In Good Shape: Robust People Detection based on Appearance and Shape

L. Pishchulin, A. Jain, C. Wojek, T. Thormählen and B. Schiele

Max Planck Institute for Informatics, Saarbrücken

30.08.11
Conventional People Detection

- State-of-the-art approaches
  - Use supervised learning methods
  - Require much training data
  - Focus on features and learning

⇒ This work changes data distribution
Conventional People Detection

- State-of-the-art approaches
  - Use supervised learning methods
  - Require much training data
  - Focus on features and learning

⇒ This work changes data distribution
Conventional People Detection

- State-of-the-art approaches
  - Use supervised learning methods
  - Require much training data
  - Focus on features and learning

⇒ This work changes data distribution
Changing Data Distribution

- Real distribution vs. “better-to-train-from” distribution

- Typical approach
  - collect more training data and hope it improves the coverage of people class variability

- Data collection
  - tedious
  - ill-defined process
    - did we capture people class variability any better?
Our Method

3D shape model

⇒ Enrich training data with complementary shape
Our Method

3D shape model → vary shape/pose

⇒ Enrich training data with complementary shape
Our Method

⇒ Enrich training data with complementary shape
Our Method

⇒ Enrich training data with complementary shape
Our Method

3D shape model → vary shape/pose → render edges → add background → training data

⇒ Enrich training data with complementary shape
Our Method

⇒ **Enrich training data with complementary shape**
Training on Synthetic Data

- **Multi-view** pose estimation
e.g. [Grauman et al., ICCV’03]

- **Depth sensor**-based pose estimation
  [Shotton et al., CVPR’11]

- **Monocular** pose estimation/
  people detection [Marin et al., CVPR’10;
Pishchulin et al., CVPR’11]

⇒ This work:
  non-photorealistic rendering for monocular people detection
3D Human Shape Model

- Proposed by [Hasler et al., Eurographics’2008]
- Learn shape from 3D laser scans of humans
- Represent shape and pose variations
Proposed Approach
Pipeline

3D shape model \rightarrow \text{vary shape/pose} \rightarrow \text{render edges} \rightarrow \text{add background} \rightarrow \text{features} \rightarrow \text{learning}

training data
Pipeline

3D shape model → vary shape/pose → render edges → add background → features → learning

training data
Shape Changes

- Sample shape parameters of semantic attributes
  - height, weight, leg lengths, waist, hips, breast

⇒ Cover major 3D shape variations
Shape Changes

- Sample shape parameters of semantic attributes
  - height, weight, leg lengths, waist, hips, breast

⇒ Cover major 3D shape variations
Pose Changes

- Sample 3D joint angles from walking poses

⇒ High degree of articulation
Pose Changes

- Sample 3D joint angles from walking poses

⇒ High degree of articulation
Rendering

- Render 3D body into a non-photorealistic edge image

⇒ Directly change the distribution of edge-based features
Rendering

- Render 3D body into a non-photorealistic edge image

⇒ Directly change the distribution of edge-based features
Adding Background

- Sample from the set of people-free images
- Combine background edges with rendered edges
Results
People Detector

- Pictorial Structures Model [Andriluka et al., CVPR 2009]

Body is represented as flexible configuration of body parts posterior over body poses

\[ p(L|D) \propto p(D|L)p(L) \]

likelihood of observations prior on body poses
Datasets

- **Training**
  - *Rendered data*: 15000 non-photorealistically rendered images
  - *Real data*: $\sim$ 3000 images [Andriluka et al., CVPR 2010]
  - *Joint data*: Real data + rendered data

- **Testing**
  - monocular images of 309 people [Andriluka et al., CVPR 2010]
Rendered Data: Shape Parameters

- ±3 sigma performs better as data covers most variation
- Uniform sampling outperforms Gauss-sampling
  - tails of shape distribution get more weight
Rendered Data: Shape Parameters

- ±3 sigma performs better as data covers most variation
- Uniform sampling outperforms Gauss-sampling
  - tails of shape distribution get more weight
- ±3 sigma performs better as data covers most variation
- Uniform sampling outperforms Gauss-sampling
  - tails of shape distribution get more weight
Training Strategies

- Rendered data suffers from missing clothes and internal edges

- Joint data outperforms Real data alone
  ⇒ synthetic samples increase the variability of training data
Training Strategies

- Rendered data suffers from missing clothes and internal edges
- Joint data outperforms Real data alone
  \[\Rightarrow\] synthetic samples increase the variability of training data
Training Strategies

- Rendered data suffers from missing clothes and internal edges
- *Joint* data outperforms *Real* data alone
  ⇒ synthetic samples increase the variability of training data
Training Strategies

- Rendered data suffers from missing clothes and internal edges

- *Joint* data outperforms *Real* data alone
  ⇒ synthetic samples increase the variability of training data
Combining Detectors

- Combine detectors by SVM

- Combination improves the results
  ⇒ complementarity of Joint data
Combining Detectors

- Combine detectors by SVM

![Graph showing precision-recall curves]

- Combination improves the results
  \( \Rightarrow \) complementarity of Joint data
Combining Detectors

- Combine detectors by SVM

Combination improves the results

⇒ complementarity of Joint data
Discussion

- Andrilukas detector fails due to partial edge evidence
- \textit{Joint} detector focuses on external edges

⇒ Joint data contains more information than real data
Discussion

- Andrilukas detector fails due to partial edge evidence
- *Joint* detector focuses on external edges

⇒ Joint data contains more information than real data
Discussion

- Andrilukas detector fails due to partial edge evidence
- *Joint* detector focuses on external edges

⇒ Joint data contains more information than real data
Discussion

- Andrilukas detector fails due to partial edge evidence
- Joint detector focuses on external edges

⇒ Joint data contains more information than real data
Conclusion

- Leveraged 3D human shape model

- Combined rendered edge data with image data

- Rendered data is complementary to real data
Thank you for your attention!