

# Multi-Pose Multi-Target Tracking for Activity Understanding

Hamid Izadinia  
University of Central Florida  
Orlando, FL  
izadinia@eecs.ucf.edu

Varun Ramakrishna, Kris M. Kitani, Daniel Huber  
Carnegie Mellon University  
Pittsburgh, PA  
vramakri, kkitani, dhuber@cs.cmu.edu

## Abstract

We evaluate the performance of a widely used tracking-by-detection and data association multi-target tracking pipeline applied to an activity-rich video dataset. In contrast to traditional work on multi-target pedestrian tracking where people are largely assumed to be upright, we use an activity-rich dataset that includes a wide range of body poses derived from actions such as picking up an object, riding a bike, digging with a shovel, and sitting down. For each step of the tracking pipeline, we identify key limitations and offer practical modifications that enable robust multi-target tracking over a range of activities. We show that the use of multiple posture-specific detectors and an appearance-based data association post-processing step can generate non-fragmented trajectories essential for holistic activity understanding.

## 1. Introduction

We explore the task of *multi-target multi-pose* person tracking for *activity-rich* surveillance videos using the current tracking paradigm of tracking-by-detection and data association. Advances in robust category-specific object detectors [5, 6] have motivated the *tracking-by-detection* paradigm, where robust detectors can act as strong observation models in tracking frameworks. In particular, recent work has shown that a single coarse part-based model (e.g., 5 to 15 parts) [7, 10, 22] is well-suited for detecting, representing and tracking upright people. While these approaches are effective for urban scenarios, such as pedestrians walking on sidewalks or people in subway stations, difficulties arise when people perform other activities like riding a bike, digging a hole, or pushing a cart. Although methods exist for full body pose estimation [21, 8, 24], they often assume full body part visibility. In this work, we target surveillance videos that contain a range of human activity, beyond walking and standing. We evaluate the strengths and limitations of state-of-the-art multi-target tracking and offer practical modifications to improve performance.



Figure 1. DARPA Mind's Eye Y2 activity dataset

We proceed with our analysis by dividing the tracking pipeline into two stages: person detection and data association. In the *person detection* stage, we compare the results of standard pedestrian detectors against richer models that encode variations in pose. In particular, we compare four different deformable part-models (DPMs) and show that training models explicitly for different postures improves performance. In the *data association* stage, we use a state-of-the-art multi-target data association framework [20] and examine how the choice of parameters affects the resulting trajectories. Specifically, we evaluate the tradeoff between the recall rate and the number of ID switches as a function of the parameters. To prevent frequent ID switching and to preserve longer trajectories, we propose an instance-specific trajectory merging process as a post-processing step, that uses appearance-based cues to make associations over long periods of time.

The contributions of this paper are as follows: (1) step-by-step analysis of detection-based data association tracking for activity-rich videos; (2) a multi-pose deformable parts model that allows for robust tracking over pose variations; and (3) long term data association using target-specific appearance-based regressors.

Work on multi-pedestrian tracking is a significant field

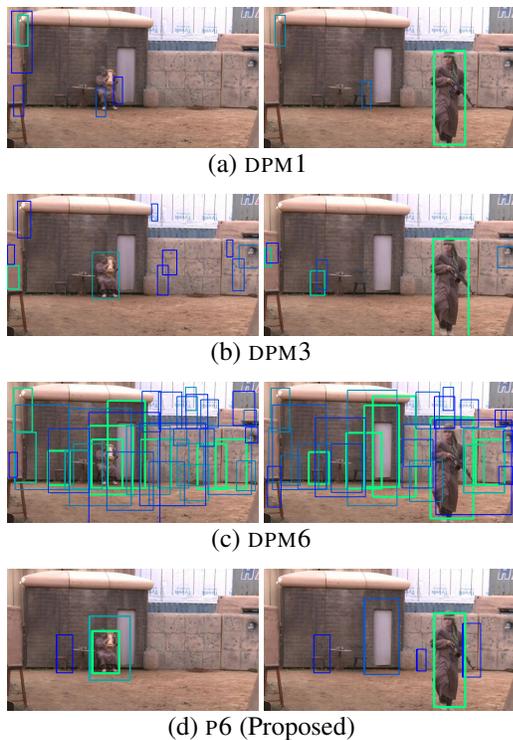


Figure 2. Examples of detections for different models: (a) INRIA single mixture DPM, (b) VOC09 three mixture DPM, (c) Mind’s Eye six mixture DPM, and (d) multi-pose DPMs (Proposed). Detection score visualized with colored bounding boxes from -0.9 (blue) to -0.3 (green), where higher scores are stronger detections.

of research, driven by many applications in surveillance, monitoring and human activity understanding. In the past decade, we have seen progress in two significant directions: (1) pedestrian-specific detection and (2) batch data association to recover trajectories.

Inspired by the success of the histogram of gradients (HOG) pedestrian detector [5], the use of a category-specific detector as an observation model in a tracking framework has become a standard approach for many pedestrian tracking algorithms. The two advantages of person-specific tracking are the discriminative power of a pedestrian detector and the elimination of a manual initialization step. Although standard pedestrian detectors, such as the DPM [7], robustly detect upright pedestrians, they faces difficulties when applied to activity-rich datasets that include non-upright postures.

The second direction of research is the transition from online filtering-based algorithms, such as Kalman filtering and particle filtering [3], to batch smoothing algorithms [2, 11, 15, 25, 19, 9, 1, 20, 23] that analyze all pedestrian detections as a batch process. While the latter approaches are not well-suited for online tracking applications, batch approaches find trajectories by allowing the algorithm to look forward and backward in time. However, most ap-

proaches assume that people are only walking or standing in the scene, and therefore use activity-agnostic features such as velocity, location and appearance.

It is interesting to note that work on pedestrian tracking has largely progressed independent of work on pose-level human activity analysis. Few researchers [17, 4] have examined the relationship between tracking and activity analysis. In this work, our goal is to evaluate the current paradigm of detection-based multi-pedestrian tracking and to understand how this approach performs applied to an activity-rich dataset. As a result of our analysis, we make the case for a recall-based evaluation metric, that gives preference to temporally persistent (non-fragmented) trajectories. On a practical level, we propose two modifications that allow for more robust multi-pose multi-target tracking, by using multiple pose detectors and an appearance-based data association post processing step.

## 2. Multi-Pose Multi-Target Tracking

Many tracking pipelines can be divided into two stages: person detection and data association. In the following section we describe the strengths and shortcomings of each stage and propose modifications that improve overall tracking performance.

### 2.1. Multi-Pose Person Detection

In a standard tracking-by-detection framework, a single person detector is applied to each image frame to create a set of potential detections. Then detections are chained together in a data association step to generate full trajectories. For many surveillance applications, such as pedestrians in a parking lot or a subway, a single model tuned to detect upright people is sufficient to capture people in the scene. However, when people engage in activities, such as bending down or sitting down, a single model is not expressive enough to represent various poses. Therefore, to enable more robust detection of people while performing such activities, it is essential that the person model be modified to include various poses. To this end, we extend the representational power of a single DPM by generating multiple DPMs. Specifically, we train six models: standing, straddling, walking, sitting, squatting and bending over. These pose categories were derived from a set of low-level pose elements used in the HOMINE ontology [18] to reason about higher level activities.

We show qualitative examples of detection with our model against several baseline models in Figure 2, and quantitative results are discussed in Section 3. Notice that our model trained for specific poses returns a strong score for a person sitting, while the single mixture DPM (DPM1) and three mixture DPM (DPM3) either miss or return only weak responses. The six mixture DPM (DPM6) learned automatically from all poses only loosely models different

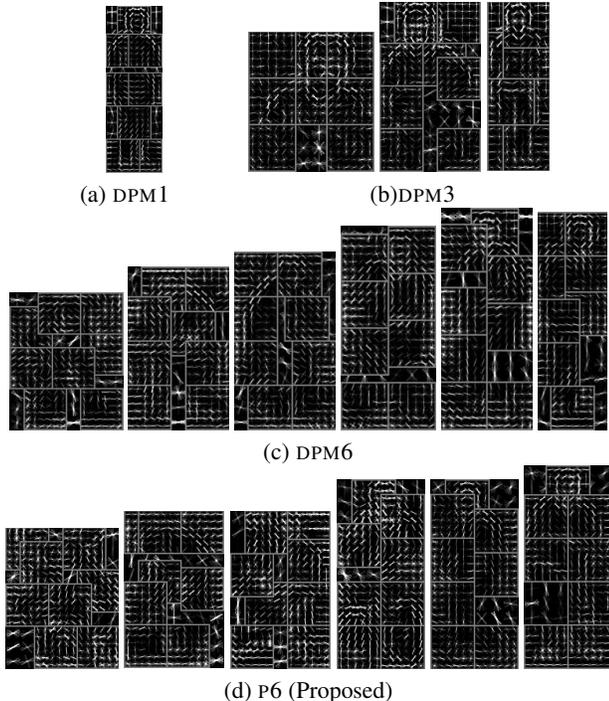


Figure 3. Evaluated DPM models: (a) DPM1: INRIA single mixture DPM, (b) DPM3: VOC09 three mixture DPM (c) DPM6: Mind’s Eye six mixture DPM and (d) P6: Proposed multi-posture DPM models of person *squatting*, *bending over*, *sitting*, *straddling*, *walking*, and *standing*.

postures and tends to generate a high number of false positives. Each model is visualized in Figure 3.

## 2.2. Graph-based Multi-Target Tracking

The second stage of the pipeline is batch data association. In this stage, each detection is treated as a node in a graph with various edge potentials. In this work, we adopt the graph-based approach of [20] to obtain continuous trajectories. Each detection is instantiated as a node in a graph. Nodes are connected by edges depending on distance and scale differences between associated detections. The optimal set of edges can be computed in a greedy way such that the obtained solution is globally optimal. In Section 3, we perform a quantitative analysis of this approach and observe the trade-off between the recall rate and the number ID switches as parameters are adjusted. Our analysis shows that this approach has a strong tendency to fragment trajectories into smaller tracklets.

## 2.3. Detection Propagation

In order to prevent the fragmentation of trajectories, we propose a pre-processing step that aims to generate a smooth set of detection hypothesis to make it easier for the data association algorithm to find continuous trajectories. Specifically, we increase the pool of confident detection

proposals by tracking the top scoring detections forward and backward in time, similar to [1]. We use a conservative template matching approach to instantiate new, yet temporally consistent detections. Propagated detections near the original detection are assigned a score that scales linearly with time (10% decrease per time step). We refer to these additional detection instances as ‘propagated detections.’ Section 3 provides a quantitative analysis of detection propagation.

## 2.4. Tracklet Association

Even with the introduction of propagated detections, the tracks produced by chaining together detections are significantly fragmented due to missed detections and occlusions. Furthermore, the framework of [20] is not able to handle large gaps, such as an individual exiting and re-entering. This limitation suggests the need for an additional level of data association to connect fragmented tracklets over long periods of time. Indeed, layered tracklet association schemes have been proposed for several multi-target algorithms [16, 9]. Traditional association approaches have used generic appearance-based features, such as color histograms, local texture histograms or part appearance models as a means of defining a generic distance metric. However, recent work has shown that discriminative object-specific descriptors can be beneficial for robust tracking [10, 12, 14, 22]. Following the direction of previous work, we implement an additional data association step that merges shorter tracklets over long temporal duration by leveraging invariant appearance information. In particular, our proposed approach is a greedy algorithm that builds discriminative appearance classifiers online from a color histogram to find the best matches between an active tracklet and a list of terminated tracklets. We used color histograms because they are invariant to large variations in pose and scale. Due to potentially large variations in posture, we have not used gradient-based spatial histogram features (*e.g.*, HOG, SIFT).

For each tracklet  $j$ , we train a random forests regressor  $f_j : \mathbf{x} \rightarrow y, y \in [-1, 1]$ , where the detection boxes are used as positive examples and negative examples are generated from neighboring patches of equal size and other detections from the same frame.

Our appearance feature is a  $(64 \times 16) \times 2$  dimensional joint color histogram over hue (64 bins) and saturation (16 bins) for the top and bottom portions of the detected bounding box. To remove the effect of the background included in the bounding box, we use a Gaussian mixture background model [26] to generate a foreground mask to retain only foreground appearance. We also found that removing trajectories that include detections with no foreground pixels greatly reduces the number of false positives.

Our greedy algorithm merges the current tracklet  $i$  with

the best matching tracklet  $\hat{j}$  according to

$$\hat{j} = \arg \max_j D(i, j), \quad (1)$$

where the score between two tracklets is computed as a two-way cumulative mean sum of regressor responses. The score is defined as

$$D(i, j) = \frac{1}{M_i} \sum_m f_j(\mathbf{x}_m) + \frac{1}{N_j} \sum_n f_i(\mathbf{x}_n), \quad (2)$$

where  $M_i$  and  $N_j$  are the number of detections in trajectory  $i$  and  $j$ , respectively. If the maximum score falls below a threshold (0.50 in our experiments), trajectory  $i$  is considered to be a new trajectory.

### 3. Experimental Analysis

To quantify tracking performance we report the CLEAR-MOT [13] metrics: Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP), on four sequences from the Mind’s Eye Y2 dataset (Figure 1). MOTA is a score that jointly quantifies the number of missed detections, false positives, and ID switches. MOTP measures the precision of object location estimates, independent of ID switches, and trajectory gaps. We also include the precision, recall and number of ID switches. The precision and recall of a trajectory is computed from the number of overlapping bounding boxes (between detection boxes and ground truth boxes) where any overlap greater than 50% is considered a true positive.

Although many tracking frameworks use the MOTA score as a general measure of overall performance, we argue that from the perspective of activity analysis, the number of ID switches and the trajectory recall rate have the greatest impact on end-to-end performance. A high number of ID switches (IDS) implies that the full trajectory has been fragmented into many small trajectories. Since traditional action recognition approaches have no mechanism for dealing with multiple trajectory fragments, it is only possible to classify the action for a single fragment (*i.e.*, losing all information about other fragments), resulting in an incomplete understanding of a specific individual’s activity. Similarly, the recall rate is also an important metric as it quantifies the number of lost detections. However, the recall rate alone can be deceiving as it does not encode ID switching.

#### 3.1. Detector Analysis

To evaluate the benefit of using our multiple pose detection approach, we compare our proposed model against (1) DPM1: a single mixture DPM trained on the INRIA pedestrian dataset, (2) DPM3: a three mixture DPM trained on the VOC 2009 person dataset [7], and (3) DPM6: a six mixture

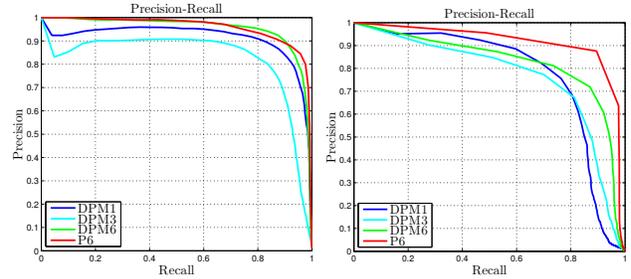


Figure 4. Precision-recall curve for DPM1, DPM3, DPM6 and our proposed P6. ROAD 1 (left) contains predominantly upright people, resulting in negligible change in performance. ROAD 2 (right) contains various postures, resulting in a significant boost in performance for our proposed model.

DPM model training on a disjoint set of the Mind’s Eye activity dataset. The root filters superimposed with part filters for each model are shown in Figure 3.

We hypothesize that adding multiple pose detectors will increase the number of detections, effectively increasing the recall rate, and will therefore allow the data association stage to track people that undergo changes in posture. However, we are also concerned about the number of false positives that are introduced by simply adding more models. Figure 4 shows the change in the precision-recall (PR) curve for the different detection models for two different videos. Since the ROAD 2 sequence contains many changes in posture, notice the significant increase in recall and precision for the proposed multi-pose person detector (P6) versus other models. However, for the ROAD 1 sequence, which contains mostly upright people, we see only modest improvements.

Table 1 shows the scores for each detector after data association using [20]. We observe that our proposed detector P6 consistently generates the highest recall rate among the models. At this step in the process, a high recall rate is important because this means that more true detections are available for the tracklet association step. As we shall see, the tracklet association step is quite robust at handling many false positives and consistently improves precision.

#### 3.2. Multi-Target Tracking and ID Switching

To gain a better understanding of the characteristics of data association using [20], we analyzed the change in performance using a parameter sweep. Figure 5 shows how the number of ID switches changes as a function of the maximum cost parameter (maximum cost of a trajectory). Notice that the minimum number of ID switches is quite large despite the change in the maximum cost parameter. This shows that the data association step is conservative and has a strong tendency to divide trajectories into smaller fragments.

Table 1. Multi-target tracking results [20] with different detector models. Scores for a single mixture (DPM1), a three mixture (DPM3), a six mixture (DPM6), and six posture model (P6). **Bold** text indicates best values.

SAFE HOUSE 1	MOTP	MOTA	Precision	Recall	IDS
DPM1	73.69	58.01	96.83	60.92	40
DPM3	74.30	<b>59.96</b>	98.50	61.63	32
DPM6	71.62	21.22	70.45	36.83	<b>7</b>
P6 (Proposed)	<b>76.47</b>	56.38	84.95	<b>68.97</b>	16
SAFE HOUSE 2	MOTP	MOTA	Precision	Recall	IDS
DPM1	75.80	37.68	<b>96.61</b>	39.70	66
DPM3	<b>76.76</b>	<b>44.65</b>	96.08	47.22	68
DPM6	75.68	20.93	91.74	23.39	<b>38</b>
P6 (Proposed)	71.87	31.65	69.34	<b>58.10</b>	80
ROAD 1	MOTP	MOTA	Precision	Recall	IDS
DPM1	74.31	<b>81.26</b>	95.51	86.31	60
DPM3	73.86	48.80	98.37	50.04	25
DPM6	<b>76.78</b>	68.35	<b>99.35</b>	69.02	<b>13</b>
P6 (Proposed)	75.95	36.12	62.63	<b>91.38</b>	45
ROAD 2	MOTP	MOTA	Precision	Recall	IDS
DPM1	75.04	64.62	92.41	73.39	46
DPM3	<b>76.68</b>	22.32	97.23	23.03	<b>1</b>
DPM6	75.17	20.23	<b>99.71</b>	20.41	2
P6 (Proposed)	74.76	<b>76.07</b>	83.19	<b>95.94</b>	8

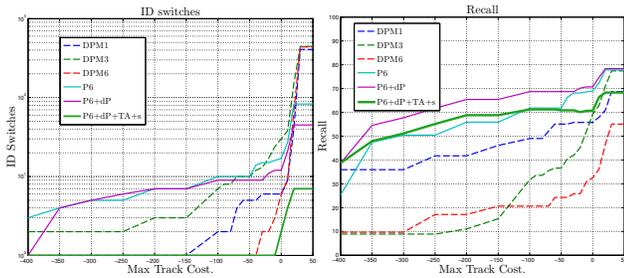


Figure 5. A parameter sweep of the maximum trajectory cost shows that a reduction in ID switches comes at the cost of the recall rate. Our proposed tracklet association (TA) shown in green, reduces the number of ID switches to zero with a minimal decrease in recall.

### 3.3. Appearance-based Tracklet Association

Based on our analysis, we observed that our multi-pose detector yields a high recall rate and a high number of ID switches after data association. While a high recall is favorable, the high number of ID switches poses a serious problem for activity recognition. Here, we analyze the benefits of supplementing the data association step with an additional tracklet association post-processing step based on a person’s appearance. Table 2 shows a significant decrease in the number of ID switches and an improvement in the overall MOTA score by adding the tracklet association (TA) post-processing step.

To gain more insight into the improvement introduced through tracklet association, we performed a parameter sweep of the maximum trajectory cost parameter of [20] and analyzed the change in the number of ID switches. Figure 5 shows that to reduce the number of ID switches to zero, the cost parameter must be lowered to at least -50 for DPM6 (red dashed line). In this regime, the recall rate drops

Table 2. Ablative analysis over tracklet association (TA), detection propagation (dP) and smoothing (s). **Bold** text indicates best values.

SAFE HOUSE 1	MOTP	MOTA	Precision	Recall	IDS
P6	76.47	56.38	84.95	68.97	16
P6+TA	<b>79.72</b>	<b>57.32</b>	<b>94.58</b>	60.85	2
P6+dP	76.53	56.50	83.51	<b>70.71</b>	11
P6+dP+TA	79.49	54.94	91.34	60.72	<b>1</b>
P6+dP+TA+s	79.62	54.94	91.34	60.72	<b>1</b>
SAFE HOUSE 2	MOTP	MOTA	Precision	Recall	IDS
P6	71.87	31.65	69.34	58.10	80
P6+TA	73.14	<b>40.81</b>	<b>92.24</b>	44.77	<b>20</b>
P6+dP	71.73	38.73	73.00	<b>62.86</b>	93
P6+dP+TA	73.50	41.67	89.47	47.49	24
P6+dP+TA+s	<b>73.73</b>	40.75	88.58	47.02	22
ROAD 1	MOTP	MOTA	Precision	Recall	IDS
P6	75.95	36.12	62.63	91.38	45
P6+TA	77.96	<b>78.48</b>	97.24	80.88	6
P6+dP	76.00	38.83	63.23	<b>94.42</b>	41
P6+dP+TA	78.26	76.72	98.45	78.02	<b>4</b>
P6+dP+TA+s	<b>78.38</b>	76.96	<b>98.59</b>	78.14	<b>4</b>
ROAD 2	MOTP	MOTA	Precision	Recall	IDS
P6	74.76	76.07	83.19	95.94	8
P6+TA	<b>76.90</b>	<b>85.68</b>	<b>95.85</b>	89.56	<b>0</b>
P6+dP	75.11	71.30	79.91	<b>96.36</b>	14
P6+dP+TA	77.32	73.21	93.92	78.28	<b>0</b>
P6+dP+TA+s	76.35	65.33	89.19	74.34	<b>0</b>

by more than 50% for DPM6. This means that many trajectory fragments are simply removed to reduce the number of ID switches. In contrast, our proposed tracklet association step can eliminate ID switches with only a 12% drop in recall. This means that more trajectory fragments are retained for the trajectory of a single person.

We also experimented with two minor modifications – detection propagation (mentioned in section 2.3) and trajectory smoothing – in an effort to improve the smoothness of trajectories. Table 2 shows that while detection propagation (dP) consistently increased the recall rate of data association, there was no significant impact on performance after the tracklet association step. It is also interesting to note that while smoothing trajectories (Gaussian smoothing after data association and after tracklet association) had a visual impact on the smoothness of the trajectories, the metrics show only minor improvement in performance since we used conservative smoothing parameters.

## 4. Discussion and Conclusion

In contrast to traditional work on multi-target tracking, we have evaluated tracking performance in the light of applications for activity analysis. As mentioned above, the recall rate and number of ID switches is a critical factor in successfully applying action recognition techniques to tracker outputs. We believe that the consideration of downstream action recognition is an important perspective when evaluating tracking algorithms.

Our use of multiple pose detectors also illustrates the advantage of coupling aspects of action recognition with tracking to improve joint performance, as shown in work

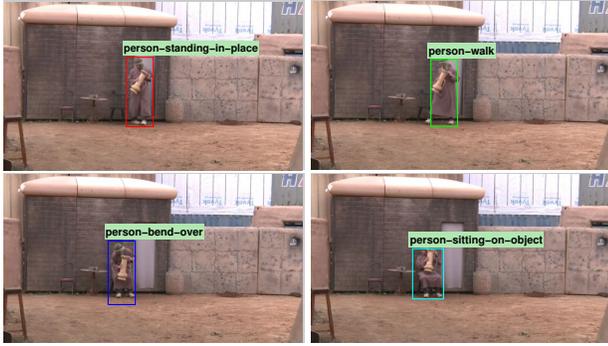


Figure 6. Action recognition via detection score aggregation.

such as [17, 4]. Since our proposed approach uses detectors explicitly trained for different postures, we are also able to use detection scores as a means of understanding low level actions. Figure 6 shows the potential for action recognition by aggregating detection scores within our tracking framework.

We have presented a step-by-step analysis of a tracking-by-detection and batch data association paradigm for multi-target tracking over activity-rich videos. We have shown qualitatively and quantitatively the characteristics of various aspect of the tracking pipeline. We have identified several key modifications that extend the state-of-the-art to multi-pose tracking. In particular, our analysis of multi-pose multi-target tracking has shown that increasing the number of pose models increases the number confident detections over a wider range of postures while adding minimal noise. We have also shown that the addition of an appearance-based data association post-processing step is effective for preserving long-term trajectories. On average, we showed that our proposed detector improves recall by 23% against the DPM3 and that our tracklet association yields an 87% decrease in the number of ID switches.

## Acknowledgements

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0061. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## References

[1] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *CVPR*, 2011. 2, 3

[2] J. Berclaz, F. Fleuret, and P. Fua. Robust People Tracking with Global Trajectory Optimization. In *CVPR*, 2006. 2

[3] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. J. V. Gool. Robust tracking-by-detection using a detector confidence particle filter. In *ICCV*, 2009. 2

[4] W. Choi and S. Savarese. A Unied Framework for Multi-Target Tracking and Collective Activity Recognition. In *ECCV*, 2012. 2, 6

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 1, 2

[6] P. Dollár, S. Belongie, and P. Perona. The fastest pedestrian detector in the west. In *BMVC*, 2010. 1

[7] P. Felzenszwalb, D. McAllester, and D. Ramaman. A Discriminatively Trained, Multiscale, Deformable Part Model. In *CVPR*, 2008. 1, 2, 4

[8] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Progressive search space reduction for human pose estimation. In *CVPR*, 2008. 1

[9] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *CVPR*, 2008. 2, 3

[10] H. Izadinia, I. Saleemi, W. Li, and M. Shah. (MP)<sup>2</sup>T: Multiple People Multiple Parts Tracker. In *ECCV*, 2012. 1, 3

[11] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In *CVPR*, 2007. 2

[12] Z. Kalal, J. Matas, and K. Mikolajczyk. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In *CVPR*, 2010. 3

[13] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *PAMI*, 31(2):319 – 336, 2009. 4

[14] C.-H. Kuo and R. Nevatia. How does person identity recognition help multi-person tracking? In *CVPR*, 2011. 3

[15] B. Leibe, K. Schindler, and L. V. Gool. Coupled detection and trajectory estimation for multi-object tracking. In *ICCV*, 2007. 2

[16] Y. Li, C. Huang, and R. Nevatia. Learning to associate: Hybrid-boosted multi-target tracker for crowded scene. In *CVPR*, 2009. 3

[17] W.-L. Lu and J. J. Little. Simultaneous tracking and action recognition using the pca-hog descriptor. In *Canadian Conference on Computer and Robot Vision*, 2006. 2, 6

[18] A. Oltramari and C. Lebiere. Mechanism meet content: Integrating cognitive architectures and ontologies. In *Advances in Cognitive Systems*, 2011. 2

[19] A. G. A. Perera, C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu. Multi-object tracking through simultaneous long occlusions and split-merge conditions. In *CVPR*, 2006. 2

[20] H. Pirsiavash, D. Ramanan, and C. Fowlkes. Globally-Optimal Greedy Algorithms for Tracking a Variable Number of Objects. In *CVPR*, 2011. 1, 2, 3, 4, 5

[21] D. Ramanan, D. Forsyth, and A. Zisserman. Strike a Pose: Tracking People by Finding Stylized Poses. In *CVPR*, 2005. 1

[22] B. Yang and R. Nevatia. Online Learned Discriminative Part-Based Appearance Models for Multi-Human Tracking. In *ECCV*, 2012. 1, 3

[23] B. Yang and R. Nevatia. An Online Learned CRF Model for Multi-Target Tracking. In *CVPR*, 2012. 2

[24] Y. Yang and D. Ramanan. Articulated pose estimation using flexible mixtures of parts. In *CVPR*, 2011. 1

[25] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network ows. In *CVPR*, 2008. 2

[26] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *ICPR*, 2004. 3