MAX-PLANCK-GESELLSCHAFT
max planck institut informatik

Learning People Detectors for Tracking in Crowded Scenes

Siyu Tang[1] Mykhaylo Andriluka[1] Anton Milan[2] Konrad Schindler[3] Stefan Roth[2] Bernt Schiele[1]

[1]Max Planck Institute for Informatics [2]TU Darmstadt [3]ETH Zürich

Goal

- Detect and track ***all*** the people in the crowded street scenes

Motivation

- Detection and tracking failures are related to frequent and long-term ***person-person occlusions***
  - ➡ Exploit person-person ***occlusion patterns***
- Person detectors used for tracking are typically trained ***independently*** from the tracker
  - ➡ ***Train detectors with trackers in the loop***, focusing on the most common tracker failures

Contribution

- A novel structural loss-based training approach for joint person detectors
- Joint person detectors are trained in a ***tracking-aware fashion***:
  - ➡ Design occlusion patterns
  - ➡ Mine occlusion patterns from tracking results
- Surpass state-of-the-art methods on several particularly challenging datasets.

Reference

[1]. S. Tang, M. Andriluka, B. Schiele Detection and tracking of occluded people. BMVC 2012. IJCV 2013.
[2]. B. Pepik, M. Stark, P. Gehler, B. Schiele Teaching 3d geometry to deformable part models. CVPR 2012.
[3]. A. Andriyenko, K. Schindler Multi-target tracking by continuous optimization energy minimization. CVPR 2011.
[4]. M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool Online multiperson tracking-by-detection from a single uncalibrated camera. PAMI 2011
[5]. G. Shu, A. Dehghan, O. Oreifej, E. Hand, M. Shah Part- based multiple-person tracking with partial occlusion handling. CVPR 2012.
[6]. A. R. Zamir, A. Dehghan, M. Shah GMCP-Tracker: Global multi-object tracking using generalized minimum clique graphs. ECCV 2012.

Joint detection

Structural learning for joint detection

- Given training images, learning the parameters of the joint detection model is formulated as the optimization problem [2]:

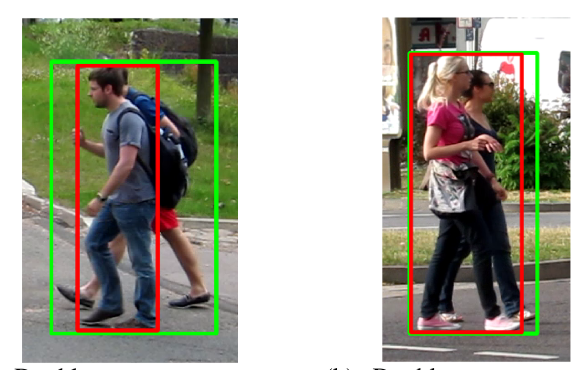
$$\min_{\beta,\xi\geq 0} \quad \frac{1}{2}\|\beta\|^2 + \frac{C}{N}\sum_{n=1}^{N}\xi_i$$
$$\text{sb.t.} \quad \max_{h}\langle\beta,\phi(I_i,y_i,h)\rangle - \max_{\hat{h}}\langle\beta,\phi(I_i,\hat{y},\hat{h})\rangle \geq \Delta(y_i,\hat{y}) - \xi_i, \qquad \forall\hat{y}\in\mathcal{Y}$$

Loss functions

- The detection with larger overlap with the ground truth bounding box has higher score than the detection with lower overlap with the ground truth bounding box

$$\Delta_{\text{VOC}}(y,\hat{y}) = \begin{cases} 0, & \text{if } y^l = \hat{y}^l = -1 \\ 1 - [y^l = \hat{y}^l]\frac{A(y^b\cap\hat{y}^b)}{A(y^b\cup\hat{y}^b)}, & \text{otherwise,} \end{cases}$$

- Teach the model to distinguish a single person and a highly occluded person pair

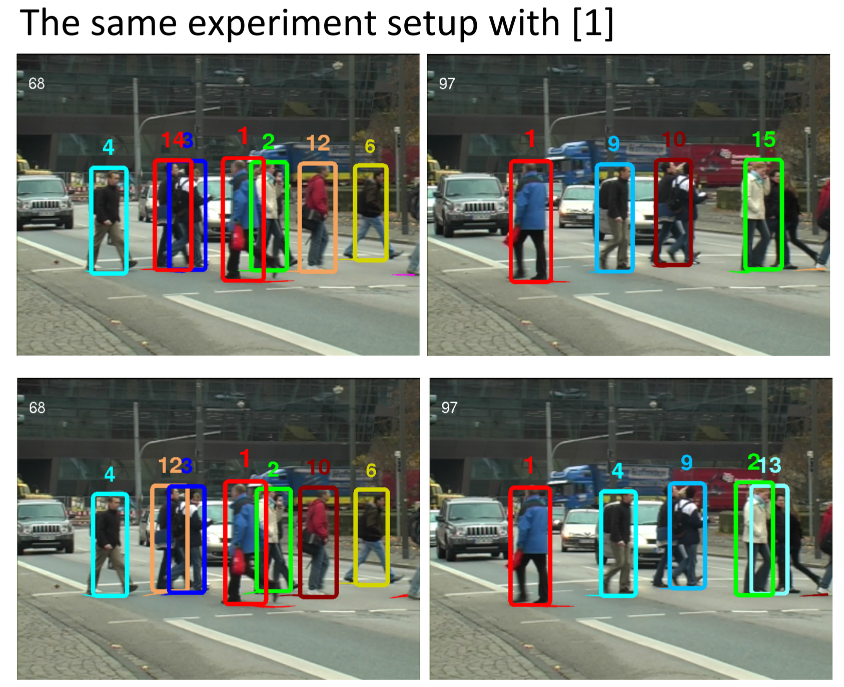


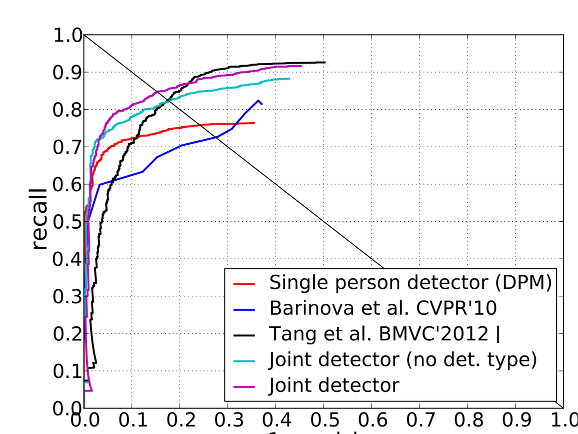
(a) Double person outscores single person with Δ_{VOC} (b) Double person outscores single person with $\Delta_{\text{VOC+DT}}$

$$\Delta_{\text{VOC+DT}}(y,\hat{y}) = (1-\alpha)\Delta_{\text{VOC}}(y,\hat{y}) + \alpha\left[y^{dt} \neq \hat{y}^{dt}\right]$$

Experimental result

The same experiment setup with [1]



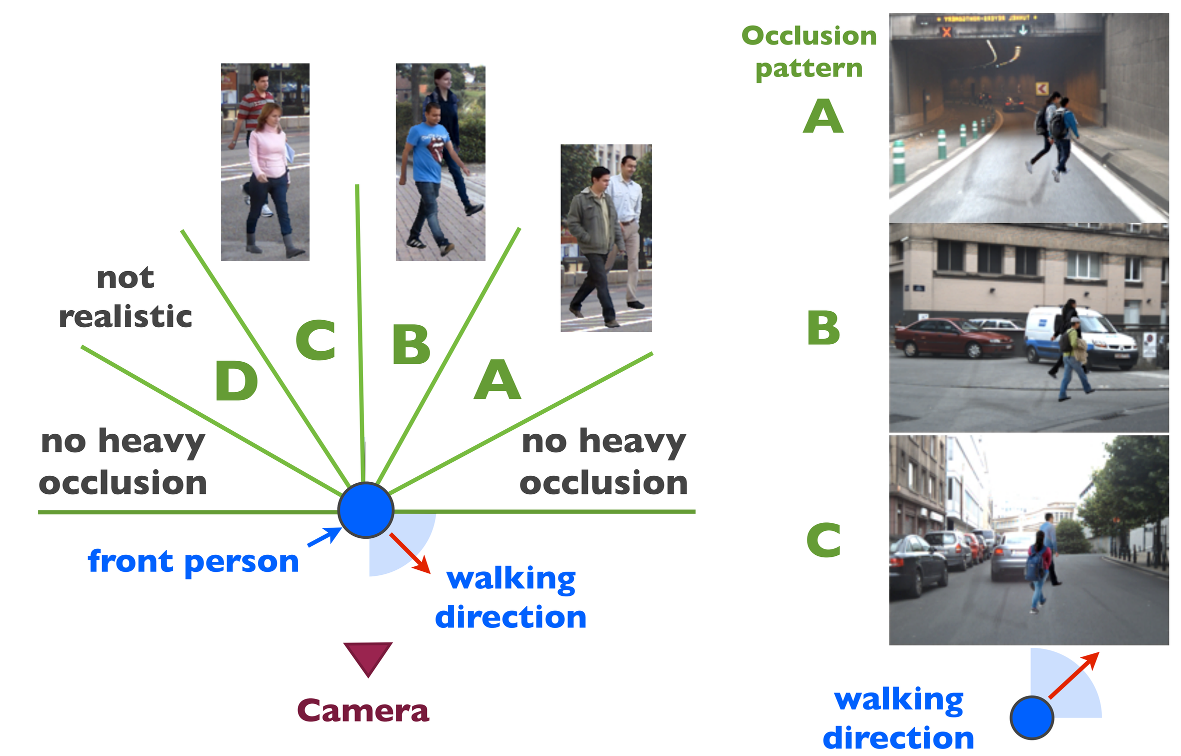


Method	Rcll	Prcsn	MOTA	MOTP
single (DPM)	78.0	94.1	72.1 %	78.5 %
Tang et al. [1]	79.9	96.5	75.6 %	79.1 %
Joint det. (no det. type)	81.9	93.2	75.1 %	79.1 %
Joint detector	82.7	93.9	76.0 %	78.6 %

Learn people detectors for tracking

Design occlusion patterns

- Manually design regular occlusion combinations that appear frequently due to long-term occlusions and therefore most relevant for tracking

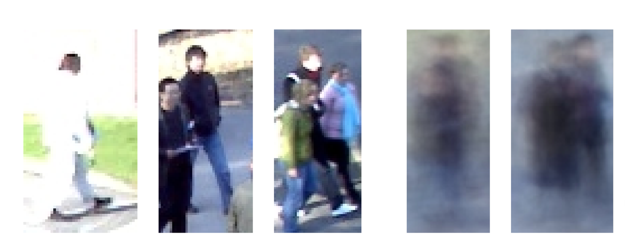


Mine occlusion patterns from tracking

Algorithm 1 Joint detector learning for tracking

Input:
Baseline detector
Multi-target tracker
Synthetic training image pool
Mining sequence
Output:
Joint detector optimized for multi-target tracking

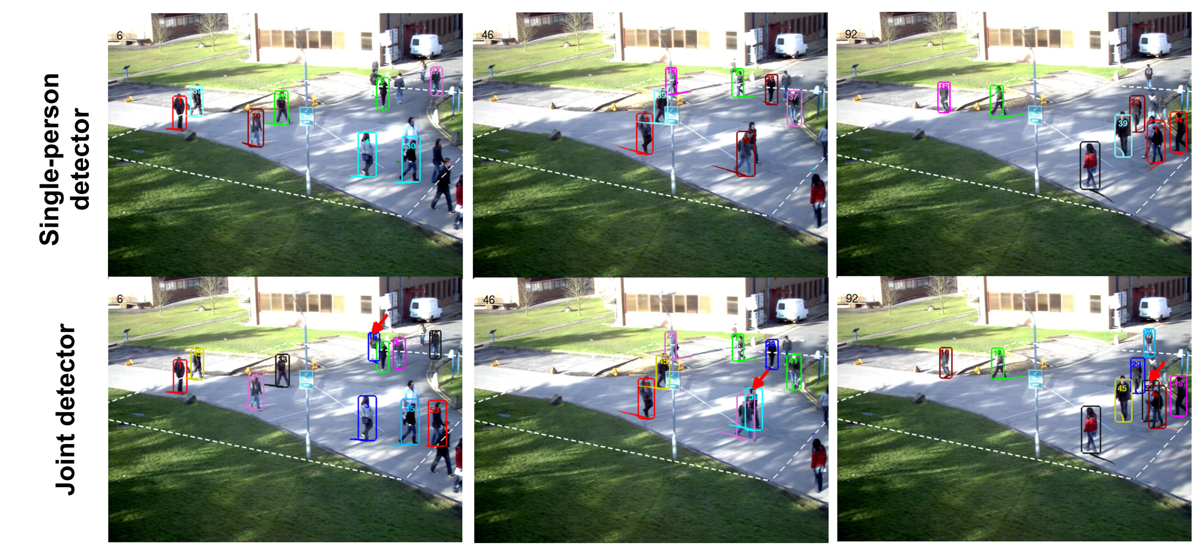
run baseline detector on *mining sequence*
run target tracker on *mining sequence*, based on the detection result from baseline detector
repeat
  collect *missing recall* from the tracking result
  cluster *occlusion patterns*
  generate *training images* for mined patterns
  train a joint detector with *new training images*
  run the joint detector on *mining sequence*
  run the target tracker on *mining sequence*
until tracking results converge



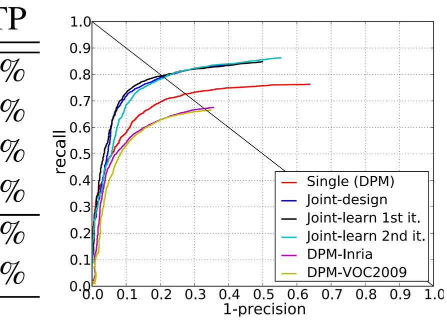
(a) (b) (c) (d) (e)

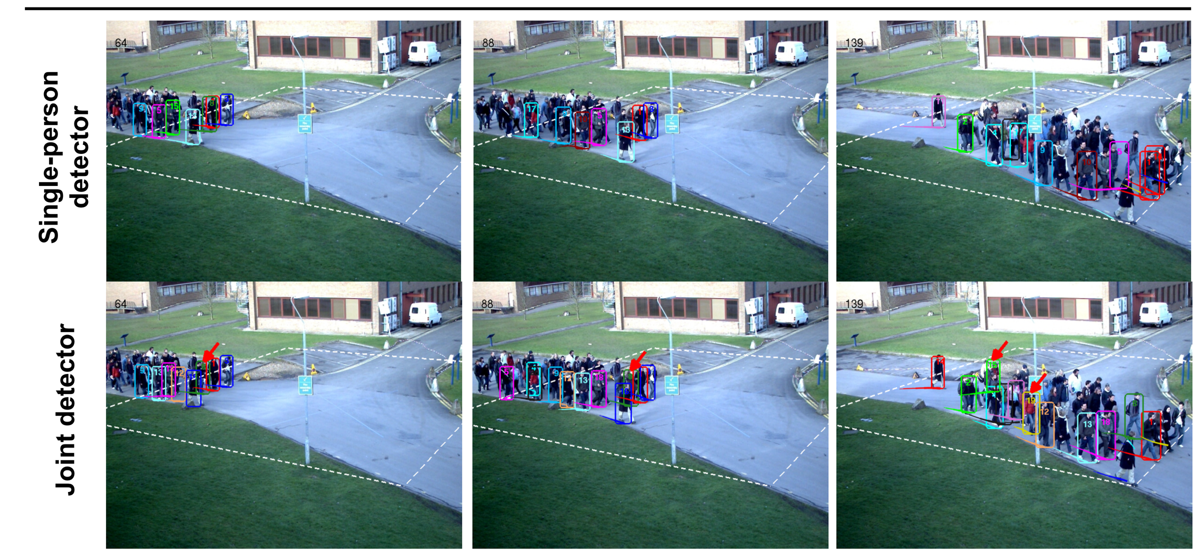
Missed targets from PETS S2.L2 mining sequence and mined occlusion patterns: (a) No person nearby; (b) interfered by one person; (c) interfered by more persons; (d) mined occlusion pattern – 1st iteration; (e) mined occlusion pattern – 2nd iteration.

Experiments

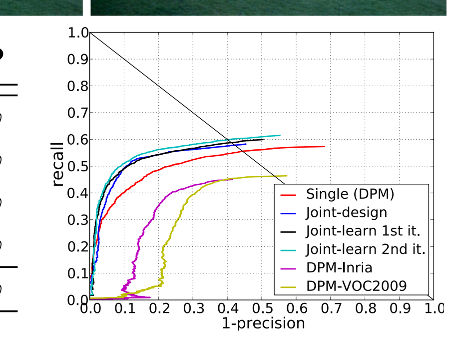


Method	Rcll	Prcsn	MOTA	MOTP
Single (DPM)	60.8	83.8	47.5 %	73.5 %
Joint-Design	65.0	91.7	57.6 %	75.2 %
Joint-Learn 1st	60.6	95.0	56.5 %	75.7 %
Joint-Learn 2nd	64.0	91.7	56.9 %	74.4 %
HOG [3]	51.0	95.5	47.8 %	73.2 %
Particle filter [4]	-	-	50.0 %	51.3 %





Method	Rcll	Prcsn	MOTA	MOTP
Single (DPM)	24.8	90.1	21.8 %	70.6 %
Joint-Design	28.5	86.3	23.0 %	70.8 %
Joint-Learn 1st	28.9	86.2	23.4 %	69.8 %
Joint-Learn 2nd	32.7	86.7	26.8 %	69.3 %
HOG [3]	24.2	83.8	19.1 %	69.6 %



Method	Rcll	Prcsn	MOTA	MOTP
Single (DPM)	90.5	97.7	87.9 %	77.2 %
Joint-Design	91.3	97.5	88.6 %	77.6 %
Joint-Learn 1st	91.0	98.5	89.3 %	77.7 %
Joint-Learn 2nd	91.0	98.0	88.7 %	76.9 %
Part-based [5]	81.7	91.3	79.3 %	74.1 %
GMCP [6]	95.0	94.2	89.1 %	77.5 %