This CVPR2015 workshop paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

Subject Centric Group Feature for Person Re-identification

Li Wei and Shishir K. Shah
Quantitative Imaging Laboratory
Department of Computer Science, University of Houston
Houston, TX 77204-3010, U.S.A.
lwei9@uh.edu, shah@cs.uh.edu

Abstract

This paper presents a subject centric group feature for person re-identification. Our approach is inspired by the observation that people often tend to walk alongside others or in a group. We argue that co-travelers' information, including geometry and visual cues, can reduce the re-identification ambiguity and lead to better accuracy, compared to approaches that rely only on visual cues. We introduce person-group feature to capture both geometry and visual information of co-travelers around a subject. We compute the dis-similarity between person-group features by solving an integer programming problem. The proposed approach is evaluated in its ability to improve the accuracy of re-identification of people traveling within groups. The results show that our approach outperforms state-of-the-art visual based as well as group information based methods.

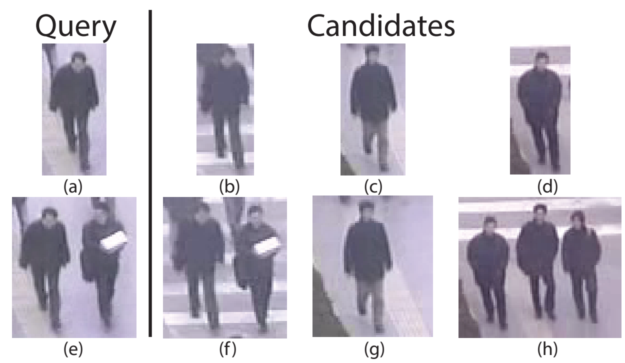

Figure 1: An example of group information assisting person re-identification. The first row is the persons' individual image, with (a) is the query and (b)˜(d) are candidates that matches with (a). (e)˜(h) are the group images query person and candidates.

1. Introduction

Person re-identification is a fundamental task in a multi-camera surveillance system to associate people across camera views at different locations and times [8]. With a growing network of cameras being used for security applications, manual re-identification that relies on a human operator is ineffective and lacking in reliability and scalability [14, 21]. Therefore, an automatic solution to person re-identification has received increasing attention from the computer vision community. Person re-identification is a challenging task and relies predominantly on visual features, such as clothing and the accessories that people carry. The visual features are intrinsically weak for matching people [8], because different people maybe dressed similarly, while the visual features of the same people may change significantly due to the changes in view angle, lighting and observed occlusions.

Many recent approaches have focused on solving the re-identification problem by developing a feature representation of a person, using low-level appearance features, such as color [22], texture [23] or their combinations [24]. Once a suitable representation is obtained, a distance metric is used to measure the similarity/dis-similarity between samples. In this paper, we refer to this methodology as the 'baseline method', on which we introduce the group information to improve the accuracy of re-identification.

The motivation of our approach is the observation that people often tend to walk alongside others or in a group. Such information can serve as context to reduce the ambiguity of person re-identification. If cameras are not geographically far apart, the same group structure would re-appear in neighboring cameras. Although the visual feature of one person could be different between cameras, by taking the co-travelers' information (geometry and visual) in to consideration, we can reduce re-identification ambiguity significantly. An intuitive example is shown in Figure 1, where (a) is the query person and (b) to (d) are the candidates' images. Considering only the individual images of candidates, it is difficult to point out the image that is most similar to (a) since all persons are dressed in dark color coats and long pants. The situation would be better if we also look into persons' group context. From (e) we can observe the query person walking with a co-traveler carrying a white object on the left side. With this information, we can tell that the first candidate has the highest possibility to match with (a), because in (f) we can observe there is a person carrying a white object walking on the left side of the first candidate, while in (g) we find that the candidate walks alone and in (h) we see that the candidate walks with two other persons.

Motivated by this example, we introduce a subject centroid feature, named person-group feature, to describe the person's profile within their belonging group. By combining the person-group feature with other approaches that measures the similarity/dis-similarity between individuals, we can improve the accuracy of re-identification. We evaluate our approach on the NLPR_MCT [1] dataset, using videos obtained from real scenarios and find an improvement in re-identification accuracy.

The main contributions of this paper include:

- We introduce a framework that can improve the baseline re-identification result using people grouping information.
- We propose a new person-group feature that encodes the person's profiles within the group, including in-group-position and co-travelers' baseline features. We also propose the metric for computing the distance between person-group features.
- We conduct rich experiments to demonstrate that our approach improves the baseline results to achieve higher accuracy (around 90% matching rate at rank 5 for group members), and out-performs other re-identification methods that also utilize group information.

2. Related Works

With the ubiquitous growth in cameras, recent approaches for re-identification have addressed both the single-shot and the multi-shot challenges. Multiple shot re-identification means that there are multiple images or video sequences that can be exploited for person re-identification. Compared to single shot re-identification, which mainly relies on appearance features [17, 4, 23, 22] from a single image, multiple shot re-identification could rely on much boarder types of features, such as spatial-temporal features [20, 7], accumulated appearance variability [11], etc. People gait information can also be extracted if video is provided, and re-identification can be solved using gait recognition [12, 15]. The time index of frames that are associated with a person can also be used to learn a probability model for non-overlapping camera tracking [13]. For a more comprehensive survey of re-identification approaches, please refer to recent surveys found in [19, 9]. Our method, presented in this paper, can take advantage of state-of-the-art approaches and improve the re-identification results by incorporating group information.

Group information based re-identification. The group information has been explored to improve re-identification in recent approaches. [25, 3] are the methods that most similar to ours. Zheng *et. al.* [25] proposed a method to associate groups of peoples in non-overlapping camera views. Their method explores group information as contextual cue for reducing the ambiguity in person re-identification if a person appears in the group. They propose a rotation invariant descriptor named Center Rectangular Ring Ratio-Occurrence Descriptor (CRRRO) to handle the person position change and camera viewpoint change. This approach addresses single shot re-identification and cannot easily be extended for multiple shot scenarios. The inputs of this method are manually selected person group images. This task in itself is time consuming, because finding groups manually in large video datasets is quite tedious and requires expertise. Our approach detects people groups automatically by clustering the person trajectories, and we introduce a person-group feature that is also robust to person position and camera view point changes. Cai *et. al.* [3] compute relative appearance context model of groups to mitigate ambiguities in individual appearance matching. Different to [25], Cai *et. al.* use a relaxed definition of group named neighboring set, which is a set of people that enter/exit at similar locations within a time frame. The groups under this definition have no social connection, therefore the assumption that same set of people will re-appear in a different camera views is weak. Cai *et. al.* [3]'; also assumes that appearance difference between pair of persons is similar across cameras, however, this assumption is also weak because the person appearance would significantly change due to the background, illumination and camera setting changes. In our approach, we use a group extraction method to detect groups that form social connections, and take advantage of state-of-the-art re-identification approach, to improve individual matching accuracy. We compare the results obtained using our approach against both of the above discussed methods [25, 3].

3. Methodology

An overview of our method is illustrated in Figure 2. Given two sets of person tracking results from Camera A and Camera B, our method computes a pair-wise score matrix that measures the dissimilarity score between persons from the two cameras. The baseline approach is a method that estimates the dis-similarity score of persons using the individual information only. There are many features that can be used in the baseline approach, such as appearance features, spatio-temporal features, and so on. The baseline

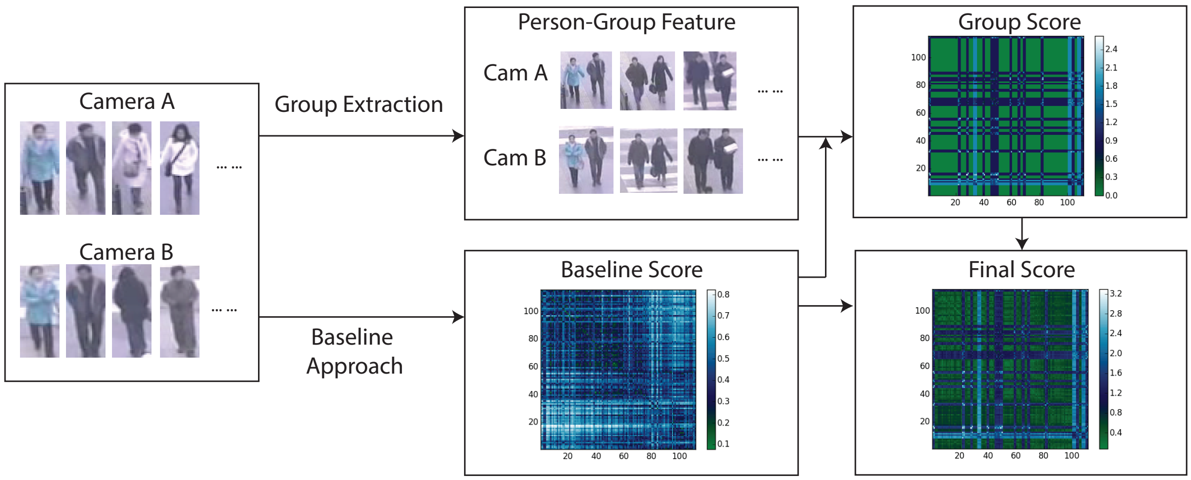

Figure 2: The overview of our approach

approach results in a pair-wise score matrix, and it serves as an initial re-id result.

Our method uses group information to improve the baseline score. First, we perform group extraction (Sec. 3.1) to extract groups from persons' trajectories. Then the person-group features (Sec. 3.2) are computed for each person. Person-group feature includes the in-group-position of a person and the information of group members. By comparing the person-group feature pair-wisely, we obtain a group score. The final score is obtained by combining group score and the baseline score (Sec. 3.3).

3.1. Group Extraction

In this section, we present a group extraction approach by clustering the person's trajectories observed in a camera view. In this paper, the group is defined as a set of person traveling together through the scene. In social science research conducted by McPhail and Wohlstein [16], they analyzed and summarized pedestrian behavior from a set of film records, and proposed the objective measure for people traveling together. The group members are determined by thresholds of difference in people's positions and velocities. Ge et. al. [6] directly applied these thresholds to automatically detect small groups in crowd automatically. However, we found that directly applying threshold does not provide robust results when persons' positions and velocities are noisy, because both are computed from person's on-ground trajectories, which is reconstructed from persons' tracking data. To improve the robustness of group extraction, we use a kernel function to compute the possibility of person grouping over frames. Next, we use affinity propagation to discover the clusters/groups of people.

Consider the trajectory of the person P_i in the scene as a set of sequence $L_i = \{(s_i^t, v_i^t)\}$, where s_i^t and v_i^t are the person's centroid (back-projected onto the ground using estimated homography) and velocity vector of P_i at frame t. Similar to [6], we compute the aggregated pairwise grouping possibility $W = [w_{ij}]$ over-time:

$$w_{ij} = \sum_{t=0}^{\infty} \delta_{ij}^t \exp\left(-\frac{\|s_i^t - s_j^t\|^2}{2\tau_s^2} - \frac{\|v_i^t - v_j^t\|^2}{2\tau_v^2}\right) / \sum_{t=0}^{\infty} \delta_{ij}^t \tag{1}$$

$$\delta_{ij}^t = \begin{cases} 1 & \text{Both } P_i \text{ and } P_j \text{ appear in the sence at frame } t \\ 0 & \text{Otherwise} \end{cases}$$

where τ_s and τ_v are the thresholds of spatial and velocity difference.

To identify the groups, we use clustering method to find the groups with the great internal grouping possibility. As we already compute the group possibilities between trajectories in Equation 1, we can use any clustering algorithm that takes pairwise distance/similarity as input, such as K-medoids or spectral clustering. However, both methods require the number of clusters as input, which is not easy to

Figure 3: Two examples of group extraction results. The images are video frames from two non-overlapping cameras. The persons' bounding boxes and trajectories of 2 seconds are shown on the figures. In each figure, the persons belong the same group are marketed using the same color.

obtain in our problem. Therefore, we use Affinity Propagation (AP) [5] to discover both the group numbers and group members. Each person forms a data point, and the grouping possibility matrix W is used as the similarity matrix, which is the input to AP. The output of AP is a set of exemplars and corresponding clusters/groups. We denote these groups as $G = \{g_i\}$. We also use $G(P_i)$ to denote the group that P_i belongs to. Figure 3 shows two examples of the group extraction results.

3.2. Person-Group Feature

In this section, we introduce the person-group feature, which describes two things about a subject within a group: who are the people that subject traveling with, and how they travel with that person. For first part, we collect the subject's co-travelers' baseline feature, and re-utilize the baseline score to evaluate similarity of co-travelers. For the second part, we propose an in-group-position signature to encode position of subject within group. We compute the local positions of co-travelers with respect to the subject's moving direction through time, the in-group-position signature is a set of co-travelers' positions. The distance measure between in-group-position signatures can be computed by solving the integer programming problem inspired by Earth Mover Distance [18].

In-group-position signature. Assume we want to construct the in-group-position signature of P_i, where P_i belongs to group $G(P_i)$. Firstly, for each $P_j \in G(P_i)$ and $P_j \neq P_i$, we have to compute the angles between P_j and the moving direction, from perspective of P_i through all frames. We denote (s_i^t, v_i^t) as P_i's position and velocity at frame t. Then the angle between P_j and moving direction

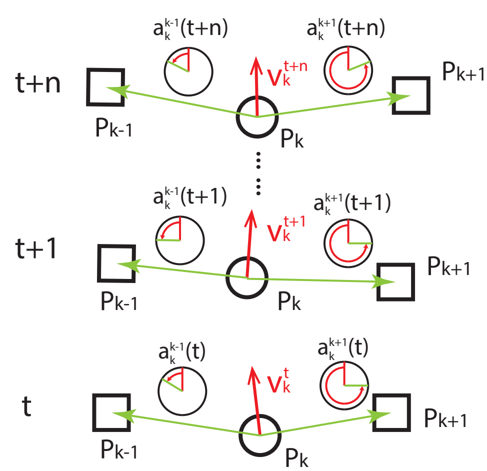

Figure 4: In-Group-Position Signature. The circle denotes the subject and rectangle detentes the co-traveler. Red arrows point to the subject moving direction.

is computed as:

$$\alpha_i^j(t) = \begin{cases} \Delta & \vec{\Gamma} \cdot \vec{Z} \geq 0 \\ 2\pi - \Delta & \text{Otherwise} \end{cases} \tag{2}$$

$$\Delta = cos^{-1} \frac{(s_j^t - s_i^t) \cdot v_i^t}{|s_j^t - s_i^t||v_i^t|}$$

$$\vec{\Gamma} = v_i^t \times (s_j^t - s_i^t)$$

$$\vec{Z} = (0, 0, 1)$$

We collect $\alpha_i^j(t)$ through all frames, which is fitted by a Gaussian distribution, and we denote this distribution as $\alpha_i^j = (\mu_i^j, \sigma_i^j)$, where μ_i^j is the mean angle and σ_i^j is the angle deviation. An illustration of in-group-position signature is shown in Figure 4.

As we collect the distributions for all group members in $G(P_i)$ except P_i, it forms a distribution set that is represented as $H_i = \{\alpha_i^j | P_j \in G(P_i), P_i \neq P_j\}$, which is in-group-position signature of P_i. We denote P_j's co-travelers baseline features as $B_i = \{\beta_i^j | P_j \in G(P_i), P_i \neq P_j\}$. Hence, we represent the person-group feature of P_i as $PG_i = (H_i, B_i)$.

Metric of Person-Group feature. Given person-group features, the distance measure between features are based on a linear combination of three terms: group size score, in-group-position score, and group baseline score. Let PG_i and PG_j denotes the person-group feature of P_i and P_j. Their distance takes the form:

$$D(PG_i, PG_j) = D_g(G(P_i), G(P_j)) + D_p(H_i, H_j) + D_b(B_i, B_j) \tag{3}$$

The first term D_g is the group size score, which return the size difference of groups that includes P_i and P_j. The group size score is computed by:

$$D_g(G(P_i), G(P_j)) = ||G(P_i)| - |G(P_j)|| \tag{4}$$

where $|G|$ is the group size (number of group members) of group G.

The second term D_p is the in-group-position score, which evaluates the difference between in-group-position signatures. As we know, $H_i = \{\alpha_i^j | P_j \in G(P_i), P_i \neq P_j\}$ is a set of distributions that encode the co-traveler's location around P_i. H_i is a distribution in metric space. The problem of computing distance between H_i and H_j becomes one of computing the distance between two distributions. There are many metrics that define distance between distributions. We found that the intuition behind Earth Mover Distance (EMD) [18] fits our problem best. EMD computes the distance between distributions in space by computing minimum cost of turning one distribution to another, where costs are assumed to be amount of weights moved, times the distance by which it is moved in space. The minimum cost can be solved as a linear programming problem. In our problem, we define the distance between in-group-position signature as the minimum amount of deformations that transfer one feature to another. However, unlike the original EMD algorithm, the person can only be transformed as a complete part, therefore integer programming is required to solve the minimum deformation in our problem.

Let $H_s = \{\alpha_s^1, ..., \alpha_s^m\}$ be the in-group-position signature of P_s, $H_t = \{\alpha_t^1, ..., \alpha_t^n\}$ be the in-group-position signature of P_t. As we mentioned above, all possible angle distribution belongs to a metric space M. The distance function of M is simply defined as the distance between the distributions' mean angle:

$$Dis(\alpha_s^m, \alpha_t^n) = \begin{cases} \frac{\Theta}{\pi} & \Theta \leq \pi \\ 2 - \frac{\Theta}{\pi} & \text{Otherwise} \end{cases}$$
$$\Theta = |\mu_s^m - \mu_t^n|$$

Let $D = [d_{ij}]$ be the difference between *i-th* element in H_s and *j-th* element in H_t. We try to find a flow $F = [f_{ij}]$, where f_{ij} is a binary variable, with $f_{ij} = 1$ when *i-th* element of H_s is moved to the same location of *j-th* element in H_t after the deformation. This optimization can be formulated as a binary integer programming problem:

$$F = \arg\min_F \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij} \tag{5}$$

subjects to the following constrains:

$$f_{ij} \in \{0, 1\}, 0 \leq i \leq m, 0 \leq j \leq n$$
$$\sum_{i=1}^{m} f_{ij} \leq 1, 1 \leq j \leq n$$
$$\sum_{j=1}^{n} f_{ij} \leq 1, 1 \leq i \leq m$$
$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(m, n)$$

After we solve the above optimization, the in-group-position signature distance is calculated using:

$$D_p(H_s, H_t) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{6}$$

The final term D_b is the group baseline score. It compute the aggregated differences of co-travelers' baseline features, under the condition that the co-traveler's correspondence is known by solving Equation 5. Let $R = [r_{ij}]$ be the pairwise baseline score matrix, where r_{ij} denotes the baseline score between *i-th* element in B_s and *j-th* element in B_t. The group baseline score takes the form:

$$D_b(B_s, B_t) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} r_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{7}$$

When a person is traveling alone, the person-group feature is empty. In this case the distance to an empty person-group feature D_p and D_b are set to zero and only group size score, G_g, contributes to the person-group feature difference.

3.3. Person Re-identification with Person-Group Feature

In Section 3.2, we introduced the person-group feature and defined the distance function between features. We argue that by combining the metric of person-group feature and baseline feature, we can improve the performance of person re-identification.

A simple way to combine two distance measurements is by linearly adding them:

$$D(P_i, P_j) = D(B_i, B_j) + D(PG_i, PG_j) \tag{8}$$

Where B_i is the baseline feature of P_i, and $D(B_i, B_j)$ means the baseline score of person P_i and P_j.

4. Results

To evaluate our approach, we test our method on the NLPR_MCT [1] dataset. The Dataset 1 and 2 of NLPR_MCT is used for evaluation. For both datasets, there are three synchronous videos (resolution: 320x240, 20 frames per-second) from three non-overlapping cameras. We use the videos produced by two outdoor cameras for evaluation. The number of people in each dataset are presented in Table 1. The dataset provides the ground truth annotation, which includes the bounding box tracking for each person. The persons' X-Y plane locations are computed by back projecting the mid-bottom of bounding boxes, and the homography is estimated interactively, off-line.

The group information is extracted using proposed algorithm in Section 3.1. In dataset 1, both Camera 1 and Camera 2 have 18 persons traveling with co-travelers and form 8 groups (with size greater than 1). In dataset 2, 35 and 31 persons travel with co-travelers and they form 16 and 15 groups in Camera 1 and Camera 2, respectively.

When one person in a camera is given, we computes their person-group feature and compute the distance to all the persons in another camera, and sort the persons in an ascending order based on the distance value. The rank score is the order of ground truth person in the sorted person list. Some examples of query and candidates person/group im-

	Camera 1	Camera 2	Common
Dataset 1	76	78	72
Dataset 2	115	111	105

Table 1: People Number of Evaluation Datasets

Figure 5: Two Examples of Re-identification Results. For each query, the image of query person and the group that person belongs to are shown. We display the matching results of baseline approach [4] and our approach. The top five candidate are shown, we display the image of candidate and the group that candidate belongs to in each grid. The ground truth matching is labeled by green boxes, where the rank is also given at right. The ranks with star symbols are the results obtain using our approach, otherwise the ranks are computed by baseline approach.

ages are demonstrate in Figure 5. The results obtained using [4] are also provided.

To test the performance of our method under difference baseline methods, we conduct experiments using the average RGB histogram of person's foreground pixels and the Symmetry-Driven Accumulation of Local Features (SDALF) [4]. Both methods require background subtraction, which is obtained using ViBe [2]. We measure the performance using Cumulative Match Curve (CMC) [10]. The results are shown in Figure 7. The matching rates comparison between our approaches and baseline methods at rank 1, 5, and 10 are given in Table 2. Our approach outperforms the baseline method, while the improvements are more significant in Dataset 2 than Dataset 1. This can be attributed to the fact that there are more persons traveling with co-travelers in Dataset 2.

Since our approach depends on the group information that given by the group extraction method, we want to discover how different group extraction algorithms affect the re-identification results. We choose Ge *et. al.* [6] as the comparing group extraction method. The results is illustrated in Figure 6. As seen, using group information extracted by either methods lead to an improvement in accuracy compared to the baseline approach. In Dataset 1, our approach provides similar accuracy as with [6], while

Dataset	Rank	RGB	Our w/ RGB	SDALF	Our w/ SDALF
1	1	0.18	0.24	0.22	0.25
	5	0.47	0.56	0.52	0.64
	10	0.67	0.75	0.74	0.80
2	1	0.33	0.36	0.37	0.44
	5	0.51	0.54	0.62	0.72
	10	0.61	0.66	0.73	0.82

Table 2: The matching rates comparison between our approach and baseline methods (RGB and SDALF)

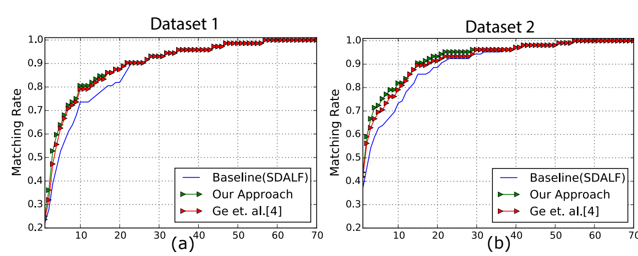

Figure 6: CMC curves for person re-identification using group information extracted using our approach and that of Ge. *et. al.* [6].

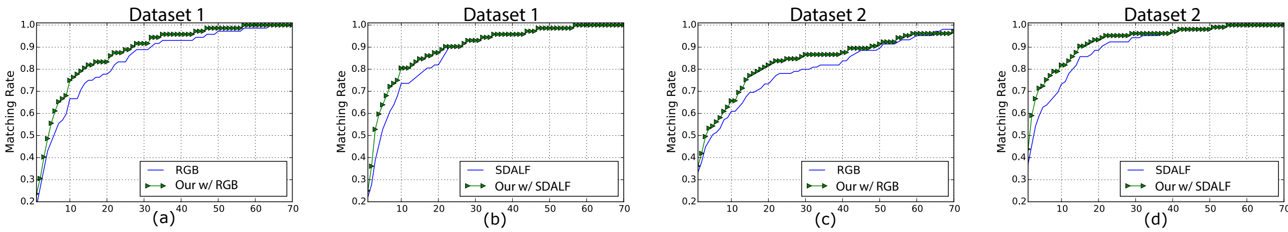

Figure 7: The comparison of CMC using baseline methods SDALF and RGB on two datasets.

in Dataset 2, our approach is slightly better. The reason is that Dataset 1 has less crowded scenes and the group extraction task is relatively easier. However, the scenes are more crowded in Dataset 2 and the performance of [6] is effected by directly using the threshold and noisy trajectories. Our group extraction algorithm handles noise better by computing the grouping probability using a kernel function.

We also compare our approach to [25] and [3], both of which use group information as context to improve the accuracy of individual re-id. The first approach [25] extracts Center Rectangular Ring Ratio-Occurrence (CRRRO) descriptor as group context feature from a manually selected static group image. Although our dataset consists of videos, we generate a group image by randomly picking one frame that includes all group members to compute CRRRO. The distance between CRRRO features are linearly combined with other appearance based distance as the final score. The second work uses Relative Appearance Context (RAC) feature as group context, which measures the appearance difference of person to the near-by people. The distance of appearance feature are also linearly combined with relative appearance context distance as the final distance value. To make sure the comparison is fair, in both comparison methods, we use SDALF to represent the individual appearance feature. We use parameters as suggested by respective authors in all our experiments.

Dataset	Rank	Our	SDALF	CRRRO	RAC
1-All	1	0.25	0.22	0.25	0.21
	5	0.64	0.53	0.57	0.53
	10	0.80	0.74	0.76	0.78
1-Group	1	0.39	0.28	0.38	0.22
	5	0.94	0.51	0.67	0.51
	10	0.94	0.67	0.78	0.83
2-All	1	0.44	0.37	0.41	0.38
	5	0.72	0.62	0.67	0.63
	10	0.82	0.73	0.78	0.77
2-Group	1	0.42	0.19	0.32	0.23
	5	0.87	0.54	0.67	0.58
	10	0.93	0.64	0.80	0.77

Table 3: Comparison of matching rates across methods that use group information for re-id.

Results are as shows in Figure 8. The matching rates at rank 1, 5, and 10 are given in Table 3. As seen, there is an overall improvement in re-identification accuracy. To further evaluate the impact of the person-group feature, we specifically restrict the dataset to those ID's that are found in a group. Results obtained on this restricted dataset are as shown in Figures 8(b) and (d).

As we can see from the results, our method provides the best performance in both datasets. By looking into the CMC for all persons (Figure 8(a) and (c)), we can observe that the accuracy is boosted through our approach. In general, the accuracy is slightly better than compared approaches. However, as seen through the CMC of group persons (Figure 8(b) and (d)), our method is able to reach accuracy of around 90% at rank 5, which is significantly better then the baseline method and compared approaches.

5. Conclusion

In this paper, we address the problem of person re-identification using subject centric group features. We proposed person-group feature that encodes the geometry and visual information of groups. The distance between

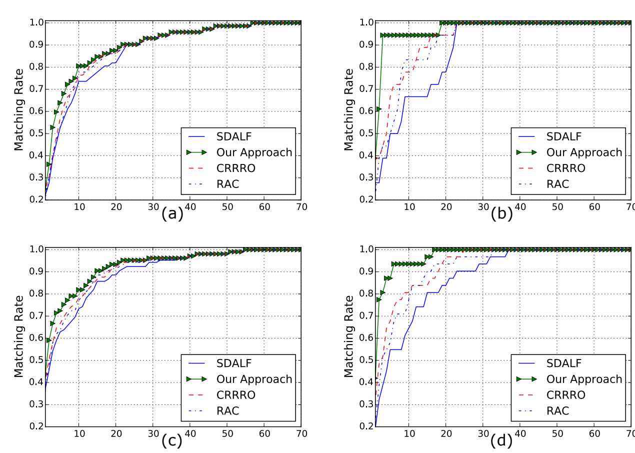

Figure 8: Compare the CMC of person re-identification using our approach, CRRRO descriptor and RAC feature.

person-group features are computed by solving an integer programming problem. The final distance ia a linear combination of person-group feature distance and a baseline distance obtained by considering feature of an individual. We demonstrate that our proposed method can always improve the accuracy of a baseline approach, and outperform the state-of-the-art group information based re-identification approaches. In current algorithm, we don't introduce weights to balance the contribution of different terms in subjects' distances because our approach is unsupervised. In the future, we plan to extend our method to be able to learn the optimal weights automatically. We also plan to explore the group/crowd behaviors to further reduce the ambiguity of person re-identification.

6. Acknowledgment

This work was supported in part by the US Department of Justice 2009-MU-MU-K004. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

References

[1] Multi-camera object tracking challenge, http://mct.idealtest.org/datasets.html, Augest 2014.

[2] O. Barnich and M. Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.

[3] Y. Cai and G. Medioni. Exploring context information for inter-camera multiple target tracking. *IEEE Winter Conference on Applications of Computer Vision*, pages 761–768, Mar. 2014.

[4] 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. IEEE, 2010.

[5] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

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

[7] N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1528–1535. IEEE, 2006.

[8] S. Gong, M. Cristani, C. C. Loy, and T. M. Hospedales. The re-identification challenge. In *Person Re-Identification*, pages 1–20. Springer, 2014.

[9] S. Gong, M. Cristani, S. Yan, and C. C. Loy. *Person re-identification*. Springer, 2014.

[10] P. Grother and P. J. Phillips. Models of large population recognition performance. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–68. IEEE, 2004.

[11] 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. IEEE, 2008.

[12] J. Han and B. Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):316–322, 2006.

[13] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking across multiple cameras with disjoint views. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 952–957. IEEE, 2003.

[14] H. Keval. Cctv control room collaboration and communication: Does it work? In *Proceedings of Human Centred Technology Workshop*, pages 11–12, 2006.

[15] R. Martín-Félez and T. Xiang. Gait recognition by ranking. In *Computer Vision–ECCV 2012*, pages 328–341. Springer, 2012.

[16] C. McPhail and R. T. Wohlstein. Using film to analyze pedestrian behavior. *Sociological Methods & Research*, 10(3):347–375, 1982.

[17] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *Proc. BMVC*, pages 21.1–11, 2010. doi:10.5244/C.24.21.

[18] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, pages 1–20, 2000.

[19] R. Vezzani, D. Baltieri, and R. Cucchiara. People reidentification in surveillance and forensics: A survey. *ACM Computing Surveys (CSUR)*, 46(2):29, 2013.

[20] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by video ranking. In *Computer Vision–ECCV 2014*, pages 688–703. Springer, 2014.

[21] D. Williams. Effective cctv and the challenge of constructing legitimate suspicion using remote visual images. *Journal of Investigative Psychology and Offender Profiling*, 4(2):97–107, 2007.

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

[23] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2528–2535. IEEE, 2013.

[24] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 144–151. IEEE, 2014.

[25] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *Proc. BMVC*, pages 23.1–23.11, 2009. doi:10.5244/C.23.23.