatial* method achieves relatively satisfactory results. This indicates that even when spatial relation features are not included, our approach can still obtain reliable performances by properly leveraging multi-grain information and importance weights.

(4) The *Proposed-auto* method has similar results to the *Proposed-GT* method. This also indicates that our multi-grain group Re-ID framework has the ability to properly handle the interferences from pedestrian misdetection. For example, in Fig. 6a, the group in camera A includes a false alarm detection (the blue rectangle) which is easily confused with the blue circled object in camera B. However, by integrating multi-grain information in our approach, we can successfully avoid this mismatch by considering the subgroup correlation in higher-level granularities (cf. Fig. 6b).

6.2 Results for Multi-Order Matching

We further evaluate the performance of our multi-order matching process by comparing five methods: (1) Only using a single level (i.e., the first order level in Fig. 4) to perform matching (*Single*); (2) Delete the effect of the inter-order potential (i.e., setting all $\mathcal{M}_{k_1 k_2}(\mathcal{C})$ to be 0 in Eq. 7, *No inter*); (3) Do not use the unmatched term (i.e., the second term in Eq. 16) when calculating matching scores (*No dis*); (4) Using a hyper-edge matching method [13] to perform matching, which integrates multi-grain information by constructing a mixed-type attribute (*Hyp-E*); (5) Our matching process (*Proposed-auto*).

Table 2 shows the individual-people matching accuracy (*MA*) [36] and Rank-1 CMC score for group Re-ID (*R*-1*(Gr)*)

Table 3: CMC results of Group Re-ID on different datasets

Rank Method	i-LIDS MCTS					DukeMTMC Group					Road Group				
	1	5	10	15	20	1	5	10	15	20	1	5	10	15	20
Saliency [41]	26.1	48.5	67.5	80.3	89.9	13.9	33.3	51.5	59.8	66.3	48.6	73.6	82.2	86.2	90.1
Mirror+KMFA [8]	28.3	58.4	69.8	80.5	90.6	11.0	31.5	49.7	62.9	70.8	25.7	49.9	59.5	66.9	72.1
CRRRO-BRO [44]	23.3	54.0	69.8	76.7	82.7	9.9	26.1	40.2	54.2	64.9	17.8	34.6	48.1	57.5	62.2
Covariance [5]	26.5	52.5	66.0	80.0	90.9	21.3	43.6	60.4	70.3	78.2	38.0	61.0	73.1	79.0	82.5
PREF [15]	30.6	55.3	67.0	82.0	92.6	22.3	44.3	58.5	67.4	74.4	43.0	68.7	77.9	82.2	85.2
BSC+CM [46]	32.0	59.1	72.3	82.4	93.1	23.1	44.3	56.4	64.3	70.4	58.6	80.6	87.4	90.4	92.1
Proposed-auto	**37.9**	**64.5**	**79.4**	**91.5**	**93.8**	**47.4**	**68.1**	**77.3**	**83.6**	**87.4**	**72.3**	**90.6**	**94.1**	**97.1**	**97.5**

of different matching methods on the Road Group dataset. In order to further demonstrate the effectiveness of our multi-order matching process, we also compare the Rank-1 scores for individual Re-ID (R-$1(In)$) by applying different matching methods to perform single person Re-ID. We observe that:

(1) Compared with the *Single* method, the *Proposed* method has improved matching accuracy and group Re-ID score by introducing multi-grain information to reduce the interferences of misdetection & mismatch. Besides, the *Proposed* method also has better individual Re-ID results than the *Single* method, this also demonstrates that introducing multi-grain information can facilitate single person Re-ID. (2) The *Proposed* method has better results than the *No inter* method. This indicates that the inter-order potential is useful in properly combining multi-order information. (3) The *Proposed* method has better group Re-ID results (R-$1(Gr)$) than the *No dis* method. This further demonstrates the importance of including the information of unmatched objects when calculating matching scores (cf. Eq. 16). (4) The *Proposed* method also has better matching accuracy & Re-ID scores than the *Hyp-E* method. This demonstrates that our multi-order matching process can make better use of the multi-grain information in groups during matching.

6.3 Comparison with the State-of-the-art Methods

Table 3 compares our approach with the state-of-the-art group re-identification methods on different datasets: *CRRRO-BRO* [44], *Covariance* [5], *PREF* [15], *BSC+CM* [46]. To further demonstrate the effectiveness of our approach, we also include the results of two state-of-the-art methods designed for single person Re-ID, which utilize patch saliency or a KMFA(R_{χ^2}) distance metric to calculate image-wise similarity (*Saliency* [41] and *Mirror+KMFA* [8]).

From Table 3, we can observe that: (1) Our approach has better results than the existing group Re-ID methods. This demonstrates the effectiveness of our approach. (2) The group Re-ID methods using global features (*CRRRO-BRO* [44], *Covariance* [5], *PREF* [15]) achieves less satisfactory results. This indicates that only using global features cannot properly handle the challenges in group Re-ID. (3) Although the *BSC+CM* method achieves improved results than the global feature-based methods by introducing fine-grain objects (i.e., patches) to handle group dynamics, its performance is still obviously lower than our approach. This implies the usefulness of including multi-grain information. (4) Our approach

also achieves better results than the methods for single person Re-ID (*Saliency, Mirror+KMFA*). This indicates that single-person Re-ID methods have limitations in handling the challenges in group Re-ID, while our approach can better address these challenges by capturing the characteristics of groups. (5) The improvement of our approach is more obvious on datasets with large group layout and group member changes (*DukeMTMC Group & Road Group*), this also demonstrates the effectiveness of our approach in handling group layout and member variations.

6.4 Computation Complexity

Finally, Table 4 shows the running time of our group Re-ID approach on different datasets (excluding object detection & feature extraction).

We list two time complexity values: (1) the running time for the entire process (*All image pairs*), and (2) the average running time for computing the similarity of a single group image pair (*Per image pair*). Table 4 shows that the complexity of our approach is acceptable.

Table 4: Running time on three datasets (evaluated on a PC with 2-core CPU and 8G RAM)

Datasets	i-LIDS MCTS	DukeMTMC Group	Road Group
All image pairs	1.1 min	18.9 min	11.5 min
Per image pair	0.06 sec	0.14 sec	0.10 sec

7 CONCLUSION

This paper introduces a novel approach to address the seldom-studied problem of group re-identification. Our approach consists of two key ingredients: 1) a multi-grain group Re-ID process which derives feature representations for multi-grain objects and iteratively evaluates their importance to handle interferences from group dynamics; 2) a multi-order matching process which integrates multi-grain information to obtain more reliable group matching results. Experiments demonstrate the effectiveness of our approach. We also release two datasets involving the realistic challenges in group Re-ID.

8 ACKNOWLEDGEMENTS

REFERENCES

[1] Shayan Modiri Assari, Haroon Idrees, and Mubarak Shah. 2016. Human re-identification in crowd videos using personal, social and environmental constraints. In *ECCV*. 119–136.

[2] Sławomir Bak and Peter Carr. 2017. One-Shot metric learning for person re-identification. In *CVPR*.

[3] Alina Bialkowski, Patrick J Lucey, Xinyu Wei, and Sridha Sridharan. 2013. Person re-identification using group information. In *DICTA*. 1–6.

[4] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: Identifying density-based local outliers. In *SIGMOD*. 93–104.

[5] Yinghao Cai, Valtteri Takala, and Matti Pietikäinen. 2010. Matching groups of people by covariance descriptor. In *ICPR*. 2744–2747.

[6] Jiaxin Chen, Zhaoxiang Zhang, and Yunhong Wang. 2015. Relevance metric learning for person re-identification by exploiting listwise similarities. *IEEE TIP* (2015), 4741–4755.

[7] Shi-Zhe Chen, Chun-Chao Guo, and Jian-Huang Lai. 2016. Deep ranking for person re-identification via joint representation learning. *IEEE TIP* (2016), 2353–2367.

[8] Ying-Cong Chen, Wei-Shi Zheng, and Jianhuang Lai. 2015. Mirror representation for modeling view-specific transform in person re-identification. In *IJCAI*. 3402–3408.

[9] Minsu Cho, Karteek Alahari, and Jean Ponce. 2013. Learning graphs to match. In *ICCV*. 25–32.

[10] Abir Das, Anirban Chakraborty, and Amit K Roy-Chowdhury. 2014. Consistent re-identification in a camera network. In *ECCV*. 330–345.

[11] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao. 2015. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition* (2015), 2993–3003.

[12] Haruyuki Iwama, Yasushi Makihara, and Yasushi Yagi. 2012. Group context-aware person identification in video sequences. *IPSJ Trans. Computer Vision and Applications* (2012), 87–99.

[13] Jungmin Lee, Minsu Cho, and Kyoung Mu Lee. 2011. Hyper-graph matching via reweighted random walks. In *CVPR*. 1633–1640.

[14] Weiyao Lin, Yang Shen, Junchi Yan, Mingliang Xu, Jianxin Wu, Jingdong Wang, and Ke Lu. 2017. Learning correspondence structures for person re-identification. *IEEE TIP* (2017), 2438–2453.

[15] Giuseppe Lisanti, Niki Martinel, Alberto Del Bimbo, and Gian Luca Foresti. 2017. Group re-identification via unsupervised transfer of sparse features encoding. In *ICCV*.

[16] Jiawei Liu, Zheng-Jun Zha, Xuejin Chen, Zilei Wang, and Yongdong Zhang. 2018. Dense 3D-Convolutional Neural Network for Person Re-Identification in Videos. *ACM Trans. Multimedia Computing, Communications and Applications* 14 (2018). Issue 4s.

[17] Jiawei Liu, Zheng-Jun Zha, Qi Tian, Dong Liu, Ting Yao, Qiang Ling, and Tao Mei. 2016. Multi-scale triplet cnn for person re-identification. In *ACM Multimedia*. 192–196.

[18] Lihang Liu, Weiyao Lin, Lisheng Wu, Yong Yu, and Michael Ying Yang. 2016. Unsupervised deep domain adaptation for pedestrian detection. In *ECCV Workshop*. 676–691.

[19] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. 2016. Large-scale vehicle re-identification in urban surveillance videos. In *ICME*. 1–6.

[20] Quynh Nguyen, Antoine Gautier, and Matthias Hein. 2015. A flexible tensor block coordinate ascent scheme for hypergraph matching. In *CVPR*. 5270–5278.

[21] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV workshop*. 17–35.

[22] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 2000. The earth mover's distance as a metric for image retrieval. *IJCV* (2000), 99–121.

[23] Chen Shen, Zhongming Jin, Yiru Zhao, Zhihang Fu, Rongxin Jiang, Yaowu Chen, and Xian-Sheng Hua. 2017. Deep siamese network with multi-level similarity perception for person re-identificatio. In *ACM Multimedia*. 1942–1950.

[24] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. 2017. Learning deep neural networks for vehicle Re-ID with visual-spatio-temporal path proposals. In *ICCV*.

[25] Francesco Solera, Simone Calderara, and Rita CucchiaraIwama. 2016. Socially constrained structural learning for groups detection in crowd. *IEEE TPAMI* (2016), 87–99.

[26] Shoubiao Tan, Feng Zheng, Li Liu, Jungong Han, and Ling Shao. 2016. Dense invariant feature based support vector ranking for cross-Camera person re-identification. *IEEE TCSVT* (2016), 356–363.

[27] Norimichi Ukita, Yusuke Moriguchi, and Norihiro Hagita. 2016. People re-identification across non-overlapping cameras using group features. *Computer Vision Image Understanding* (2016), 228–236.

[28] Rahul Rama Varior, Gang Wang, Jiwen Lu, and Ting Liu. 2016. Learning invariant color features for person reidentification. *IEEE TIP* (2016), 3395–3410.

[29] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. 2016. Joint learning of single-image and cross-image representations for person re-identification. In *CVPR*. 1288–1296.

[30] Hanxiao Wang, Shaogang Gong, and Tao Xiang. 2014. Unsupervised learning of generative topic saliency for person re-identification. In *BMVC*.

[31] Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, and Qi Tian. 2017. Global-local-alignment descriptor for pedestrian retrieval. In *ACM Multimedia*. 420–428.

[32] Jimin Xiao, Yanchun Xie, Tammam Tillo, Kaizhu Huang, Yunchao Wei, and Jiashi Feng. 2017. IAN: The individual aggregation network for person search. *CoRR abs/1705.05552* (2017).

[33] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. 2016. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*. 1249–1258.

[34] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. 2017. End-to-end deep learning for person search. In *CVPR*.

[35] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. 2014. Person re-identification using kernel-based metric learning methods. In *ECCV*. 1–16.

[36] Junchi Yan, Changsheng Li, Yin Li, and Guitao Cao. 2017. Adaptive discrete hypergraph matching. *IEEE Trans. Cybernetics* (2017), 765–779.

[37] Mang Ye, Chao Liang, Zheng Wang, Qingming Leng, and Jun Chen. 2015. Ranking optimization for person re-identification via similarity and dissimilarity. In *ACM Multimedia*. 1239–1242.

[38] Li Zhang, Tao Xiang, and Shaogang Gong. 2016. Learning a discriminative null space for person re-identification. In *CVPR*. 1239–1248.

[39] Ziming Zhang and Venkatesh Saligrama. 2017. Prism: Person reidentification via structured matching. *IEEE TCSVT* (2017), 499–512.

[40] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. 2013. Unsupervised salience learning for person re-identification. In *CVPR*. 3586–3593.

[41] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. 2017. Person re-identification by salience learning. *IEEE* (2017), 356–370.

[42] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. 2016. Mars: A video benchmark for large-scale person re-identification. In *ECCV*. 868–884.

[43] Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, and Qi Tian. 2017. Person re-identification in the wild. In *CVPR*.

[44] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. 2009. Associating groups of people. In *BMVC*.

[45] Zhen Zhou, Yan Huang, Wei Wang, Liang Wang, and Tieniu Tan. 2017. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification. In *CVPR*. 6776–6785.

[46] Feng Zhu, Qi Chu, and Nenghai Yu. 2016. Consistent matching based on boosted salience channels for group re-identification. In *ICIP*. 4279–4283.