From Groups to Co-traveler Sets: Pair Matching based Person Re-identification Framework

Min Cao[1] Chen Chen[1*] Xiyuan Hu[1] Silong Peng[12]
[1]Institute of Automation Chinese Academy of Sciences Beijing, China
[2]Beijing Visytem Co. Ltd.
{caomin2014,chen.chen,xiyuan.hu,silong.peng}@ia.ac.cn

Abstract

In video surveillance, group refers to a set of people with similar velocity and close proximity. Group members can provide visual clues for person re-identification. In this paper, we discuss the essentials of group-based person re-identification and relax the group definition towards a concept of "co-traveler set", keeping constraints on velocity differences while loosening the distance constraint. Accordingly we propose a pair matching scheme to measure the distance between co-traveler sets, which tackles the problems caused by dynamic change of group across camera views. The final individual matching score is weighted by the obtained distance measurements between co-traveler sets. A proof of concept shows the rationality of introducing the concept of co-traveler relation into person re-id. Experiments were conducted on four different datasets. Our co-traveler set based framework shows promising improvement compared with the group-based methods and the individual-based methods.

Figure 1. An intuitive example of re-identifying query person with the help of the group context. (a)-(d) show the query person and the candidates with similar appearance; (e)-(h) show the corresponding group images with different appearance.

1. Introduction

Person re-identification (re-id) [29] aims to match people across non-overlapping camera views and has attracted increasing attention in recent years, because of the application of intelligent video surveillance system.

Researchers mainly focus on two steps for person re-id: developing robust feature representations [11, 36, 21, 27] and learning a proper distance/similarity metric with these features [19, 20, 25], mostly based on only individual pedestrian. However, the appearance of people often undergoes dramatic change across different camera views due to changes in illuminations, human poses, viewpoints and backgrounds. Furthermore, the surveillance network captures tens of thousands of people every day and many different people are similar in appearance. Hence, if the features and corresponding metrics based on only individual appearance are utilized for re-identifying person, the performance enhancement is limited.

In order to improve the accuracy of person re-id, an intuitive solution is to utilize the context information, for instance, other pedestrians in the surrounding area. Empirical study suggests that about 50-70% of people (depending on the environment) are walking with other people [24]. If cameras are not geographically far apart, the same group structure will appear again in neighboring cameras, the group context can provide clues for person re-id. Figure 1 illustrates one intuitive example. The top row shows the query person in (a) and the candidates in (b-d), and it is difficult to differentiate the true candidate in (b) from others since they look much similar in appearance. When the group information is introduced as (e-h) showed in bottom row, it is relatively easy to identify the correspondence of query person in (f) rather than the rest in (g-h) due to the huge appearance difference provided by the group members.

*Corresponding author.

Several group-based person re-id methods [3, 18, 37, 28, 1] have been proposed recently. These methods have shown decent results in terms of accuracies. In these methods, the group is regarded as a set of people with close proximity and similar velocity. The essential effect brought by groups for person re-id includes the following aspects: 1) groups establish the specific neighbouring relations between a number of query persons with their surrounding pedestrians; 2) these relations introduce extra useful visual information; 3) these relations keep stable across the camera views. The success of utilizing group information leads naturally to an inference: if more query persons are associated with such stable relations across camera views, better person re-id accuracy should be achieved.

In this paper, we discuss the definition of group and extend groups to stable but more relaxed relations among people, i.e. the concept of *co-traveler set*. A target person's co-traveler set is defined as a set of people who walk with similar velocity and keep a certain (relatively large) distance with the target person. Compared with the traditional group definition, due to the loose constraint of the walking distance between pedestrians, more stable relations among pedestrians are established and keep consistent across camera views.

In addition, with the stable relations among pedestrians across camera views in the format of groups or co-traveler sets, the feature representations of these relations should also ensure the stability across views. Many existing group-based methods represent the group features as the relative positions of group members and group size. However, the group is highly non-rigid and the relative positions of group members are likely to change across views, and sometimes the group size may also change. In order to reduce the instability of representations of stable relations, we propose a co-traveler set based pair matching scheme. The pairs are constructed by combing the query person with each of his/her co-travelers, on the basis, the pair matching strategy to some extend is immune to the changes of co-traveler set or group size and relative position of co-travelers or group members.

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 discusses the definition of groups and the proposed *co-traveler set*. Section 4 describes co-traveler set based person re-id framework. Section 5 reports the experimental results, and Section 6 concludes this paper.

2. Related Works

The existing person re-id methods can be divided into individual-based methods and group-based methods. We briefly review them as follows.

2.1. Individual-based Person Re-identification

For re-identifying people across different camera views, most works [34, 39, 25, 32] have focused on individual-based techniques, which usually consist of two steps: feature design and metric learning.

In the feature design stage, several works [11, 36, 21, 27] are devoted to extract features of individuals accurately and discriminatingly. For example, L. Bazzani *et al*. [11] proposed symmetry-driven accumulation of local features including the weighted color histograms, maximally stable color regions and recurrent high-structured patches. T. Matsukawa *et al*. [23] considered that mean information of pixel features is the major discriminative information of person images and proposed Gaussian Of Gaussian (GOG) descriptor by adding the mean of image pixel features into the covariance descriptor. Compared with these hand-crafted features, deep CNN as one of the most popular methods in computer vision field, is powerful to learn the robust and discriminative features. For example, D. Cheng *et al*. [7] designed a multi-channel parts-based CNN model to jointly learn both the global full-body and local body-parts features of people. However, CNN requires a mass of labeled samples in the training phase. To this end, the researchers trained a CNN with data from all the domains [31]. Compared with these single-shot based methods that extract individual features from a single static image, multi-shot based methods [14, 10, 16] incorporate the motion information of people and extract additional features from multi-frame images, which has received much attention in recent years.

In the metric learning stage [38, 25, 19, 20], most researches have focus on supervised settings. Their works learned either a distance metric or a discriminative subspace aiming to obtain a larger between-class distance than within-class distance. Zheng *et al*. [38] introduced a Probabilistic Relative Distance Comparison (PRDC) model that maximises the probability of a pair of true match having a smaller distance than that of a wrong match pair. Liao *et al*. [20] considered dimension reduction before metric learning and learned a discriminant low dimensional subspace by cross-view quadratic discriminant analysis.

Although several profound progresses have been made in individual-based person re-id, it remains a challenging task. One major reason is that the appearance and motion of a person can be far different, which results from different camera views and pose variations of a person. As a compensation, some of the person re-id methods have been built on the idea of group.

2.2. Group-based Person Re-identification

Recently, there is a growing interest in using group information to improve the performance of person re-id [3, 18, 37, 28, 1]. The frameworks of these methods are similar. Firstly, groups are obtained either manually or by employing an automatic technique. Then, the information of groups is described by the extracted group features. Finally, the group features, together with the individual features, are incorporated in the person re-id task. More specifically, Zheng *et al*. [37] defined two group features named Center Rectangular Ring Ratio-Occurrence Descriptor (CRRRO) and Block based Ratio-Occurrence Descriptor (BRO) to be extracted. The former treated with changes in the relative positions of people in a group and the latter addressed variations in illumination and viewpoint across camera views. In the work of Cai *et al*. [3], the feature of the group images are represented by the covariance descriptor, which captures both appearance and statistical properties of image regions. More recently, Li *et al*. [18] proposed to extract the group features based on both geometry and visual information of a subject's partners. Compared with manually selecting group images in the works [37, 3], Li *et al*. [18] used the Affinity Propagation (AP) clustering algorithm [12] to discover the group members. In addition, Chen et.al [4] used outstanding contextual pedestrians to help reidentify the query person. By utilizing group information, these methods have shown decent results in terms of accuracies.

What's more, researchers often utilize the group information for human tracking. Cai *et al.* [2] proposed a relaxed definition of group named neighboring set for tracking, which is closely related to the proposed *co-traveler set*. However, the definition of neighboring set does not constrain the velocity between people, which is the essential factor for group-based person re-id method (refer to section 3) and is considered in the co-traveler set. In [6], elementary groups in which two targets are contained is proposed and its information is explored to improve multi-target tracking. Compared with using the information of complete group, the elementary group is more flexible to the dynamic change of group in the real world. Similarly, we introduce the pair structure into co-traveler set (or group) matching and propose a co-traveler set based pair matching strategy.

3. From Groups to Co-traveler Sets

3.1. Essentials of Group-based Methods

A set of people who have similar speed and close proximity are regarded as a group [9]. Groups, commonly appeared in video surveillance, play an important role for various research goals including crowd understanding [26], people tracking [33] and person re-id [28]. In these applications, a group is defined and detected mainly by four trajectory factors: the velocity difference, the walking direction difference, the inter-people distance, and the temporal overlap. For group-based person re-id, only some factors are essential to enhance the performance.

A pedestrian's speed and walking direction are usually assumed as constant across two adjacent cameras. Therefore, if a set of people are associated as a group due to their similar velocity and walking direction in one camera, the associations will be stably sustained in the adjacent camera view. The stability of these associations ensures these extra visual cues useful and reliable. Compared with the individual-based person re-id methods, the essentials of group-based methods can be summaries as:

1) a number of query persons have group associations;
2) these associations bring extra useful information;
3) these associations remain stable across camera views.

Checking with the four trajectory factors with the summarized essentials, we can conclude that for the purpose of remaining group stability across views, the constraints on velocity and walking direction must be kept to guarantee the reliable associations; while the requirement for close proximity is not strictly necessary. If the walking distance constraint is loosened, more query persons will be associated with surrounding pedestrians and the accuracy of person re-id can be improved.

3.2. Relaxation towards Co-traveler Sets

We introduce a concept of *co-traveler set* into person re-id in this paper. Let the trajectory of a person p in a video scene be a set of tuples $(\boldsymbol{s_t}, \boldsymbol{v_t}, t)$ at frame t, where $\boldsymbol{s_t}$ is the position vector of person p and $\boldsymbol{v_t}$ is the velocity vector at frame t, $t \in [T_1, T_2]$, while T_1 and T_2 denote the starting frame and ending frame of person p appeared in the video. The average velocity $\boldsymbol{v}$ of person p during the period $[T_1, T_2]$ can be estimated accordingly.

For two persons p_i and p_j appeared successively in the video scene, without loss of generality, we assumed $T_1^i \leq T_1^j, T_2^i \leq T_2^j$, the pairwise measuring features are defined as follows:

- the velocity difference:

$$\nu_{ij} = \|\boldsymbol{v_i} - \boldsymbol{v_j}\| \ ,$$

- the angle of walking direction difference:

$$\theta_{ij} = arccos\,(\frac{\boldsymbol{v_i} \cdot \boldsymbol{v_j}}{\|\boldsymbol{v_i}\|\,\|\boldsymbol{v_j}\|}) \ ,$$

- the distance between p_i and p_j:

$$\rho_{ij} = \frac{\|\boldsymbol{v_i}\| \cdot |T_1^i - T_1^j| + \|\boldsymbol{v_j}\| \cdot |T_2^i - T_2^j|}{2} \ .$$

For each pair of individuals, we compute these three measuring features $(\nu_{ij}, \theta_{ij}, \rho_{ij})$ and set corresponding thresholds $(\tau_\nu, \tau_\theta, \tau_\rho)$ to evaluate the existence of stable relation between p_i and p_j. Person p_i and person p_j are defined

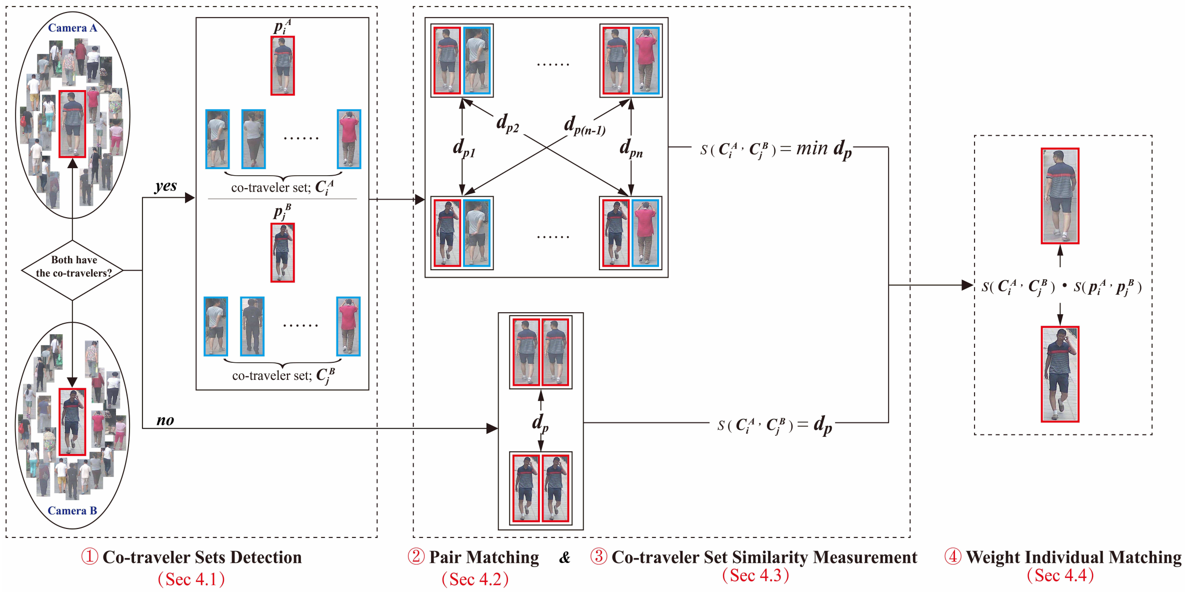

Figure 2. Overview of the proposed co-traveler set based framework for person re-id. Better viewed in colour.

as *co-travelers* if $\nu_{ij} < \tau_\nu$, $\theta_{ij} < \tau_\theta$ and $\rho_{ij} < \tau_\rho$. The *co-traveler set* of person p_i is defined as:

$$\mathcal{C}_i = \{p_j | \nu_{ij} < \tau_\nu, \theta_{ij} < \tau_\theta, \rho_{ij} < \tau_\rho, i \neq j\}.$$

The thresholds τ_v and τ_θ are set to small values, similarly with the traditional group definition, to ensure the established co-traveler relations stable. The larger threshold τ_ρ and the absence of the requirement for temporal overlap loosen the traditional group definition. Under such threshold settings, more query persons are associated with surrounding people and these associations remain stable across camera views by definition, and these stable relations can improve the accuracy of person re-id. Ideally, if the speed constraints are perfectly satisfied, the threshold τ_ρ can be set very large. In very rare cases in the real world scenario, few individual pedestrians might change velocity greatly even across adjacent camera views, a large distance threshold τ_ρ and strict constraints on velocity by thresholds τ_ν and τ_θ may lead to wrong match of co-traveler sets and corresponding pedestrians.

4. Person Re-identification Framework

The flowchart of our proposed person re-id framework is illustrated in Figure 2.

4.1. Co-traveler Sets Detection

Consider some video clips from two neighbouring cameras, based on camera calibration or other homography projection techiniques, the trajectory information of a person p, including the location vector $\boldsymbol{s_t}$ at frame t and time duration $[T_1, T_2]$, can be easily obtained by basic detection and tracking. The velocity vector $\boldsymbol{v_t}$ is estimated by $\dfrac{\boldsymbol{s_{t'}} - \boldsymbol{s_t}}{\Delta t}(\Delta t = t' - t)$ with smoothing by Savitzky-Golay filters. The average velocity $\boldsymbol{v}$ of person p is computed using the smoothed velocity $\boldsymbol{v_t}$.

With the relative trajectory information, given person p and the surrounding people, we compute features (ν, θ, ρ) and the corresponding thresholds $(\tau_\nu, \tau_\theta, \tau_\rho)$ to identify the co-travelers and the co-traveler set $\mathcal{C}$ according to the definition of Section 3.2.

4.2. Pair Matching

With co-traveler set detection, person p_i^A captured in camera A is associated with a co-traveler set $\mathcal{C}_i^A$, and person p_j^B in camera B with a co-set $\mathcal{C}_j^B$, we use $P_{im}^A = (p_i^A, p_m^A)$ to denote a pair of co-travelers, similarly, $P_{jn}^B = (p_j^B, p_n^B)$.

The distance between pair P_{im}^A and pair P_{jn}^B is computed based on the fusion of the individual distance metrics and we use the weighted sum to realize the fusion:

$$d(P_{im}^A, P_{jn}^B) = \alpha s(p_i^A, p_j^B) + (1 - \alpha) s(p_m^A, p_n^B), \quad (1)$$

where α denotes the weight, $0 \leq \alpha \leq 1$, and $s(\cdot, \cdot)$ is the distance metric of features extracted from individual pedestrians obtained by the individual-based method.

4.3. Co-traveler Set Distance Measurement

If a candidate in gallery is the true match of the query person in probe, their surrounding people, i.e. their co-traveler sets should also match and the distance measure between the two corresponding co-traveler sets should be relatively low. Similarly with the group-based methods, the proposed approach utilizes the distance measures of the co-traveler sets to improve the individual person re-identification.

According to the definition of co-traveler set, if $\mathcal{C}$ is person p's co-traveler set, each person in $\mathcal{C}$ and p are co-travelers. We can rewrite person p's co-traveler set $\mathcal{C}$ in the format of a set of pairs containing person p and each of all his/her co-travelers: $\mathcal{P} = \{(p, q) | q \in \mathcal{C}\}$. In the context of set matching, the distance value of two co-traveler sets $\mathcal{C}_i^A$ and $\mathcal{C}_j^B$ essentially equals to the distance between the pair sets $\mathcal{P}_i^A$ and $\mathcal{P}_j^B$:

$$s(\mathcal{C}_i^A, \mathcal{C}_j^B) = d(\mathcal{P}_i^A, \mathcal{P}_j^B) \tag{2}$$

where the distance $d(\mathcal{P}_i^A, \mathcal{P}_j^B)$ can be computed due to the set theory:

$$d(\mathcal{P}_i^A, \mathcal{P}_j^B) = \min_{n,m} \left\{ d(P_{im}^A, P_{jn}^B) \mid P_{im}^A \in \mathcal{P}_i^A, P_{jn}^B \in P_j^B \right\}, \tag{3}$$

as the minimum of scores corresponding to all possible pair matching which described in Section 4.2.

Compared with those group-based methods, we measure the distance between co-traveler sets using pair matching strategy and the set distance metric. The proposed method can perfectly avoid the low matching rate caused by members changing their relative spatial locations within groups/co-traveler sets and be robust against the variation in size of groups/co-traveler sets across camera views.

If person p_i^A and person p_j^B walk alone in camera A and B respectively, their co-traveler set $\mathcal{C}_i^A$ and $\mathcal{C}_j^B$ are empty by definition. In this case p_i and p_j form two virtual pairs by themselves respectively, *i.e.* $P_{ii}^A = (p_i^A, p_i^A)$ and $P_{jj}^B = (p_j^B, p_j^B)$. The pair distance is computed by $d(P_{ii}^A, P_{jj}^B)$ and the distance measurement between the two virtual sets according to Eq.3 actually degenerates to the distance value between the two individuals.

4.4. Weighted Individual Matching

This section describes the strategy of integrating the co-traveler set information into the individual query person re-identification.

The distance value $d(p_i^A, p_j^B)$ between the query person p_i^A in probe set and the candidate p_j^B in gallery set is defined as:

$$d(p_i^A, p_j^B) = \widetilde{s}(\mathcal{C}_i^A, \mathcal{C}_j^B) \cdot s(p_i^A, p_j^B), \tag{4}$$

where

$$\widetilde{s}(\mathcal{C}_i^A, \mathcal{C}_j^B) = \mathcal{N}(s(\mathcal{C}_i^A, \mathcal{C}_j^B)) \tag{5}$$

is the normalized distance measurement between the query person's co-traveler set $\mathcal{C}_i^A$ and the candidate's co-traveler set $\mathcal{C}_i^B$, $\mathcal{N}(\cdot)$ denotes the min-max normalization operator that linearly scales the distance values into the range $[0.1, 1]$, and $s(p_i^A, p_j^B)$ is the individual distance values obtained by the individual-based method.

In addition, the results of co-traveler set detection can also facilitate in improving person re-id accuracy. After co-traveler set detection, each person is assigned a label of "with co-travelers" or "with no co-travelers". For a query person p_i^A and a candidate p_j^B, 1) if both of them have the same label, i.e. p_i^A and p_j^B are both with co-travelers or both with no co-travelers; 2) they have different labels, i.e. one is with co-travelers and the other with no co-travelers. Obviously, p_i^A and p_j^B in the first case are more likely to be the same person than in the second case. Therefore, we add a penalty factor λ in Eq.4 to differentiate the two situations:

$$d(p_i^A, p_j^B) = \lambda \cdot \widetilde{s}(\mathcal{C}_i^A, \mathcal{C}_j^B) \cdot s(p_i^A, p_j^B) \tag{6}$$

where

$$\lambda = \begin{cases} 1 & p_i^A \ and \ p_j^B \ have \ the \ same \ label \\ C & otherwise \end{cases} \tag{7}$$

and C is a constant, $C > 1$. By introducing the penalty factor λ, a query person with co-travelers tend to match with candidates with co-travelers while those with no travelers tend to match with candidates walking alone.

5. Experiments

5.1. Datasets and Settings

We have conducted extensive experiments to evaluate the proposed co-traveler set based framework (CTS framework) on three video datasets: i-LIDS MCTS [37], NLPR_MCT [18] and PRID2011 [15] and a new dataset CYBJ-G[1].

i-LIDS MCTS [37] monitors an airport arrival hall in the busy time under a multi-camera CCTV network. Zheng *et al.* [37] extracted 65 groups including 274 group images with different sizes from the dataset. Most of the groups have 4 images, either from different camera views or from the same camera but captured at different locations and time. These group images were resized to 182×60 pixels. They are very challenging due to the severe occlusions and the variations of relative position of group memebers.

NLPR_MCT [18] contains threes datasets. We use the videos (resolution: 320×240, frames per-second: 20) produced by two outdoor cameras in dataset 1 and 2. 73 and 106 persons were captured by both cameras in dataset 1 and 2, respectively. The datasets provide the ground truth annotation including the bounding box for each person. The homography between the image plane and the ground surface has been estimated off-line in the experiments.

PRID2011 [15] includes person images recorded from two cameras and full surveillance videos. 385 and 749 persons were recorded in camera views, respectively. Following the

[1] `http://mda.ia.ac.cn/people/huxy/patent.htm`

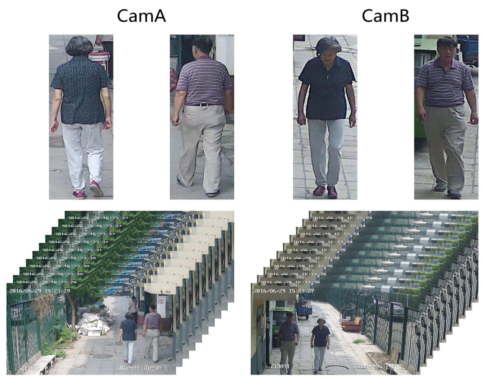

Figure 3. Examples of CYBJ-G dataset: the cropped images of pedestrians are shown in the top row and their corresponding video clips in the bottom row.

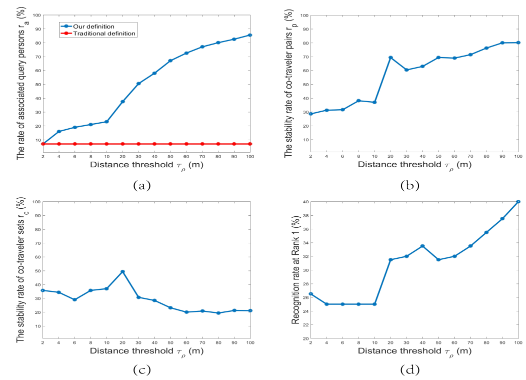

Figure 4. The change tendency of r_a, r_p, r_c and Rank 1 of our proposed method under different distance threshold.

protocol used in [30], we use 200 persons appeared in both cameras. People tracking and camera calibration have been conducted for co-traveler sets detection.

CYBJ-G consists 194 pedestrians captured by two surveillance cameras from the frontal view and the back view in a residential area. For each camera view, every pedestrian's data consists one cropped image of the person and the sequential images of the corresponding video clip. The sequential images contain 9-199 frames of video clips. The cropped images of persons have various sizes originally and have been resized to 384×144 pixels. For co-traveler sets detection, people tracking and camera calibration have been made. Some examples of the CYBJ-G dataset can be found in Figure 3.

Settings: We employ the Cumulated Matching Curve (CMC) [13] to evaluate and compare the performance of different methods. Most individual-based methods can be used as the baseline of the proposed CST framework. For the baseline methods that need training step, we randomly split dataset 10 times. A continuous time window randomly slides to collect all successive pedestrians and form the test set in order to keep the temporal sustainability of pedestrians in test set, while the rest pedestrians form the training set. In our experiments, the parameter α in Eq.1 are set to 0.5, and the penalty factor C in Eq.7 is set to $C = 2.0$.

5.2. A Proof of Concept: Co-traveler Set

In this section, we validate the effectiveness of proposed *co-traveler set* for person re-id. We first define three measurement as follows and use them to evaluate how many association co-traveler sets establish, how stable there associations are, and how they are linked with person re-id accuracy: the rate of associated query persons: $r_a = \frac{N_a}{N_*}$, the stability rate of co-traveler pairs: $r_p = \frac{N_p}{N_a}$, and the stability rate of co-traveler sets: $r_c = \frac{N_c}{N_a}$. N_a is the number of associated query persons, N_* is the number of query persons, N_p is the number of query persons keeping stable relations with at least one co-traveler across views and N_c is the number of query persons whose co-traveler set remains the same across views.

We conduct the experiments on PRID2011 dataset and use GOG [23] as baseline feature representation and Euclidean distance as the distance metric between features. Thresholds τ_v and τ_θ are set as a fixed value $\tau_v = 0.37m/s$ and $\tau_\theta = 19°$, the distance threshold τ_ρ is changed from the range of 2 to 100 (in meters). The velocity differences of each pedestrian across views are firstly computed: only about 8% people have 40% of the range of difference in speed, indicating that most people's speed are similar across views in PRID2011.

Figure 4 show the change tendency of r_a, r_p, r_c and Rank 1 of our proposed CTS framework under different distance threshold τ_ρ. We can see that **1)** with the increase of τ_ρ, both r_a and r_p almost continue to rise, showing that more query persons can be associated with such stable relations across camera views by the loose constraint of the walking distance between pedestrians; **2)** with the increase of τ_ρ, r_c increases at first and then decreases, while Rank 1 has increasing tendency, indicating that **a)** proposed *co-traveler set* is favourable for the performance of person re-id, **b)** our proposed method is robust to change of co-traveler set stability and it benefits from the pair strategy when measuring co-traveler sets. In the following experiment on PRID2011 dataset, we set $\tau_v = 0.37m/s$, $\tau_\theta = 19°$ and $\tau_\rho = 90m$.

5.3. Comparison with Group-based Methods

In this section, we compare our method with group-based methods including CRRRO [37] and SCGF[18] on i-LIDS MCTS [37] and NLPR_MCT [18] datasets. For NLPR_MCT, we set $\tau_v = 0.38m/s$, $\tau_\theta = 10°$ and $\tau_\rho = 10m$ for co-traveler sets detection. For the purpose of fair com-

Table 1. Comparison with group-based methods on i-LIDS MCTS, NLPR_MCT dataset 1 (d1) and dataset 2 (d2). Results are shown as matching rates (%) at $Rank = 1; 5; 10; 20$. The best and second results are shown in red and blue. Better viewed in colour.

Methods	i-LIDS MCTS			NLPR_MCT d1			NLPR_MCT d2		
Rank	r=1	r=5	r=10	r=1	r=5	r=10	r=1	r=5	r=10
SDALF [11]	16.0	31.0	38.4	22.0	53.0	74.0	37.0	62.0	73.0
CRRRO [37]	12.5	28.4	36.4	25.0	57.0	76.0	41.0	67.0	78.0
SCGF [18]	-	-	-	25.0	64.0	80.0	44.0	72.0	82.0
CTS	22.1	40.6	48.5	26.0	67.1	82.2	44.3	66.0	79.2

Table 2. Comparison with the classic and existing state-of-the-art methods on PRID2011 and CYBJ-G. The best and second best results (%) are respectively shown in red and blue. Better viewed in colour.

		PRID2011				CYBJ-G			
Methods	**Reference**	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20
SDALF	CVPR2010[11]	4.1	20.6	31.6	41.9	32.2	57.2	69.8	79.9
Salience	CVPR2013[36]	25.8	43.6	52.6	62.0	40.8	64.2	75.5	83.4
LOMO+XQDA	CVPR2015[20]	39.0	68.0	83.0	91.0	67.2	87.4	91.7	95.2
SCSP	CVPR2016[5]	12.7	32.7	51.0	66.0	21.7	39.1	50.0	67.4
DNS	CVPR2016[35]	38.4	66.6	79.0	92.1	58.2	84.2	91.5	94.1
GOG	CVPR2016[23]	59.2	83.5	92.2	96.8	76.5	93.3	97.2	98.3
DTDL	ICCV2015[17]	41.0	70.0	78.0	86.0	-	-	-	-
PaMM	CVPR2016[8]	45.0	72.0	85.0	92.5	-	-	-	-
STA	ICCV2015[22]	64.1	87.3	89.9	92.0	-	-	-	-
CTS+SCSP	Ours	16.3	39.8	53.2	68.8	33.8	54.8	63.3	74.7
CTS+DNS	Ours	53.9	83.5	92.6	98.0	81.7	91.5	92.4	95.1
CTS+GOG	Ours	75.8	92.5	96.2	98.6	91.9	98.0	98.7	99.3

parison, SDALF [11] is adopted as the baseline method for all experiments in this section.

The comparison results are shown in Table 1, as we can see, most group-based methods improved the performance of baseline method, which shows that the group context is useful to performance enhancement of person re-id.

i-LIDS MCTS Our method is superior to CRRRO on this dataset, in which the images are all group images and most people in the group images changed the relative spatial location with partners across views. CRRRO is proposed to solve this problems. However, the pair matching strategy in our method is immune to the change of relative positions within the co-traveler set.

NLPR_MCT dataset 1 We detected 61 people associated with surrounding people and other group-based methods [11, 18] detected 18 group members in probe set. The results show that our proposed method outperforms other group-based methods.

NLPR_MCT dataset 2 Our proposed method gets the best result at Rank 1. 46 query persons are associated with other people according to our co-traveler set detection and 35 according to other group-based methods of detection [11, 18]. More query persons associated with surrounding people are introduced into the process of person re-id, it should bring great performance enhancement. It does not improve very significantly since most of the detected co-travelers are similar in appearance, the information of these co-travelers is not helpful for re-identifying people.

5.4. Comparison with the State-of-the-art Methods

In Table 2, we reported the comparison of our proposed method with the classic and existing state-of-the-art individual-based person re-id methods including SDALF [11], Salience [36], LOMO+XQDA [20], SCSP [5], DNS [35] and GOG [23] which are based on a single static image, as well as DTDL [17], PaMM [8] and STA [22] based on multi-frame images. The experiments were conducted on PRID2011 and CYBJ-G datasets. For CYBJ-G , we set $\tau_v = 0.5m/s$, $\tau_\theta = 20^\circ$ and $\tau_\rho = 32m$ for co-traveler sets detection. In our proposed framework, all individual-based methods including both single-shot and multi-shot based methods can be used as the baseline method. We choose several single-shot methods as the baseline CTS framework including SCSP, DNS and GOG.

From Table 2, we can see that **1)** the performance enhancement (for all ranks $r = 1, 5, 10, 20$) is achieved by applying the proposed framework on the baseline methods, and our proposed framework reaches best results with the baseline method of GOG and outperforms all the other state-of-art methods at all rank level; **2)** the baseline methods with high rates usually obtain larger increases; **3)** in PRID2011, Rank 1 results increase by $3.6\% \sim 16.2\%$. In

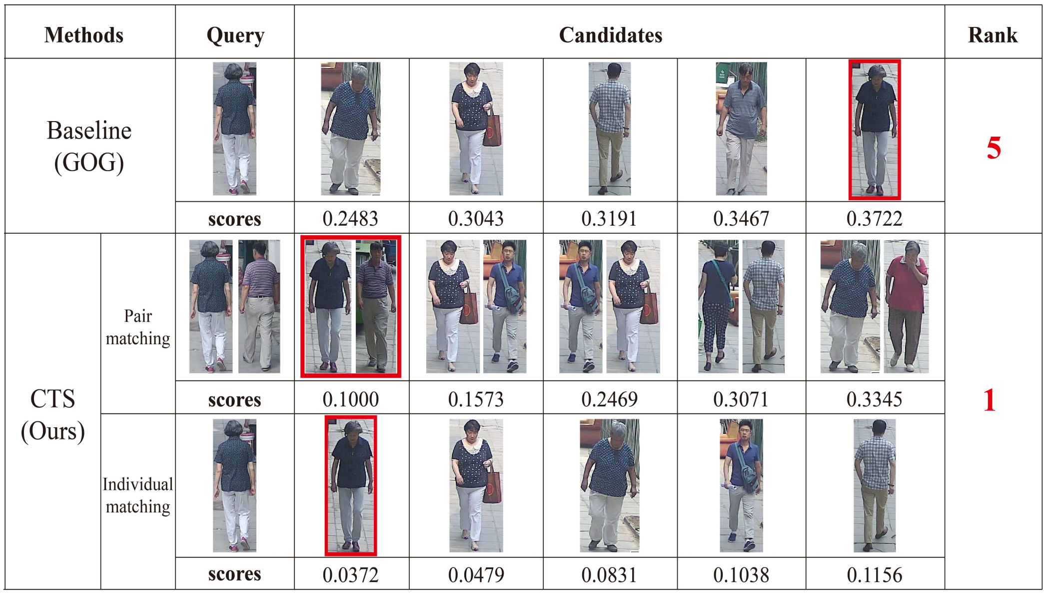

Figure 5. One example of re-id results from baseline method GOG and our proposed method. The matching results and ranks of co-traveler set based pair matching (Pair matching for short in figure) and weighted individual matching (Individual matching for short in figure) are displayed. The scores below the candidate are matching distances between query person (pair) and that candidate. We display the image of the top five candidate people and pairs in each grid. The ground truth matching is labeled by red boxes.

Table 3. The statistics of co-traveler sets on PRID2011 and CYBJ-G datasets in probe set. r_a, r_p and r_c are the rate of associated query persons, the stability rate of co-traveler pairs and the stability rate of co-traveler sets (%), respectively.

Datasets	r_a	r_p	r_c
PRID2011	82	63	22
CYBJ-G	100	91	94

CYBJ-G, Rank 1 results increase by $12.1\% \sim 23.5\%$ that is larger than the increase in PRID2011, because of more stable pairs in CYBJ-G. The statistics of co-traveler sets detected by our method on two datasets are list in Table 3.

An example of person re-id in CYBJ-G is shown is Figure 5. Using the baseline method, the true candidate is ranked the 5th among people in gallery. Applying the the co-traveler set based pair matching step and weighted individual matching step of CTS framework, the pair matching leads to the correct pair candidate and the true individual candidate moves up to top 1.

6. Conclusion

In this paper, we proposed a novel co-traveler set based pair matching framework for person re-id problem. It significantly differs from other existing methods in that 1) it is based on greatly relaxed groups, i.e. *co-traveler sets*, which allow more query person are associated with their surrounding pedestrians and keep the associations stable across camera views 2) it utilizes an effective pair matching mechanism to pass the co-traveler set distance measurement onto the final individual distance measurement in the format of normalized weights, which is robust to the changes in relative positions within co-traveler set and in size variation of co-traveler sets. The proposed framework can take most individual-based person re-id methods as baseline method, and experiments show that the proposed framework can dramatically increase the person re-id accuracy compared to various individual-based state-of-art methods on different datasets.

Acknowledgment

This work is supported by the National Key R&D Program of China under Grant 2017YFC0803505, China Postdoctoral Science Foundation, and National Natural Science Foundation of China with Grant No.61571438.

References

[1] A. Bialkowski, P. Lucey, X. Wei, and S. Sridharan. Person re-identification using group information. In *International Conference on Digital Image Computing: Techniques and Applications*, pages 1–6, 2013.

[2] Y. Cai and G. Medioni. Exploring context information for inter-camera multiple target tracking. In *Applications of Computer Vision*, pages 761–768, 2014.

[3] Y. Cai, V. Takala, and M. Pietik?inen. Matching groups of people by covariance descriptor. In *International Conference on Pattern Recognition*, pages 2744–2747, 2010.

[4] C. Chen, M. Cao, X. Hu, and S. Peng. Key person aided re-identification in partially ordered pedestrian set. In *Conference the British Machine Vision Conference*, 2017.

[5] D. Chen, Z. Yuan, B. Chen, and N. Zheng. Similarity learning with spatial constraints for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1268–1277, 2016.

[6] X. Chen, Z. Qin, L. An, and B. Bhanu. An online learned elementary grouping model for multi-target tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1242–1249, 2014.

[7] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1335–1344, 2016.

[8] Y. J. Cho and K. J. Yoon. Improving person re-identification via pose-aware multi-shot matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1354–1362, 2016.

[9] R. T. Collins, W. Ge, and R. B. Ruback. Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 34(5):1003–1016, 2012.

[10] D. N. T. Cong, C. Achard, L. Khoudour, and L. Douadi. Video sequences association for people re-identification across multiple non-overlapping cameras. *Traitement Du Signal*, 27(3):179–189, 2009.

[11] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2360–2367, 2010.

[12] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–6, 2007.

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

[14] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, pages 1–6, 2008.

[15] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian Conference on Image Analysis*, pages 91–102, 2011.

[16] S. Karaman and A. D. Bagdanov. Identity inference: Generalizing person re-identification scenarios. In *International Conference on Computer Vision*, pages 249–250, 2012.

[17] S. Karanam, Y. Li, and R. J. Radke. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In *IEEE International Conference on Computer Vision*, pages 4516–4524, 2015.

[18] W. Li and S. K. Shah. Subject centric group feature for person re-identification. In *Computer Vision and Pattern Recognition Workshops*, pages 28–35, 2015.

[19] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *IEEE Conference on Computer Vision & Pattern Recognition*, pages 3610–3617, 2013.

[20] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206, 2015.

[21] C. Liu, S. Gong, C. L. Chen, and X. Lin. Person re-identification: What features are important? In *European Conference on Computer Vision, International Workshop on Re-Identification*, pages 391–401, 2012.

[22] K. Liu, B. Ma, W. Zhang, and R. Huang. A spatio-temporal appearance representation for viceo-based pedestrian re-identification. In *IEEE International Conference on Computer Vision*, pages 3810–3818, 2015.

[23] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato. Hierarchical gaussian descriptor for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1363–1372, 2016.

[24] R. Mazzon, F. Poiesi, and A. Cavallaro. Detection and tracking of groups in crowd. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 202–207, 2013.

[25] P. M. Roth, M. Hirzer, M. K?stinger, C. Beleznai, and H. Bischof. *Mahalanobis Distance Learning for Person Re-identification*. Springer, 2014.

[26] J. Shao, C. C. Loy, and X. Wang. Scene-independent group profiling in crowd. In *CVPR*, pages 2227–2234, 2014.

[27] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *Computer Vision and Pattern Recognition*, 2015.

[28] N. Ukita, Y. Moriguchi, and N. Hagita. People re-identification across non-overlapping cameras using group features. *Computer Vision & Image Understanding*, 144(C):228–236, 2016.

[29] R. Vezzani and R. Cucchiara. Person re-identification. *Visual Analysis of Behaviour*, 42(7):301–313, 2013.

[30] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by video ranking. In *European Conference on Computer Vision*, pages 688–703, 2014.

[31] T. Xiao, H. Li, W. Ouyang, and X. Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[32] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *Computer VisionCeccv*, pages 1–16, 2014.

[33] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *Computer Vision and Pattern Recognition*, pages 1345–1352, 2011.

[34] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision*, pages 536–551, 2014.

[35] L. Zhang, T. Xiang, and S. Gong. Learning a discriminative null space for person re-identification. pages 1239–1248, 2016.

[36] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593, 2013.

[37] W. S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *British Machine Vision Conference, BMVC 2009, London, UK, September 7-10, 2009. Proceedings*, 2009.

[38] W. S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *The IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, Co, Usa, 20-25 June*, pages 649–656, 2011.

[39] W. S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 35(3):653–668, 2013.