Detection and Tracking of Occluded People

Siyu Tang
Mykhaylo Andriluka
Bernt Schiele

Max Planck Institute for Informatics
Saarbrücken, Germany
Motivation

• **Goal:** Detect and track *all* the people in crowded street scenes
Motivation

- **Goal:** Detect and track *all* the people in crowded street scenes
- **Main challenge:** *Occlusion*
Motivation

- **Goal:** Detect and track *all* the people in crowded street scenes
- **Main challenge:** *Occlusion*
  - Detection: significant partial occlusions
Motivation

- **Goal:** Detect and track *all* the people in crowded street scenes
- **Main challenge:** *Occlusion*
  - Detection: significant partial occlusions
  - Tracking: long-term occlusions *even for the entire sequence*
Motivation

Goal: Detect and track all the people in crowded street scenes

Main challenge: Occlusion
- Detection: significant partial occlusions
- Tracking: long-term occlusions even for the entire sequence

Our approach:
- Dominant occlusion cases: person/person occlusions
- Person/person occlusion: characteristic, explicitly used
Related work: Tracking

- Multi-people tracking in crowded street scenes
  - Tracking by detection
  - Reason across long-term occlusions
  - Require **sufficient visibility** for a certain frame

References:
- Andriyenko et al. CVPR’12
- Andriluka et al. CVPR’10
- Breitenstein et al. ICCV’09
- Huang et al. ECCV’08
Related work: Tracking

- Multi-people tracking in crowded street scenes
  - Tracking by detection
  - Reason across long-term occlusions
  - Require **sufficient visibility** for a certain frame

[Andriyenko et al. CVPR’12]  
[Andriluka et al. CVPR’10]  
[Breitenstein et al. ICCV’09]  
[Huang et al. ECCV’08]
Related work: Tracking

- **Multi-people tracking in crowded street scenes**
  - Tracking by detection
  - Reason across long-term occlusions
  - Require *sufficient visibility* for a certain frame

[Andriyenko et al. CVPR’12]
[Andriluka et al. CVPR’10]
[Breitenstein et al. ICCV’09]
[Huang et al. ECCV’08]
Related work: Tracking

- **Multi-person tracking in crowded street scenes**
  - Tracking by detection
  - Reason across long-term occlusions
  - Require *sufficient visibility* for a certain frame

[Andriyenko et al. CVPR’12]
[Andriluka et al. CVPR’10]
[Breitenstein et al. ICCV’09]
[Huang et al. ECCV’08]
Related work: Tracking

- Multi-people tracking in crowded street scenes
  - Tracking by detection
  - Reason across long-term occlusions
  - Require **sufficient visibility** for a certain frame

- [Andriyenko et al. CVPR’12]
- [Andriluka et al. CVPR’10]
- [Breitenstein et al. ICCV’09]
- [Huang et al. ECCV’08]
Related work: Tracking

• Multi-people tracking in crowded street scenes
  ▸ Tracking by detection
  ▸ Reason across long-term occlusions
  ▸ Require *sufficient visibility* for a certain frame

[Andriyenko et al. CVPR’12]
[Andriluka et al. CVPR’10]
[Breitenstein et al. ICCV’09]
[Huang et al. ECCV’08]
Related work: Tracking

- Multi-people tracking in crowded street scenes
  - Tracking by detection
  - Reason across long-term occlusions
  - Require **sufficient visibility** for a certain frame
    - [Andriyenko et al. CVPR’12]
    - [Andriluka et al. CVPR’10]
    - [Breitenstein et al. ICCV’09]
    - [Huang et al. ECCV’08]

- Occluded people detection: essential for multi-people tracking
Related work: Detection

- Multi-people detection in crowded street scenes
  - State-of-the-art people detectors
    - Robust to pose and viewpoint variations
    - Fails at strong occlusion levels

[Dollar et al. PAMI’11]
[Felzenszwalb et al. PAMI’10]
Related work: Detection

- Multi-people detection in crowded street scenes
  - State-of-the-art people detectors
    - Robust to pose and viewpoint variations
    - Fails at strong occlusion levels
    - Explicit occlusion reasoning
      - Treat occlusions as *distractions* or *nuisance*
      - Use *characteristic patterns* of person/person occlusions
      - Train a detector to detect the presence two people
Related work: Detection

- Multi-people detection in crowded street scenes

  - State-of-the-art people detectors
    - Robust to pose and viewpoint variations
    - Fails at strong occlusion levels

  - Explicit occlusion reasoning
    - Treat occlusions as *distractions* or *nuisance*

    Here

    - Use *characteristic patterns* of person/person occlusions
    - Train a detector to detect the presence of two people

- Baseline: Deformable Part Model (DPM)
Related work: Detection

- **Multi-people detection in crowded street scenes**

  - **State-of-the-art people detectors**
    - ✓ Robust to pose and viewpoint variations
      - Fails at strong occlusion levels
    - ✓ Baseline: Deformable Part Model (DPM)

  - **Explicit occlusion reasoning**
    - Treat occlusions as *distractions* or *nuisance*
      - ✓ Use *characteristic patterns* of person/person occlusions
      - ✓ Train a detector to detect the presence two people

  - Here

**How can we quantize the DPM performance for occluded people detection?**
MPII 2People dataset

- 900 person/person occlusion images
MPII 2People dataset

- 900 person/person occlusion images
- Categorized by 10 *occlusion levels*
MPII 2People dataset

- 900 person/person occlusion images
- Categorized by 10 occlusion levels
- Explicitly evaluate the detection performance for person/person occlusion cases
MPII 2People dataset

- DPM single-person detector evaluation results
MPII 2People dataset

- DPM single-person detector evaluation results
MPII 2People dataset

- DPM single-person detector evaluation results
MPII 2People dataset

- DPM single-person detector evaluation results
MPII 2People dataset

- DPM single-person detector evaluation results
MPII 2People dataset

- DPM single-person detector evaluation results

Only 50% recall!
Overview
Overview

• **Double-person detector**  
  [first contribution]
  
  ▶ Starting point: DPM  
  [Felsenszwalb et al. PAMI’10]
  
  ▶ Detect the presence of two people
  
  ▶ Predict bounding box for individual person
Overview

• Double-person detector
  [first contribution]
  » Starting point: DPM
    [Felsenszwalb et al. PAMI’10]
  » Detect the presence of two people
  » Predict bounding box for individual person
Overview

- Double-person detector
  [first contribution]
  - Starting point: DPM
    [Felsenszwalb et al. PAMI’10]
  - Detect the presence of two people
  - Predict bounding box for individual person
Overview

• **Double-person detector**
  [first contribution]
  - Starting point: DPM
    [Felsenszwalb et al. PAMI’10]
  - Detect the presence of two people
  - Predict bounding box for individual person

• **Joint-person detector**
  [second contribution]
  - *Jointly* train a single model to detect multiple people
  - Application: Multi-person tracking
Overview

- **Double-person detector**
  [first contribution]
  - Starting point: DPM
    [Felsenszwalb et al. PAMI’10]
  - Detect the presence of two people
  - Predict bounding box for individual person

- **Joint-person detector**
  [second contribution]
  - *Jointly* train a single model to detect multiple people
  - Application: Multi-person tracking
Overview

• **Double-person detector**
  [first contribution]
  ➔ Starting point: DPM
  [Felsenszwalb et al. PAMI’10]
  ➔ Detect the presence of two people
  ➔ Predict bounding box for individual person

• **Joint-person detector**
  [second contribution]
  ➔ *Jointly* train a single model to detect multiple people
  ➔ Application: Multi-person tracking
Overview

- **Double-person detector**
  [first contribution]
  - Starting point: DPM
  - Detect the presence of two people
  - Predict bounding box for individual person

- **Joint-person detector**
  [second contribution]
  - Jointly train a single model to detect multiple people
  - Application: Multi-person tracking
Single-person detector (DPM)
[Felsenszwalb et al. PAMI’10]

- **Mixture of components:** \( C = (C_1, C_2, \ldots, C_M) \)

- **Initialization:**

- **Training:**

- **Bounding box prediction:**
Single-person detector (DPM)
[Felsenszwalb et al. PAMI’10]

- **Mixture of components:** \( C = (C_1, C_2, \ldots, C_M) \)

  ![Images of templates and deformed templates]

- **Initialization:**

- **Training:**

- **Bounding box prediction:**
Single-person detector (DPM)
[Felsenszwalb et al. PAMI’10]

- **Mixture of components:** \(C = (C_1, C_2, \ldots, C_M)\)

- **Initialization:**
  - Bounding box aspect ratio

- **Training:**
  - Latent SVM + hard-negative mining

- **Bounding box prediction:**
Single-person detector (DPM)
[Felsenszwalb et al. PAMI’10]

- **Mixture of components:** \( C = (C_1, C_2, \ldots, C_M) \)

- **Initialization:**
  - Bounding box aspect ratio

- **Training:**
  - Latent SVM + hard-negative mining

- **Bounding box prediction:**
  - Linear regression function \( \times M \)

---

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Double-person detector

- **Mixture of components:** \( C = (C_1, C_2, \ldots, C_M) \)

- **Initialization:**

- **Training:**

- **Bounding box prediction:**
Double-person detector

- **Mixture of components:** \[ C = (C_1, C_2, \ldots, C_M) \]

- **Initialization:**
  - *Occlusion levels:* 0%~25%, 25%~55%, 55%~85%

- **Training:**

- **Bounding box prediction:**
Double-person detector

- **Mixture of components:** $C = (C_1, C_2, \ldots, C_M)$

- **Initialization:**
  - **Occlusion levels:** 0%~25%, 25%~55%, 55%~85%

- **Training:** Standard DPM training approach with large number of training images

- **Bounding box prediction:**
Double-person detector

- **Mixture of components:** \( C = (C_1, C_2, \ldots, C_M) \)

- **Initialization:** *Occlusion levels:* 0%–25%, 25%–55%, 55%–85%

- **Training:** *Standard DPM training approach with large number of training images*

- **Bounding box prediction:** *Linear regression function* \( \times M \times 2 \)
Synthetic image generation

- Double-person appearance variation
  - Large number of training images
- Various backgrounds, relative positions and scales
- Occlusion level initialization
  - Accurate occlusion level estimation
Synthetic image generation

- Double-person appearance variation
  - Large number of training images
- Various backgrounds, relative positions and scales
- Occlusion level initialization
  - Accurate occlusion level estimation

→ Synthetic Images
Synthetic image generation

- Double-person appearance variation
  - Large number of training images
- Various backgrounds, relative positions and scales
- Occlusion level initialization
  - Accurate occlusion level estimation

**Synthetic Images**

Annotated foreground person map  →  Foreground person  →  Generated synthetic training image

Annotated background person map  →  Background person
Synthetic image generation

- Double-person appearance variation
  - Large number of training images
- Various backgrounds, relative positions and scales
- Occlusion level initialization
  - Accurate occlusion level estimation

→ Synthetic Images
Synthetic image generation

- Double-person appearance variation
  - Large number of training images
- Various backgrounds, relative positions and scales
- Occlusion level initialization
  - Accurate occlusion level estimation

\[ \text{Synthetic Images} \]

- Synthetic training images are categorized by occlusion levels

  Occlusion level: 5%~15%
  Occlusion level: 35%~45%
  Occlusion level: 75%~85%
Experiment result

- MPII 2People dataset

Single person detector

Double person detector
Experiment result

- MPII 2People dataset

Single person detector

![Images with bounding boxes and percentages]

- ~25%
- ~45%
- ~55%
- ~85%

Double person detector

![Images with bounding boxes and percentages]

- ~25%
- ~45%
- ~55%
- ~85%
Experiment result

- MPII 2People dataset

![Graph showing equal error rate vs occlusion levels for single and double person detectors.]

- Single person detector
  - ~25%
  - ~45%
  - ~55%
  - ~85%

- Double person detector
  - ~25%
  - ~45%
  - ~55%
  - ~85%

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Experiment result

- MPII 2People dataset
Experiment result

- MPII 2People dataset

Single person detector

Double person detector

Occlusion levels

Equal Error Rate

Single-person detector
Double-person detector

0.99
0.91
Experiment result

- MPII 2People dataset

**Single person detector**

- ~25%
- ~45%
- ~55%
- ~85%

**Double person detector**

- ~25%
- ~45%
- ~55%
- ~85%

**Best performance**

<table>
<thead>
<tr>
<th>Occlusion levels</th>
<th>Equal Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>~25%</td>
<td>0.99</td>
</tr>
<tr>
<td>~45%</td>
<td>0.91</td>
</tr>
<tr>
<td>~55%</td>
<td>0.85</td>
</tr>
<tr>
<td>~85%</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Overview

- **Double-person detector**  
  [first contribution]
  - Starting point: DPM  
    [Felsenszwalb et al. PAMI'10]
  - Detect the presence of two people
  - Predict bounding box for individual person

- **Joint-person detector**  
  [second contribution]
  - Jointly train a single model to detect multiple people
  - Application: Multi-person tracking
Overview

- **Double-person detector**  
  [first contribution]
  - Starting point: DPM  
    [Felsenszwalb et al. PAMI’10]
  - Detect the presence of two people
  - Predict bounding box for individual person

- **Joint-person detector**  
  [second contribution]
  - Jointly train a single model to detect multiple people
  - Application: Multi-person tracking
### Joint detector

<table>
<thead>
<tr>
<th>Single-person detector</th>
<th>Double-person detector</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>$C_1$</td>
</tr>
<tr>
<td>$C_M$</td>
<td>$C_N$</td>
</tr>
</tbody>
</table>

- **Root template**
- **Part template**
- **Deformations**
- **Example**

**Example**

![Example](image_url)
Joint detector

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Joint detector

\[ C_1 \]

\[ C_M \]

\[ C_1 \]

\[ C_N \]
Joint detector

- Root template
- Part template
- Deformations
- Example

\( C_1 \)

\( C_M \)

\( C_1 \)

\( C_N \)
Joint detector

\[ C_1 \]

\[ C_M \]

\[ C_N \]
Joint detector

\[ C_1 \]

\[ C_M \]

\[ C_1 \]

\[ C_N \]
Joint detector

\[ C_1 \]
\[ C_M \]
\[ C_1 \]
\[ C_N \]
Joint detector

$C_1$

$C_M$

$C_1$

$C_N$
Joint detector

\[ C_1 \]

\[ C_M \]

\[ C_N \]
Joint detector

- **Mixture of components:**

\[
C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N)
\]
Joint detector

- **Mixture of components:**
  \[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

- **Initialization:**
  *Bounding box aspect ratio + occlusion level*

---

**Single-person components**
- Root template
- Example

**Double-person components**
- \( C_1 \)
- \( C_M \)
- \( C_{M+1} \)
- \( C_{M+N} \)
Joint detector

• **Mixture of components:**

\[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

• **Initialization:**

*Bounding box aspect ratio + occlusion level*

• **Joint Training:**

*Latent SVM + hard-negative mining*
Joint detector

- **Mixture of components:**
  \[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

- **Initialization:**
  *Bounding box aspect ratio* + *occlusion level*

- **Joint Training:**
  *Latent SVM* + *hard-negative mining*

- **Bounding box prediction:**
  *Component-dependent linear regression functions*: \( M + 2N \)

---

**Single-person components**
- **root template**
- **example**

- \( C_1 \)
- \( C_M \)

**Double-person components**
- \( C_{M+1} \)
- \( C_{M+N} \)
Joint detector

• **Mixture of components:**
  \[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

• **Initialization:**
  *Bounding box aspect ratio + occlusion level*

• **Joint Training:**
  *Latent SVM + hard-negative mining*

• **Bounding box prediction:**
  *Component-dependent linear regression functions: M + 2N*

• **Two-level non-maximum suppression:**
Joint detector

- **Mixture of components:**
  \[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

- **Initialization:**
  *Bounding box aspect ratio + occlusion level*

- **Joint Training:**
  *Latent SVM + hard-negative mining*

- **Bounding box prediction:**
  *Component-dependent linear regression functions: M + 2N*

- **Two-level non-maximum suppression:**
  
  
  **Bounding box prediction**
Joint detector

- **Mixture of components:**
  \[ C = (C_1, C_2, \ldots, C_M, C_{M+1}, \ldots, C_{N-1}C_N) \]

- **Initialization:**
  *Bounding box aspect ratio + occlusion level*

- **Joint Training:**
  *Latent SVM + hard-negative mining*

- **Bounding box prediction:**
  *Component-dependent linear regression functions: M + 2N*

- **Two-level non-maximum suppression:**
Two-level Non-Maximum Suppression
Two-level Non-Maximum Suppression

✓ Single person detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection
Two-level Non-Maximum Suppression

- Single person detection
- Person pair detection
- Three and more people detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Two-level Non-Maximum Suppression

- Single person detection ✓
- Person pair detection ✓
- Three and more people detection ?

Double-person Component

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
?
Three and more people detection

Double-person Component

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection

Double-person Component

[Image of a street scene with pedestrians and highlighted boxes indicating detection areas]
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection

Double-person Component

Suppressed
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
? Three and more people detection
Two-level Non-Maximum Suppression

✓ Single person detection
✓ Person pair detection
?
 Three and more people detection
Experiments

- **TUD-pedestrians dataset**
  - 250 images with 311 fully visible people

- Joint person detector
- Single person detector

![Graph showing performance metrics for joint and single person detectors.]

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Experiments

- **TUD-pedestrians dataset**
  - 250 images with 311 fully visible people

Joint person detector

Single person detector

![Graph](image)

- **Joint-person detector**
- **Single-person detector**
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

![Graph showing recall vs. 1-precision for Single-person detector]
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

![Image of people crossing the street](image)

![Graph showing recall vs. 1-precision](graph)

**Single-person detector**

**Double-person detector**

**Barinova et al. CVPR’10**
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

Joint person detector

[Barinova et al. CVPR’10]

![Graph showing performance metrics for different detection methods.](image)
Experiments

• **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

Joint person detector

[Barinova et al. CVPR’10]

Single-person detector
Double-person detector
Barinova et al. CVPR’10
Joint-person detector
Experiments

- **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

Joint person detector

[Barinova et al. CVPR’10]
Experiments

• **TUD-crossing dataset**
  - 201 images with 1008 annotated people
  - Frequently partial or even fully occlusion

Joint person detector

[Barinova et al. CVPR’10]

S. Tang, M. Andriluka, B. Schiele | Detection and Tracking of Occluded People | BMVC 2012
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

| Single person detector | Joint person detector |
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

![Single person detector](image1.jpg) ![Joint person detector](image2.jpg)
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

![Single person detector](image1)

![Joint person detector](image2)
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

Single person detector | Joint person detector
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

![Single person detector](image1.png) ![Joint person detector](image2.png)
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

Single person detector

Joint person detector
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.
Result: Multi-person tracking

- Multi-person tracking by detection [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

Single person detector  Joint person detector
Result: Multi-person tracking

- **Multi-person tracking by detection** [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

![Single person detector](image1)

![Joint person detector](image2)
Result: Multi-person tracking

- Multi-person tracking by detection [Andriluka et al. CVPR’10]
  - Compare the performance of single-person and joint detector in the context of multiple people tracking.

Single person detector

Joint person detector
Conclusions

• Double-person detector
  [first contribution]
  ▸ Characteristic person/person occlusion patterns
  ▸ Detection based approach to occlusion handling
  ▸ Outperform DPM detector by a large margin on MPII 2people dataset
Conclusions

- **Double-person detector**  
  [first contribution]
  - Characteristic person/person occlusion patterns
  - Detection based approach to occlusion handling
  - Outperform DPM detector by a large margin on MPII 2people dataset

- **Joint-person detector**  
  [second contribution]
  - *Jointly* train a single model to detect multiple people
  - Best performance on TUD-Crossing benchmark
Thank you for your attention!

Acknowledgement:
Thank Bojan Pepik for the code and helpful discussions on DPM.