SOCIAL GROUPING FOR TARGET HANDOVER IN MULTI-VIEW VIDEO

Zhen Qin *Christian R. Shelton* *Lunshao Chai*[*]

University of California, Riverside
Beijing University of Posts and Telecommunications[*]
{zqin001,cshelton}@cs.ucr.edu chailunshao@bupt.edu.cn

ABSTRACT

This paper uses a social grouping model to improve target handover across multiple non-overlapping cameras to enable wide-area video understanding. Prior work focuses on modeling appearance and spatial-temporal cues for target handover. In cameras with different conditions, these cues are weak, at best. We provide a complete generative social grouping model which generalizes a recent single-camera case. Our extension requires strengthening the probabilistic interpretations and the resulting optimization over track handovers and social groupings can be formulated in terms of standard fast algorithms. We demonstrate the effectiveness of the method over existing techniques on challenging real-world multi-camera video.

Index Terms— Tracking, Video Understanding, Social Grouping, Optimization

1. INTRODUCTION

Multi-target tracking aims to maintain the trajectories and identities of multiple moving objects across video frames. It lays the foundation for understanding high-level events such as activity recognition and event detection, enabling applications such as automatic surveillance, content-based video retrieval and recommendation, and human-computer interaction [1, 2]. Compared to single camera tracking, multiple camera tracking—especially inter-camera tracking with non-overlapping fields of views (FOVs)—is a very challenging but less explored topic. However, the value of understanding wide-area videos makes it of great practical importance.

Multi-camera tracking systems conduct intra-camera tracking first, and then associate or "handover" targets across different cameras to achieve consistent labeling of multiple targets across large areas. The handover task with multiple non-overlapping camera views is difficult as traditional visual evidence for intra-camera tracking is *very* weak. Object appearance is unreliable across different cameras due to differing characteristics, view-points, or illumination conditions.

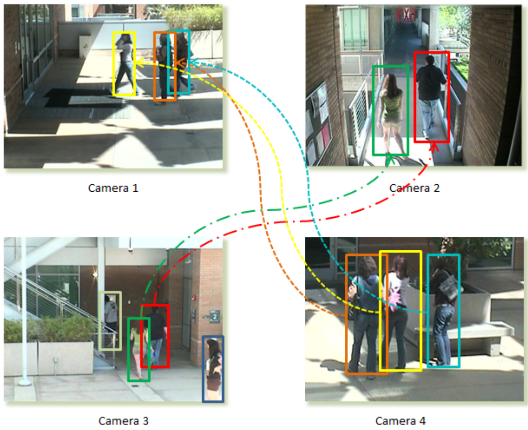


Fig. 1. An illustration of the multi-camera target handover problem and the dataset we use. Notice the severe appearance change, strong within-camera illumination change, and the heavy interactions among individuals. However, social grouping behavior generally exists.

To establish more robust appearance matching, researchers often learn the brightness transfer functions (BTFs) between each pair of cameras, which may not be invariant over time. Furthermore, the illumination conditions are not necessarily consistent within each camera. Spatial-temporal cues (such as location, time and velocity) are less reliable due to open blind areas between cameras. Finally, a multi-camera tracking system must be able to identify whether a track belongs to a newly entered person or should be linked to an old one. This can be challenging under the ambiguity inherent in the problem. In this work, we look beyond traditional visual cues and model social grouping behavior to mitigate ambiguities in the multi-camera handover problem. Sociology research shows that in natural scenes, up to 70% of people walk in groups as friends, couples or families, staying close to each other and possessing similar speeds and trajectories [3]. We note that social grouping behavior is essentially invariant to factors that make multi-camera handover difficult, such as camera conditions.

In this work, we consider the quality of target grouping for target handover for videos captured in multiple cameras. Our

Lunshao Chai conducted this work as a visiting student at UR-Riverside major contributions are (1) We propose to use social grouping information for the multi-camera tracking scenario, as a principled regularizer for the visually ambiguous tracking solution. This social model may be combined with existing methods. (2) We derive the model by building a complete generative model for multi-camera tracking using social grouping, which strengthens the probabilistic understanding of the system. We show in a unified framework that this model generalizes a recently proposed case for single camera and answers questions about implementation concerns of building social groups across multiple cameras.

2. RELATED WORK

Compared to single camera tracking, multi-camera tracking with non-overlapping FOVs is a less explored topic. As noted in the previous section, researchers typically employ spatial-temporal and appearance cues to handover targets across cameras with non-overlapping FOVs.

For spatial-temporal information, [4] used a Parzen window density estimator to jointly model the inter-camera travel time intervals, locations of exit/entrances, and velocities of objects. [5] proposed an unsupervised learning method to validate the camera network model. [6] described the observed pattern of pedestrian motion via a stochastic transition matrix.

In terms of appearance similarity, [7] showed that the BTFs between cameras lie in a low dimensional subspace and proposed a method to learn them with labeled correspondences. A cumulative brightness transfer function (CBTF) was proposed by [8] for mapping color between cameras using sparse training set. [9] used Multiple Instance Learning (MIL) to learn a discriminative appearance affinity model online. [10] proposed a new illumination-tolerant appearance representation. [11] evaluated several BTFs and showed that they demonstrated similar behaviors and limitations. Some notable work focuses on the general optimization or learning framework, but again, apply them only to appearance and spatial-temporal cues, either for Bayesian path reconstruction [12] or unsupervised incremental learning for better time invariance [13, 14].

Simultaneously, social behavior has attracted more attention in tracking. Researchers have used various social factors to improve tracking performance, including a pedestrian's destination, desired speed, repulsion from other individuals, and social grouping behavior [15, 16, 17, 18]. However, all these approaches are designed for intra-camera tracking and whether they can be applied to wide area video understanding is not clear. Recent work [19] employed a social force model to model *common paths* to help the multi-camera handover task without using appearance information, but it only focuses on individual tracks. In this paper, we build on [18] to use social *grouping* to improve multi-camera tracking. To the best of our knowledge, this work is the first to generally use other tracks as social grouping context for the multi-camera tracking problem.

3. GENERATIVE SOCIAL GROUPING FOR MULTI-CAMERA TRACKING

We first derive a complete generative model for target handover across cameras using social context. We then describe the choice for each component of the model, some of which are naturally derived from existing work.

3.1. Problem Formulation

Let the within-camera track set $T = \{T_1, \dots, T_N\}$ have N total input tracks in a time window. Each track T_i, over the time interval $t \in [t_i^{start}, t_i^{finish}]$, is a sequence of $\{(a_i(t), x_i(t))\}$ with $a_i(t)$ denoting the camera for track i at time t and $x_i(t)$ denoting the position of the track in the camera-centric coordinate system. (Note that for any t_1 and $t_2 \in [t_i^{start}, t_i^{finish}]$, $a_i(t_1) = a_i(t_2)$ as the track is within the same camera.) Traditional track handover can be modeled as a maximum a posteriori (MAP) problem [12] with the objective function

$$\phi^* = \underset{\phi \in \Phi}{\arg\max} P(\phi|T), \tag{1}$$

where the association result is represented by a correspondence matrix ϕ such that

$$\phi_{i,j} = \begin{cases} 1 & \text{if } T_j \text{ is linked to } T_i \text{ as the same target,} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

with the non-overlapping constraint that $\sum_j \phi_{i,j} = 1$ and $\sum_i \phi_{i,j} = 1$, meaning that each track can only precede or follow one track. Φ is the set of valid correspondence matrices.

In this work, we take the quality of social grouping behavior into consideration to help eliminate visual ambiguities in the multi-camera tracking system. We assume people form social groups denoted by the set $G = \{g_1 \dots, g_{|G|}\}$ where $|G|$ is the number of social groups and g_k is the description of the kth group. Each track should be assigned to one social group, as represented by a social grouping matrix ψ:

$$\psi_{i,k} = \begin{cases} 1 & \text{if track } i \text{ is assigned to group } k, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Again there is an added constraint that $\sum_k \psi_{i,k} = 1$ stating that each track can only be assigned to one group and we let Ψ be the set of valid social grouping matrices.

Then we can model the MAP formulation of track handover across cameras as

$$(\phi^*, \psi^*, G^*) = \underset{\phi \in \Phi, \psi \in \Psi, G}{\arg\max} P(\phi, \psi, G|T), \tag{4}$$

where

$$
\begin{aligned}
&P(\phi,\psi,G|T) \propto P(\phi,\psi,G)\; P(T|\phi,\psi,G)\\
&= P(G)\; P(\psi|G)\; P(\phi|\psi,G) \prod_{T_i \in T} P(T_i|\phi,\psi,G),
\end{aligned}
\tag{5}
$$

assuming that the likelihood of input tracks are conditionally independent given the track correspondences and group assignments. The likelihood of one track models the probability of false alarms. In this work, we set $P(T_i|\phi,\psi,G) = 1$ for all i as reliable single camera tracking algorithms shall largely remove false alarms. If a model of intra-camera tracking is available, it can be incorporated. The other three terms indicate the probability of a social group set and how this social setting generates the track group assignment and track handover.

3.2. Description of MAP Components

First, we model the probability of social groups as

$$
P(G) \propto e^{-\kappa|G|},
\tag{6}
$$

penalizing large numbers of social groups to avoid overfitting (such as placing each person in a separate group).

$P(\psi|G)$ models the probability of track-group assignment given social groups as

$$
P(\psi|G) \propto \prod_{i,k|\psi_{ik}=1} P(T_i|g_k),
\tag{7}
$$

where $P(T_i|g_k)$ is the likelihood that track i comes from group k, which we decompose across time as

$$
P(T_i|g_k) = \prod_{t=t_i^{start}}^{t_i^{finish}} P(a_i(t)|g_k)\; P(x_i(t)|a_i(t),g_k).
\tag{8}
$$

$P(a_i(t)|g_k)$ is the probability that group k appears at camera $a_i(t)$, a parameter of the model for group k which we denote as $b_{k,a}(t)$. $P(x_i(t)|a_i(t),g_k)$ is the probability that at time t, a member of the group in camera $a_i(t)$ will appear at position $x_i(t)$, which we model as a Gaussian centered around the mean $u_{k,a}(t)$, the position for group k in camera a at time t, also a parameter of the model for group k. We use a fixed variance σ for all such Gaussians.

$P(\phi|\psi,G)$ measures the probability of track handover given the social group information. Compared to other track handover methods, this adds a group constraint that if two tracks are linked (they are the same person), they belong to the same group (one group per person, but note that people-group assignment is free to change across different time windows):

$$
\begin{aligned}
P(\phi|\psi,G) = &\prod_{i|\forall m,\phi_{m,i}=0} P_{init}(T_i) \prod_{j|\forall m,\phi_{j,m}=0} P_{term}(T_j)\\
&\prod_{i,j|\phi_{i,j}=1} \begin{cases} P_{link}(i,j) & \text{if } \forall k, \psi_{i,k}=\psi_{j,k},\\ 0 & \text{otherwise.}\end{cases},
\end{aligned}
\tag{9}
$$

where $P_{init}(T_i)$ is the likelihood of T_i being an initial track, and $P_{term}(T_j)$ the likelihood of T_j being the last track. $P_{link}(i,j)$ is the likelihood that track j is the first instance following track i. This part is usually called the basic affinity model in track handover.

4. OPTIMIZATION

In the previous section, we proposed a complete generative model for multi-camera tracking using social grouping in the literature. Here we show that the resulting optimization can be formulated in terms of standard fast algorithms.

4.1. Optimization Reformulation

We perform optimization of Eq. 5 in the negative log-likelihood space (a minimization problem). First, since $-\ln P(T_i|\phi,\psi,G) = 0$, we drop that term. Ignoring an additive constant from the proportionality in Eq. 6,

$$
-\ln P(G) = \kappa|G|.
\tag{10}
$$

Ignoring a similar additive constant, for $P(\psi|G)$ (Eq. 7), we have $-\ln P(\psi|G) =$

$$
\begin{aligned}
&\sum_{i,k|\psi_{ik}=1} \sum_{t=t_i^{start}}^{t_i^{finish}} -\alpha \ln b_{k,a_i(t)}(t) + \beta|x_i(t) - u_{k,a_i(t)}(t))|^2\\
&= \sum_{i,k|\psi_{ik}=1} D(T_i,g_k)
\end{aligned}
\tag{11}
$$

from Eq. 8 where α and β are weighting parameters relating to the variance of the Gaussian. For simplicity, we use $D(T_i,g_k)$ to denote the "distance" of track i from group k.

$P(\phi|\psi,G)$ can be transformed to an assignment problem by defining a $2N \times 2N$ handover matrix

$$
H = \left(\begin{array}{c|c} H^{link}_{N\times N} & H^{term}_{N\times N} \\ \hline H^{init}_{N\times N} & \infty_{N\times N} \end{array}\right)
\tag{12}
$$

with $H^{link}_{i,j} = -\ln P_{link}(i,j)$, $H^{init}_{i,i} = -\ln P_{init}(T_i)$, $H^{term}_{i,i} = -\ln P_{term}(T_i)$ and infinity ($-\ln 0$) elsewhere. Eq. 9 is 0 if any assignments violate the constraint that linked tracks must be in the same social group. Therefore, if we add this constraint ($\forall i,j,k\ \phi_{i,j}(\psi_{i,k} - \psi_{j,k}) = 0$), the resulting equation can be written in terms of H:

$$
-\ln P(\phi|\psi,G) = \sum_{i,j} \phi_{i,j} H_{i,j}
\tag{13}
$$

Our optimization's outer loop tries different numbers of social groups. Inside, we can drop Eq. 10 and minimize the sum of Eq. 13 and Eq. 11 with the above constraint:

$$
\begin{aligned}
&\underset{\phi\in\Phi,\psi\in\Psi,G}{\arg\min} \quad \sum_{ij} \phi_{i,j} H_{i,j} + \sum_{ik} \psi_{i,k} D(T_i,g_k)\\
&\text{s.t.} \quad \forall i,j,k \quad \phi_{i,j}(\psi_{i,k} - \psi_{j,k}) = 0.
\end{aligned}
\tag{14}
$$

We call Eq. 14 the primal problem.

4.2. A Two-stage Alternating Minimization Algorithm

We apply the two-stage iterative alternative optimization algorithm proposed in [18] to solve Eq. 14 by first applying Lagrange theory, yielding

$$\begin{aligned} L(\phi,\psi,G,\mu) =& \sum_{ij}\phi_{i,j}H_{i,j} + \sum_{ik}\psi_{i,k}D(T_i,g_k) \\ &+ \sum_{ijk}\mu_{ijk}\phi_{i,j}(\psi_{i,k}-\psi_{j,k}), \end{aligned} \tag{15}$$

in which the μs are the Lagrange multipliers. The dual of this problem is

$$\begin{aligned} &\max q(\mu) \\ \text{where} \quad &q(\mu) = \min_{\phi\in\Phi,\psi\in\Psi,G} L(\phi,\psi,G,\mu). \end{aligned} \tag{16}$$

The resulting correspondence ϕ of the optimization is the output of the method. For a fixed μ, let

$$(\phi^{\mu},\psi^{\mu},G^{\mu}) = \mathop{\arg\min}_{\phi\in\Phi,\psi\in\Psi,G} L(\phi,\psi,G,\mu). \tag{17}$$

To solve Eq. 16, we use a quasi-Newton strategy with limited-memory BFGS updates and Wolfe line search conditions guided by the subgradient:

$$\left.\frac{\partial q}{\partial \mu_{ijk}}\right|_{\mu} = \phi^{\mu}_{i,j}(\psi^{\mu}_{i,k}-\psi^{\mu}_{j,k}). \tag{18}$$

To calculate the subgradient, we use a two-stage block coordinate-minimization algorithm to solve Eq. 17. The first stage minimizes over ϕ (the track correspondence result) from Eq. 15 with ψ and G fixed:

$$\phi^{\mu} = \mathop{\arg\min}_{\phi\in\Phi}\sum_{ij}\phi_{i,j}[H_{i,j}+\sum_{k}\mu_{ijk}(\psi_{i,k}-\psi_{j,k})]. \tag{19}$$

This amounts to adding a penalty term to the matrix scores (compare with Eq. 13). So Eq. 19 is a standard assignment problem and can be efficiently solved by the Hungarian algorithm.

The second stage minimizes Eq. 15 over ψ and G, with ϕ fixed: $(\psi^{\mu},G^{\mu}) =$

$$\mathop{\arg\min}_{\psi\in\Psi,G}\sum_{ik}\psi_{i,k}[D(T_i,g_k)+\sum_{j}(\mu_{ijk}\phi_{i,j}-\mu_{jik}\phi_{j,i})]. \tag{20}$$

This amounts to a standard K-means clustering problem. If the "centers", G, are fixed, the assignments, ψ, are made to minimize the augmented distance. When the assignments are fixed, the centers can be placed to minimize their distances to the captured points. Several initial group assignments are tried, as K-means converges to local minimum. The output of the one with the minimum value for Eq. 16 for one specific $|G|$ (number of groups) is maintained.

At the end we add the linear penalty of $|G|$ indicated by Eq. 10 and the outer loop (over $|G|$) selects the solution with the minimal negative log-likelihood score.

5. IMPLEMENTATION DETAILS

In this section we describe some implementation details of the system and how our model generalizes the single-camera case.

5.1. Building the Basic Affinity Model

We build the basic affinity model (H in Eq. 13) using a Brightness Transfer Function (BTF) for appearance (app) cues and a Parzen window density estimator for spatial-temporal (st) cues:

$$-\ln P_{link}(i,j) = -\ln p^{app}_{i,j} - \ln p^{st}_{i,j}. \tag{21}$$

We use the BTF model in [7] for $-\ln p^{app}_{i,j}$ and the Parzen window technique in [4] for spatial-temporal information $-\ln p^{st}_{i,j}$. Readers could refer to [7] and [4] for more detail. $P_{init}(T_i)$ and $P_{term}(T_i)$ are set to be a single constant (from training) for simplicity.

5.2. Augmented Social Grouping Model for Multi-camera Tracking

The remaining problem is how to implement the two steps of K-means clustering: group update (when the group assignments are given) and track assignment (when group parameters are given).

Recall that we modeled the group mean trajectory for g_k as, at each time t, a distribution over which camera a member of the group appears in ($b_{k,\cdot}(t)$) and a mean position within each camera a that a group member would appear ($u_{k,a}(t)$). Track assignment (finding ψ given a fixed G) is simple: for each track T_i, compute $D(T_i,g_k)$ from Eq. 11 for each group g_k and select the one that minimizes the negative log-likelihood.

For group update of g_k with the assignment ψ fixed, we must find the parameter assignments to $b_{k,\cdot}$ and $u_{k,\cdot}$ that maximize the likelihood. The log-likelihood is a sum across time, so the maximization can be done independently at each time point. $b_{k,a}(t)$ is a multinomial parameter and therefore its maximum likelihood estimate is proportional to the number of tracks assigned to group k at time t that are in camera a.

$u_{k,a}(t)$ is the conditional mean for group k at time t in camera a. Therefore, its maximum likelihood parameter is the average position of all tracks assigned to group k at time t in camera a. If at any point there are no tracks for group k and camera a, we use linear interpolation or extrapolation to generate a mean. If no tracks in camera a are ever assigned to group k, we place $u_{k,a}(t)$ in the middle of the image for all t.

If there is only one camera, the distribution $b_{k,a}(t)$ is degenerate and drops out of the equations. The remaining model is the same as in [18], thus this is a generalization to multiple cameras.

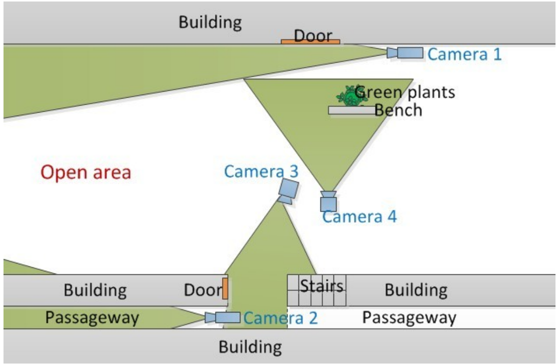


Fig. 2. Topology of the cameras in the experiments.

6. EXPERIMENT

Compared to single camera tracking, datasets publicly available for real-world multi-camera tracking with non-overlapping FOVs are extremely limited. In this work, we test our method using two sets of videos on the VideoWeb dataset [20]. We choose Cam27, Cam20, Cam36 and part of Cam21 (indexed by 1–4) to establish the desired non-overlapping setting, the topology of which is shown in Fig. 2. Multi-camera tracking in this setting is very challenging for the following reasons. (1) We use 4 cameras, unlike most prior work that use 2–3. (2) This is an outdoor dataset with a cluttered environment. As shown in Fig. 1, there is severe within-camera illumination change, which makes traditional methods that establish one single transformation between each camera pairs, such as BTFs, much less reliable. (3) Since this dataset is mainly designed for complex real-world activity recognition, there exist heavy interactions among individuals, unlike "designed" tracking datasets, *e.g.* [7].

We compare our proposed multi-camera social grouping behavior tracking (MulSGB) to directly using the Bhattacharyya distance between RGB color histograms without BTF transformation or spatial-temporal information (Color), Parzen window estimation for spatial-temporal information and the original color histogram for appearance (Parzen Window) in [4], and the BTF plus Parzen window estimation framework in [7](Parzen Window + BTF).

Due to the availability of the dataset, we gather 9 videos using all 4 cameras and 4 videos with camera 1–3. We use 5 videos from the first set for training and all the other videos for testing (note the second set of videos contains a subset of cameras of the first set so no additional training is needed). All other videos in the dataset either had no inter-camera motion or were missing data for more cameras. The data used has roughly 40,000 frames (25fps) for each of the four cameras for training and 80,000 frames for each camera for testing. For detection, we use a state-of-art pedestrian detector [21] to get detection responses and generate reliable intra-camera tracks using [18]. We use time windows of length 4000 frames and allow frame gaps as long as 1200 frames (only tracks with a time gap of less than 1200 frames can be linked). We hand-labeled ground truth and measure the percentage of correctly linked pairs for the eight testing scenes (which consist of 244 single-camera tracks in total). Fig. 3 and Fig. 4 show the results for each set of videos.

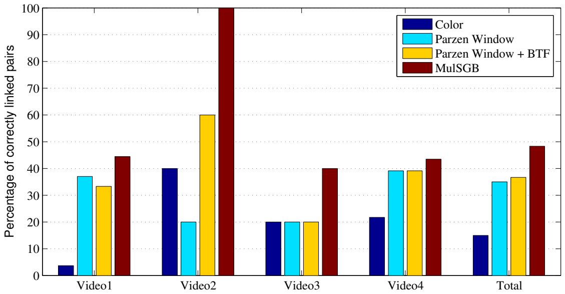


Fig. 3. Percentage of correctly linked pairs on the four video sequences with four cameras. Each video sequence consists of 27, 5, 5 and 23 (60 in total) ground truth linked pairs respectively.

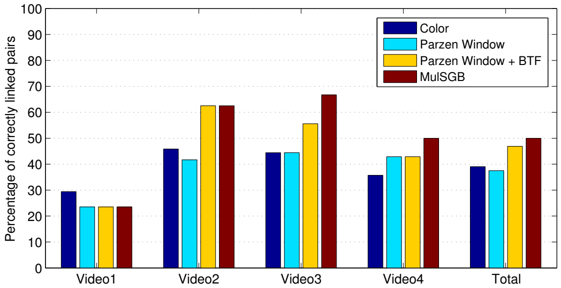
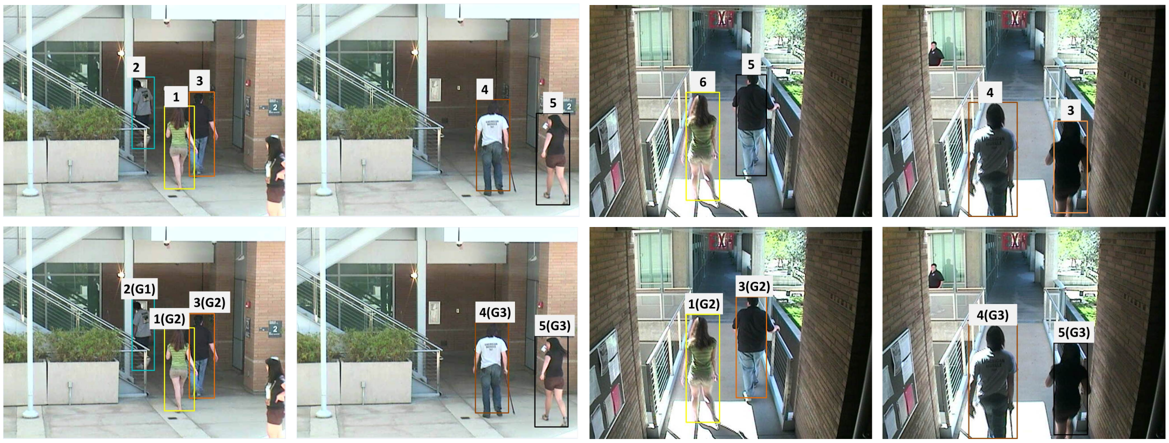


Fig. 4. Percentage of correctly linked pairs on the four video sequences with three cameras. Each video sequence consists of 17, 24, 9 and 14 (64 in total) ground truth linked pairs respectively.

We have the following observations. (1) Given the poor color histogram result especially for the four-camera setting (demonstrating the difficulty of the dataset), the overall performance is very promising, as our MulSGB model indeed improves tracking performance over competing methods. (2) The example in Fig. 5 shows a representative example where social grouping helps to disambiguate, while other methods fail under this challenging sequence. (3) Since our social grouping model serves as a regularizer, the basic affinity model upon which we built social grouping model is sometimes a bottleneck, especially for challenging sequences as in our case. For example, we observe no improvement upon the baseline model for two sequences in Fig. 4. We observed that in such cases, although the optimization usually heads toward a good solution, it could not recover wrong links since the basic model provides very unlikely handover possibility between the correct pairs. For example, when the illumination condition changes between the testing set and training set, the learned BTF may even hurt the performance comparing to pure color histogram comparison, as is the case for video1 in Fig. 4.



Frame 5809 (Cam 3) Frame 6222 (Cam 3) Frame 6295 (Cam 2) Frame 6689 (Cam 2)

Fig. 5. Example tracking results (first row: [7] without social grouping, second row: ours with social grouping, where G indicates group number). Because people form groups and show proximity to group members, social grouping provides powerful contextual information to improve multi-camera tracking. Other methods tend to identify a new person (Frame 6295 target 1) or output an identity switch (target 3 and 5) on this sequence, because traditional evidences are highly unreliable

7. CONCLUSION

We offer an explicit generalization of a single-camera social grouping behavior model to the multi-camera tracking problem with non-overlapping FOVs to enable wide-area video understanding. We propose a generative social grouping model that strengthens the probabilistic interpretation of social grouping generation, for which the single camera tracking is a special case. Our experiments on a very challenging publicly available real-world dataset show improvements over other methods.

8. REFERENCES

[1] H. Sabirin, J. Kim, and M. Kim, “Graph-based object detection and tracking in h.264/avc bitstreams for surveillance video,” in *ICME*, 2011.

[2] C. Fagiani, M. Betke, and J. Gips, “Evaluation of tracking methods for human-computer interaction,” in *WACV*, 2002.

[3] M. Moussaid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, “The walking behavior of pedestrian social groups and its impact on crowd dynamics,” *PLoS ONE*, vol. 5, 2010.

[4] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, “Tracking across multiple cameras with disjoint views,” in *ICCV*, 2003.

[5] D. Makris, T. Ellis, and J. Black, “Bridging the gaps between cameras,” in *CVPR*, 2004.

[6] A. Dick and M. Brooks, “A stochastic approach to tracking objects across multiple cameras,” in *Proc. Australian Conf. Artificial Intelligence*, 2004.

[7] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, “Modeling inter-camera space-time and appearance relationships for tracking across non overlapping views,” in *CVIU*, 2008.

[8] B. Prosser, S. Gong, and T. Xiang, “Multi-camera matching using bi-directional cumulative brightness transfer functions,” in *BMVC*, 2008.

[9] C.-H. Kuo, C. Huang, and R. Nevatia, “Inter-camera association of multi-target tracks by on-line learned appearance affinity models,” in *ECCV*, 2010.

[10] C. S. Madden, E. Cheng, and M. Piccardi, “Tracking people across disjoint camera views by an illumination-tolerant appearance representation,” in *Mach. Vis. Appl.*, 2007.

[11] T. D’Orazio, P.L. Mazzeo, and P. Spagnolo, “Color brightness transfer function evaluation for non-overlapping multi-camera tracking,” in *ICDSC*, 2009.

[12] V. Kettnaker and R. Zabih, “Bayesian multi-camera surveillance,” in *CVPR*, 1999.

[13] A. Gilbert and R. Bowden, “Tracking objects across cameras by incrementally learning inter-camera color calibration and patterns of activity,” in *ECCV*, 2010.

[14] K.-W. Chen, C.-C. Lai, Y.-P. Huang, and C.-S. Chen, “An adaptive learning method for target tracking across multiple cameras,” in *CVPR*, 2008.

[15] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in *ICCV*, 2009.

[16] P. Scovanner and M.F. Tappen, “Learning pedestrian dynamics from the real world,” in *ICCV*, 2009.

[17] Stefano Pellegrini, Andreas Ess, and Luc Van Gool, “Improving data association by joint modeling of pedestrian trajectories and groupings,” in *ECCV*, 2010.

[18] Z. Qin and C.R. Shelton, “Improving multi-target tracking via social grouping,” in *CVPR*, 2012.

[19] R. Mazzon and A. Cavallaro, “Multi-camera tracking using a multi-goal social force model,” *Neurocomputing*, 2012.

[20] G. Denina, B. Bhanu, H. Nguyen, C. Ding, A. Kamal, C. Ravishankar, A. Roy-Chowdhury, A. Ivers, and B. Varda, “Videoweb dataset for multi-camera activities and non-verbal communication,” *Distributed Video Sensor Networks*, 2010.

[21] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part based models,” *IEEE Trans. PAMI*, 2010.