Joint Tracking and Gait Recognition of Multiple People in Video

Maryam Babaee, Gerhard Rigoll, Mohammadreza Babaee

Institute for Human-Machine Communication, TU Munich, Germany

{maryam.babaee, rigoll, reza.babaee}@tum.de

Introduction

Abstract: We propose a novel approach to address the problem of jointly tracking and gait recognition of multiple people in a video sequence. The most state of the art algorithms for gait recognition consider the cases where there is only one person without any occlusion in a very constrained environment. However, in real scenarios such as in airports, train stations, etc, there are many people in the environment that make these algorithms inapplicable. Although first tracking of each person and then gait recognition could be a solution, we argue that the multipeople tracking and the gait recognition in a video are two sub-problems that can help each other. Hence, we propose a joint tracking and gait recognition of multiple people as one framework that can improve gait recognition accuracy and decrease the ID switching in tracking. Experimental results confirm the validity of proposed approach.

Gait Recognition

- Appearance-based features for gait recognition

- Gait Energy Image (GEI) : The average over a complete gait cycle silhouettes [1]
- Gradient Histogram Energy Image (GHEI): The average over a complete gait cycle after computing gradient histograms at all locations of the original images [2]
- Gait Flow Image (GFI): The average of the optical flow field from the binary silhouettes of each cycle[3]
- Gait Entropy Image (GEnI): Encoding in a single image the randomness of pixel values in the silhouette images over a complete gait cycle [4]

Tracking by Detection as Min-Cost Flow Graph

The tracking problem is achieved by finding an optimal set of tracks T , which has the Maximum a-Posteriori (MAP) probability given detections (D) as observations [5].

$$T^* = \arg\max_T P(T \mid D) = \arg\max_T P(D \mid T)P(T) \Leftrightarrow T^* = \arg\min_T \sum_{D_k \in T_u} -\log P(D_k \mid T) + \sum_{T_u \in T} -\log P(T_u)$$

$$\text{s.t.} \quad T_u \cap T_v = \phi, \forall u \neq v$$

Mapping to a constrained Min-Cost Flow Graph:

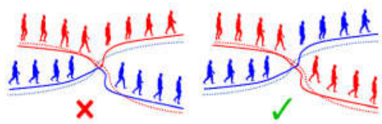
$$T^* = \arg\min_T \sum_k C_k f_k + \sum_k C_{en,k} f_{en,k} + \sum_{k,l} C_{k,l} f_{k,l} + \sum_k C_{ex,k} f_{ex,k}$$

$$\text{s.t.} \quad f_{en,k} + \sum_l f_{l,k} = f_k = f_{ex,k} + \sum_l f_{k,l}$$

D: person detections
T_u: a track/trajectory (a set of detections)
T^* : an optimal set of tracks
$f_{k,l}$: flow through an edge (value: 0 or 1)
C_k: Costs

Gait recognition of Multiple People in Video

- Most approaches proposed for gait recognition are able to recognize a person from images containing only one person.
- To identify a person in a multiple people video sequence, obviously, it is required to first track the underlined person along the video
- Reducing ID-switching error during tracking with the help of prior gait information



(a) (b)

Fig. 1. The problem of ID-switching in tracking.
(a) Id-switching happens and the gaits are not consistent.
(b) there is no ID switching and the gaits are consistent.
The dot lines show ground truth trajectories and the solid lines show the trajectories obtained by tracking.

Approach

Joint Tracking and Gait Recognition

Approach: Combining tracking and gait recognition into one single framework (MAP formulation)

$$T^* = \arg\max_T \; P(T \mid D, G) = \arg\max_T \; P(D \mid T)P(G \mid T)P(T) = \arg\max_T \prod_{D_k \in D} P(D_k \mid T) \prod_{G_i \in G} P(G_i \mid T) \prod_{T_u \in T} P(T_u)$$

$$\text{s.t.} \quad T_u \cap T_v = \phi, \forall u \neq v$$

Converting above MAP problem into a cost-flow graph and optimizing by Binary Integer Programming (BIP):

$$T^* = \arg\min_T \sum_k C_k f_k + \sum_i C_{G_i} G_i + \sum_k C_{en,k} f_{en,k} + \sum_{k,l} C_{k,l} f_{k,l} + \sum_k C_{ex,k} f_{ex,k}$$

$$\text{s.t.} \quad f_{en,k} + \sum_l f_{l,k} = f_k = f_{ex,k} + \sum_l f_{k,l}$$

Costs:

$$C_k = \omega_{\det} (\log \frac{\beta}{1-\beta})$$

$$C_{G_i} = -\log(Score_{SVM} (G_i))$$

$$C_{k,l} = \omega_{geo} Z_{geo} (k,l) + \omega_{gait} Z_{gait} (k,l)$$

$$Z_{gait} (k,l) = -\log(1 - dist(D_l, D_k) / dist_{\max})$$

$$dist(D_l, D_k) = \begin{cases} 0, & \text{if GID}_l = GID_k \\ 3, & \text{if GID}_l \neq GID_k \\ 1, & \text{if no ID has been assigned} \end{cases}$$

G_i: a set of detections shwoing one complelt gait cycle
$C_{k,l}$: cost of the edge between detections k and l
β: the false positive rate of person detector
$\omega_{det}, \omega_{geo}, \omega_{gait}$: constant values
GID_k: the predicted gait ID of k^{th}detection obtained by SVM
$Score_{SVM}(G_i)$: the predicted score of G_i obtained by SVM (in range [0,1])

Flowchart of Proposed Technique

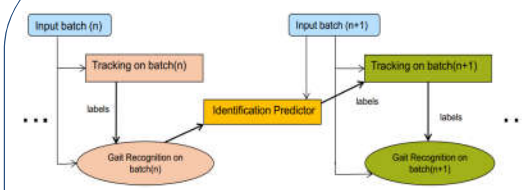
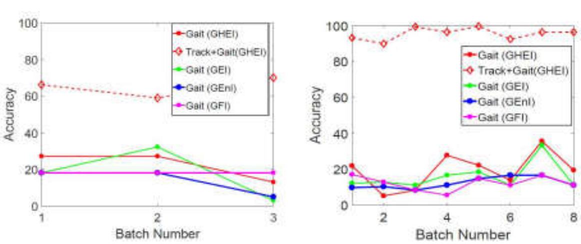


Fig. 2. The flowchart of the proposed joint tracking and gait recognition

As for joint tracking and gait recognition, we utilize the gait features to improve the association of the detections of the same person along the sequence. After applying tracking on batch n, the detections with assigned labels are preprocessed in order to be fed to a Support Vector Machine (SVM) as training data. The SVM gets segmented-GHEI images (GHEI images after background subtraction) of detection boxes labeled by the tracking algorithm as training data. After training, the SVM predicts the label of all possible sub-trajectories for the next batch (i.e. Batch $n + 1$). The SVM gives a score S_i to each sub-trajectory showing how likely this sub-trajectory belongs to the class i. If two different sub-trajectories have the same GID (Gait ID), one that has higher score is selected as the trajectory corresponding to that person. This prior gait ID information is used for tracking in next batch.

Experiments

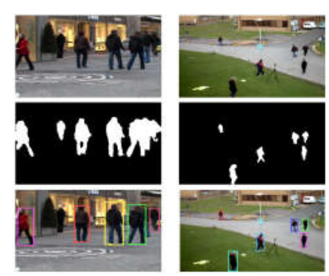


(a) TUD-Stadtmitte (b) S2L1-View 1

Fig. 3. The comparison of the two types of methods; 1) pure gait recognition (Gait) with different gait features and 2) joint gait-tracking (Track + Gait).

Table 1. Tracking results of S2L1 and TUD-Stadtmitte

Seq.	Method	MOTA	MOTP	IDS	MT	PT	ML
S2.L1	Berclaz et al. [6]	80.3	72.0	13	73.9	17.4	8.7
	Milan et al. [7]	90.3	74.3	22	78.3	21.7	0
	Pirsiavash et al. [8]	77.4	74.3	57	60.9	34.7	4.3
	Andriyenko et al. [9]	88.3	79.6	18	82.6	17.4	0
	Chari et al. [10]	85.5	76.2	56	94.7	0	0
	Milan et al. [11]	85.3	77.5	9	100	0	0
	Ours-Track	96.0	81.3	11	94.74	5.26	0
	Ours-Track-Gait	**96.0**	**81.8**	**8**	**100**	**0**	**0**
Stadtmitte	Milan et al. [7]	56.2	61.6	15	44.4	55.5	0
	Chari et al. [10]	51.6	**61.7**	15	20.0	**80.0**	0
	Ours-Track	51.21	60.58	13	30.0	60.0	10.0
	Ours-Track-Gait	**63.06**	57.77	**9**	**50.0**	50.0	**0**



(a) TUD-Stadtmitte (b) S2L1-View 1

Fig. 4. Identification results on the two datasets; The input images(first row), the segmented images (second row), and the recognition results (third row).

References

[1] J. Han and B. Bhanu , " Individual recognition using gait energy image ," PAMI, Feb 2006, pp:316–322.
[2] M. Hofmann and G. Rigoll, "Exploiting gradient histograms for gait-based person identification," 20th IEEE International Conference on Image Processing (ICIP). IEEE, 2013, pp. 4171–4175
[3] T.H. Lam, K.H. Cheung, J.N Liu, " Gait flow image: A silhouette-based gait representation for human identification ," Pattern Recognition. 2011;44, pp.:973–987.
[4] K. Bashir, T. Xiang, S. Gong , " Gait recognition using gait entropy image ," Proceedings of 3rd International Conference on Crime Detection and Prevention (ICDP 2009); London, UK. 3 December 2009; pp. 1–6.
[5] M. Hofmann, D. Wolf, and G. Rigoll, "Hypergraphs for joint multi-view reconstruction and multi-object tracking," in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE International Conference on, pp.3650–3657.
[6] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," IEEE transactions on pattern analysis and machine intelligence, 2011, vol. 33, no. 9, pp. 1806–1819.
[7] A. Milan, K. Schindler, and S. Roth, "Detection-and trajectory-level exclusion in multiple object tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3682–3689.
[8] H. Pirsiavash, D. Ramanan, and C. Fowlkes, "Globallyoptimal greedy algorithms for tracking a variable number of objects," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 1201–1208.
[9] A. Andriyenko, K. Schindler, and S. Roth, "Discretecontinuous optimization for multi-target tracking," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 1926–1933.
[10] V. Chari, S. Lacoste-Julien, I. Laptev, and J. Sivic, "On pairwise costs for network flow multi-object tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5537–5545.
[11] A. Milan, L. Leal-Taix´e, K. Schindler, and I. Reid, "Joint tracking and segmentation of multiple targets," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5397–5406.