Learning People Detection Models from Few Training Samples

Leonid Pishchulin, Arjun Jain, Christian Wojek, Mykhaylo Andriluka
Thorsten Thormälen and Bernt Schiele
Max Planck Institute for Informatics, Saarbrücken
{leonid, ajain, cwojek, andriluk, thormae, schiele}@mpi-inf.mpg.de

max planck institut informatik

MAX-PLANCK-GESELLSCHAFT

Goal

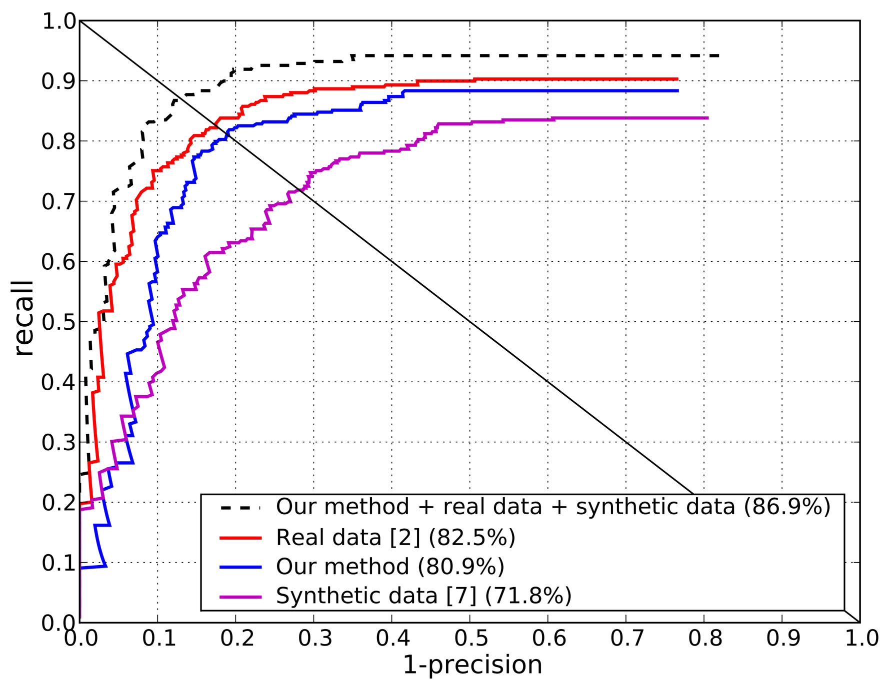
- Propose a novel technique to train people detectors from only a few observed training subjects
- Push the performance of current detection systems trained on hundreds of manually annotated pedestrians
- Approach the lack-of-training-data problem by automatically generating realistic training samples

Contributions

- Explore the applicability of state-of-the-art 3D human model to learn people detectors
- Compare the results to prior work (e.g. [2, 7])

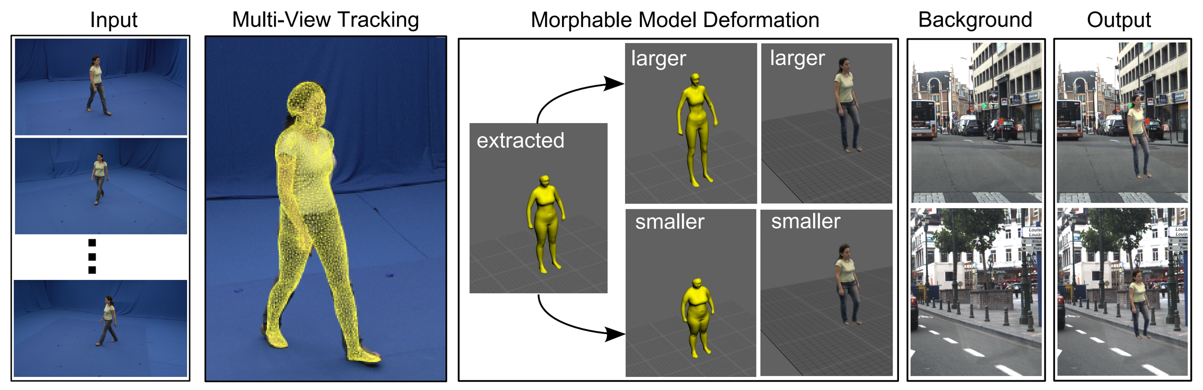
- Analyze various combinations of synthetic and real training data

$\Rightarrow$ outperform current methods which use real training data only



Proposed Approach

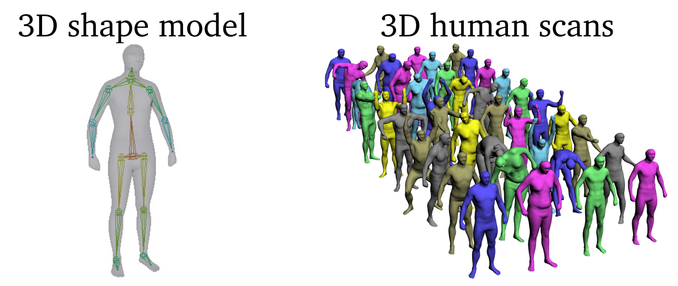
1. Generate realistic synthetic data by MovieReshape [6]
2. Combine reshaped humans with backgrounds
3. Automatically obtain 2D part annotations from known 3D joint positions



$\Rightarrow$ **Realistic distributions of human appearance and shape**

Statistical 3D human shape model [5]

- Learn shape from 3D laser scans of humans
- Represent shape variations via PCA
- Embed kinematic skeleton with linear blend skinning



Automatic model fitting

- Fit the parameters of 3D body model to silhouettes
  $\Rightarrow$ particle filter-based estimator

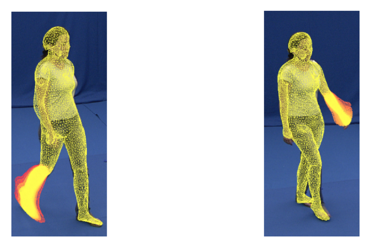
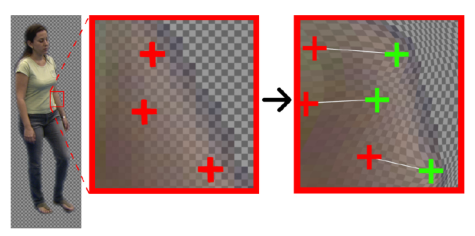


Image deformation

- Sample 3D shape parameters $\pm 3\sigma$ from the mean shape
- Use 3D offset vectors to drive 2D image warping



Composition with background

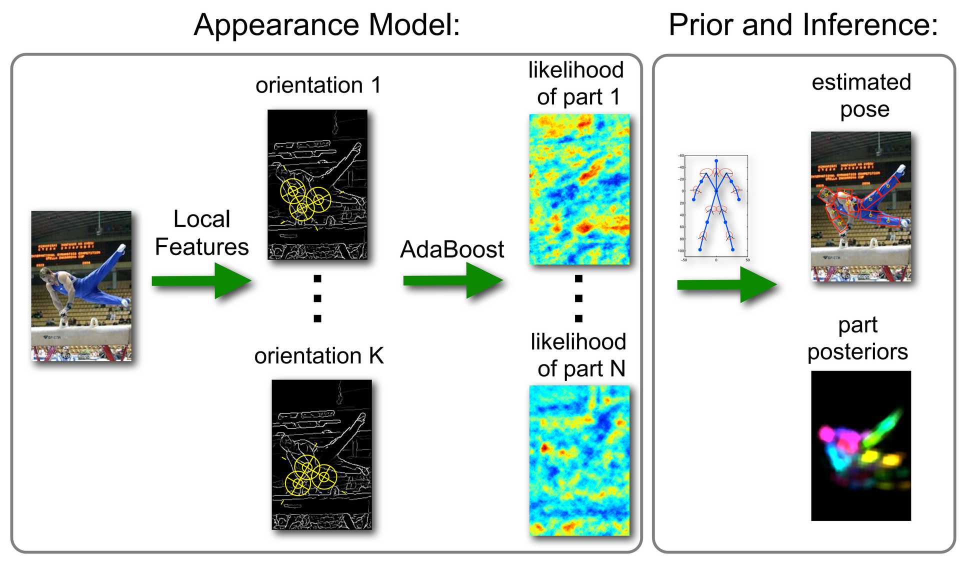
- Adjust color distribution of pedestrian w.r.t. background

Sample output images with gradual height changes

People Detection Models

Pictorial structures (PS) [1, 4]

- Flexible configuration of body parts with pose prior
- AdaBoost part detectors learned from dense shape context descriptor
- Inference by sum-product belief propagation

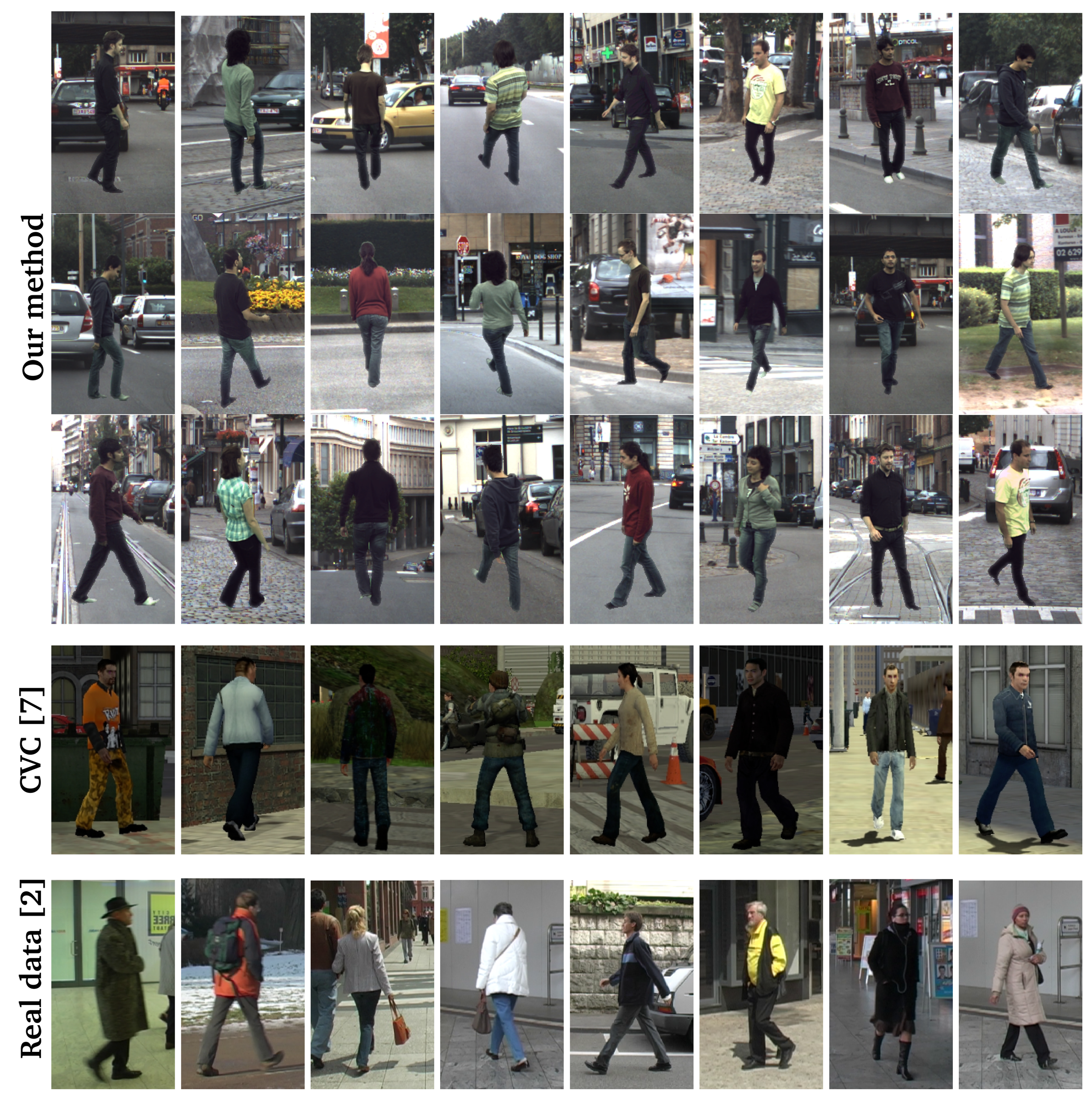


Histograms of oriented gradients (HOG) [3]

- Sliding window detection
- Monolithic template based on HOG features
- Histogram intersection kernel SVM

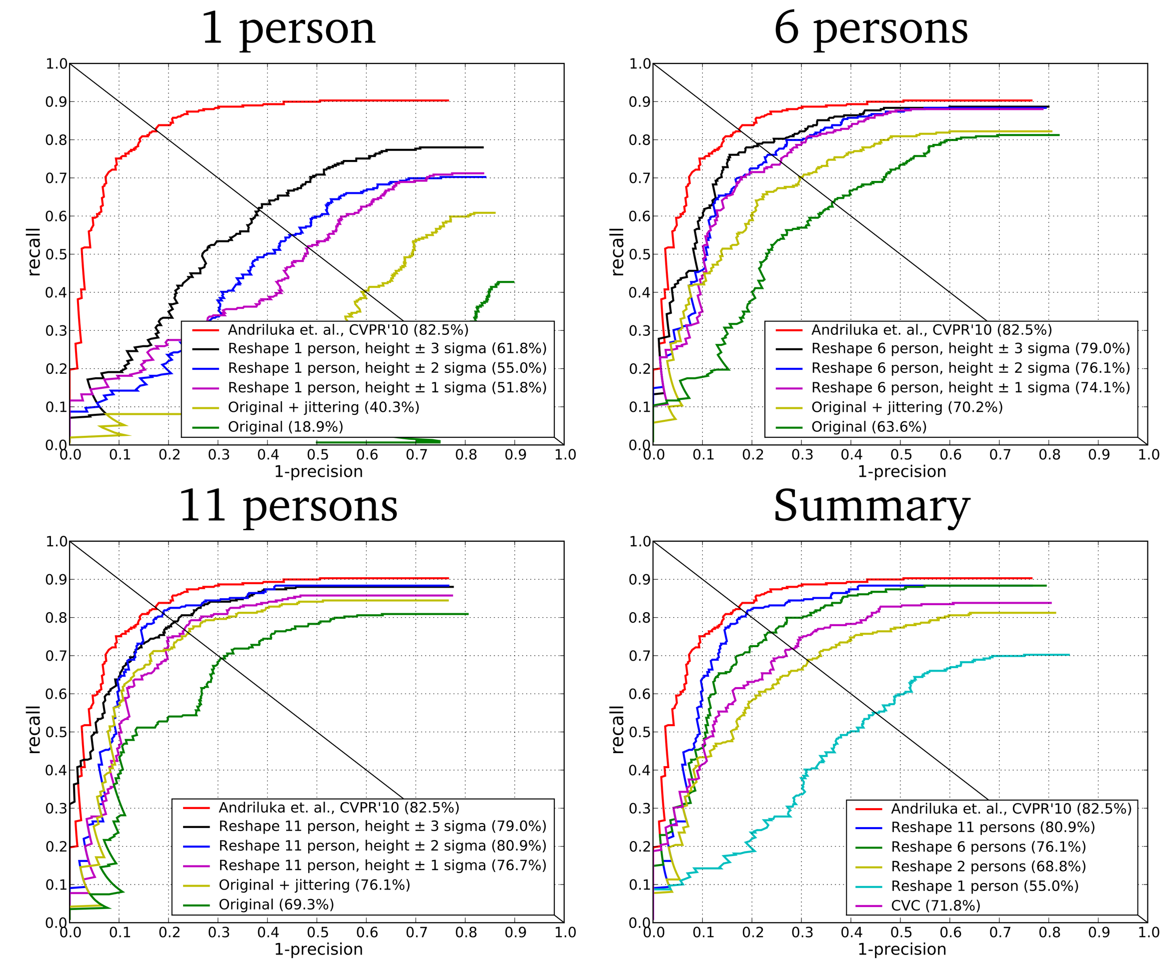
Datasets

- *Reshape data (our method)*: 11 persons, ~ 2000 reshaped images per person
- *CVC (virtual pedestrians) [7]*: 3432 images total
- *Multi-viewpoint real data [2]*: 2972 train images, 248 test and 248 validation images

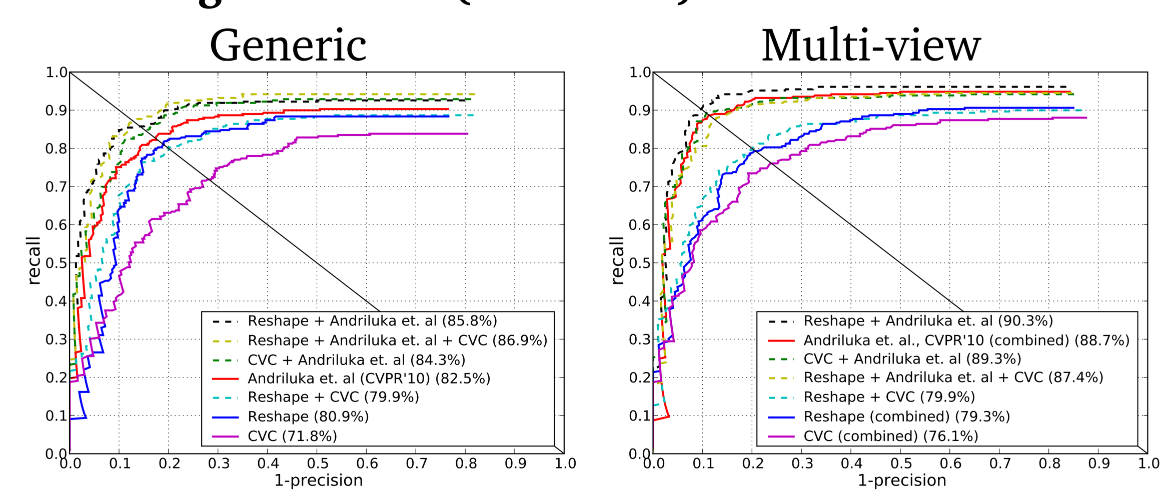


Results

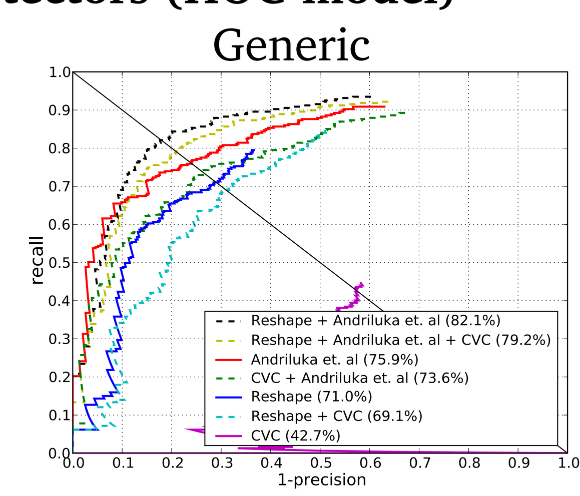
Using Reshape data (PS model)



Combining detectors (PS model)



Combining detectors (HOG model)



References

[1] M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *CVPR*, pages 1014–1021, 2009.

[2] M. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estimation and tracking by detection. In *CVPR*, pages 623–630, 2010.

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

[4] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61:55–79, 2005.

[5] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. *CGF (Proc. Eurographics 2008)*, 2(28), 2009.

[6] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt. Moviereshape: Tracking and reshaping of humans in videos. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 29(5), 2010.

[7] J. Marin, D. Vazquez, D. Geronimo, and A. Lopez. Learning appearance in virtual scenarios for pedestrian detection. In *CVPR*, pages 137–144, 2010.