Deep 3D Human Pose Estimation under Partial Body Presence

Saeid Vosoughi and Maria A. Amer

Electrical and Computer Engineering Department
Concordia University
Montreal, Quebec

IEEE International Conference on Image Processing, October 2018, Athens, Greece

Agenda

- 3D human pose estimation under partial presence
- Our method: Network architecture
- Our method: Experimental setup and data preparation
- Results: Objective evaluation
- Results: Subjective evaluation
- Conclusion

3D Human Pose Estimation under Partial Presence

3D human pose: Body's main joints' positions in the 3D space

$\boldsymbol{S} = \tau\{\boldsymbol{I}\};$

$\boldsymbol{S} \in \mathbb{R}^{\boldsymbol{3}\times\boldsymbol{j}}$: Estimated human body pose in the 3D space

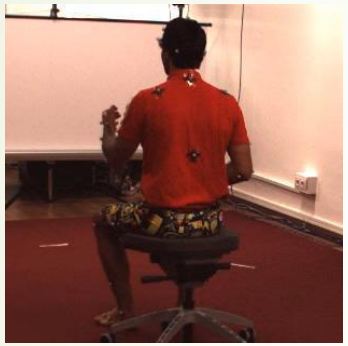
$\boldsymbol{j}$: Number of main body joints (=17 in this paper)

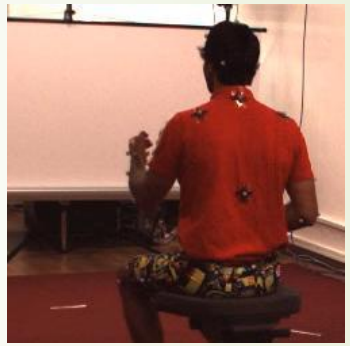
τ : Transformation from the 2D imagery to 3D human poses

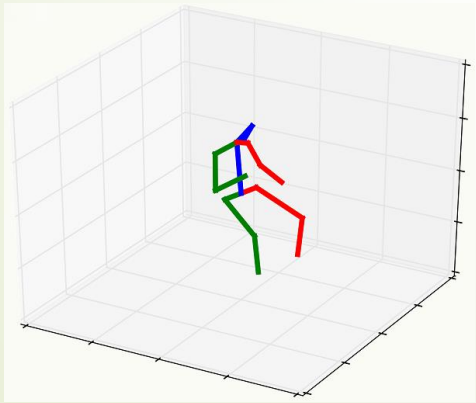
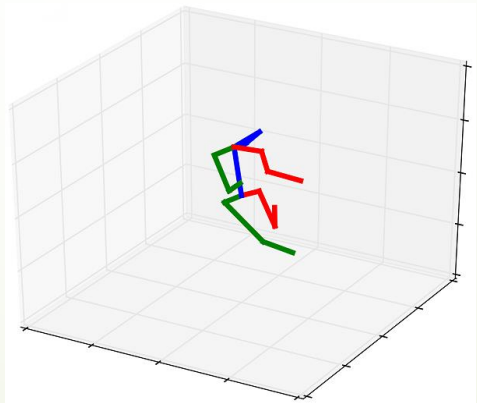
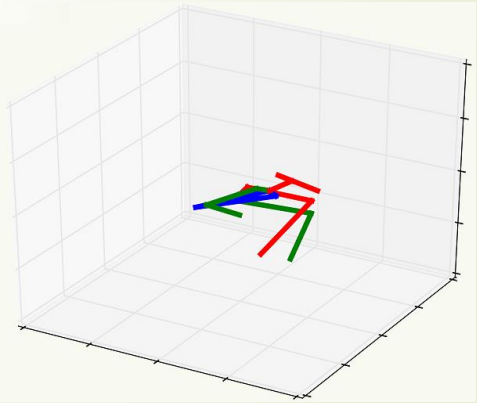
$\boldsymbol{I}$: Digital intensity image

3D Human Pose Estimation under Partial Presence

Partial body presence: Missing some parts of the body





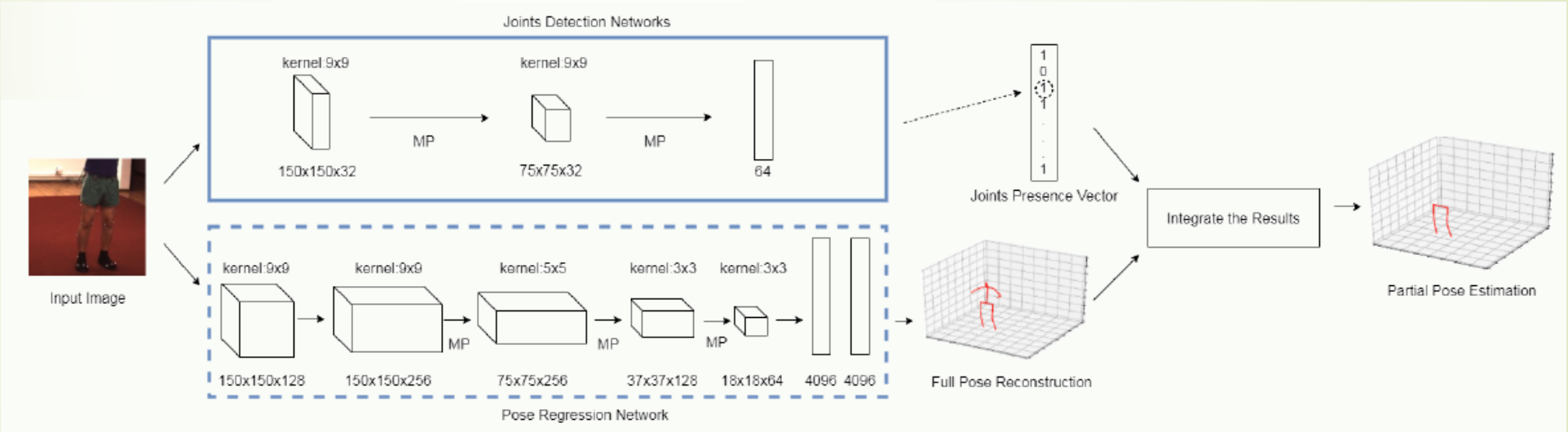
3D Human Pose Estimation under Partial Presence

imperfect segmentation

zoomed-in photography

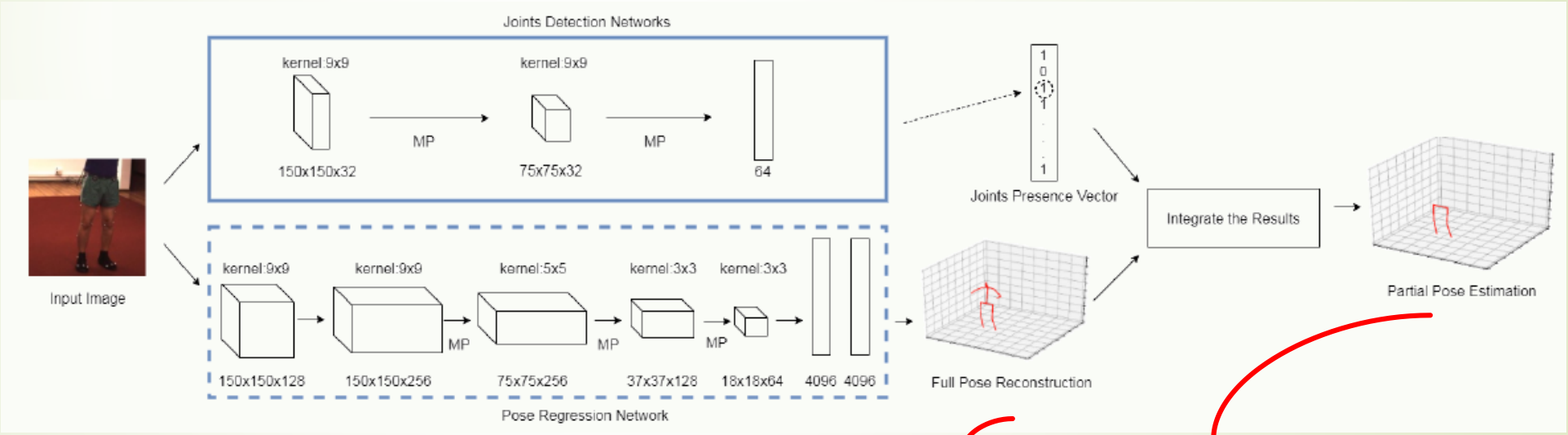
Our method: Network architecture

Deep Convolutional Neural Network: 1) Joints Detection 2) Pose Regression



Our method: Network architecture

Deep Convolutional Neural Network: 1) Joints Detection 2) Pose Regression



Outputs: 1. **Full reconstruction**

2. **Partial reconstruction**

Our method: Experimental setup and data preparation

Experimental Setup:

- Adam optimization
- Batch size 32
- Human3.6M dataset
- Downsampled by a factor of 5

- Detection
- Learning rate 0.001
- Loss function: Cross-Entropy

Regression
- Learning rate 0.0001
- Loss function: Mean Square Error

Our method: Experimental setup and data preparation

Data Preparation:

- Random window selection
- Uniform distribution
- Covering more than one quarter of the subject region
- Spanned over the four quarters

Random Window Selection:

Results: Objective evaluation

Compared to:

1. **VNect:** Mehta, Dushyant, et al. "Vnect: Real-time 3d human pose estimation with a single rgb camera." ACM Transactions on Graphics (TOG) 36.4 (2017): 44.

2. **InWild:** Zhou, Xingyi, et al. "Towards 3d human pose estimation in the wild: a weakly-supervised approach." IEEE International Conference on Computer Vision. 2017.

Results: Objective evaluation

Method	Direction	Discussion	Eating	Average
Vnect	286.64	329.20	350.67	338.01
InWild	300.00	329.96	338.91	332.48
Ours	**143.5**	**180.25**	**144.72**	**173.6**

mean-per-joint-error of **present** joints

$$L(y_{gt}, y_{est}) = \frac{1}{N}\sum_{j=1}^{N} ||C_j^{y_{gt}} - C_j^{y_{est}}||_2;$$

y_{gt}: ground truth joints′ position matrix
y_{est}: estimated joints' position matrix
N: number of the joints
$C_j^{y_{gt}}$: the vector of the j^{th} column of the matrix y

Results: Objective evaluation

Method	Direction	Discussion	Eating	Average
Vnect	370.57	387.86	403.85	396.44
InWild	392.1	394.96	405.00	400.50
Ours	**156.02**	**190.71**	**157.71**	**184.94**

mean-per-joint-error of **full** body recovery

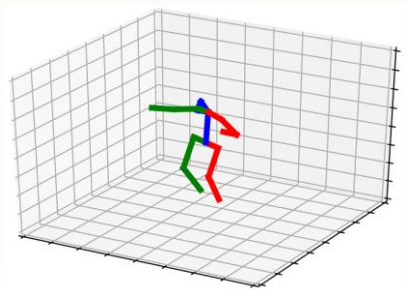
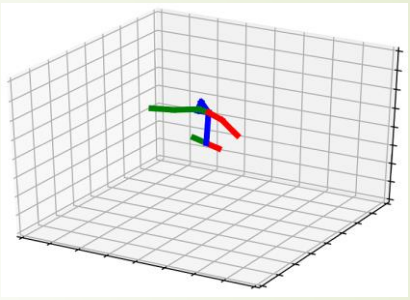
Results: Objective evaluation

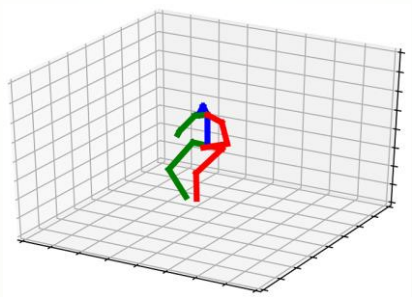
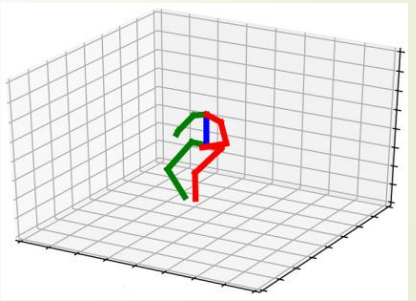
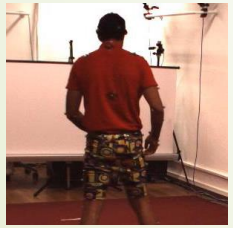
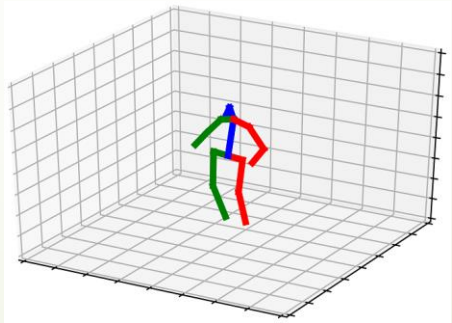
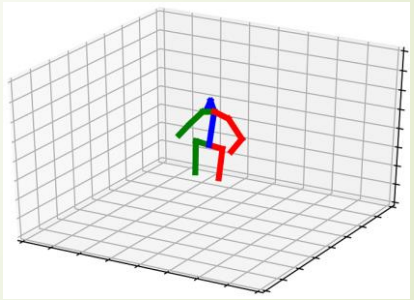
Joint	Pelvis	Left Ankle	Right Shoulder	Average
Accuracy (%)	92.63	86.70	90.41	88.00

performance of the **detection** stage based on binary accuracy

Results: Subjective evaluation

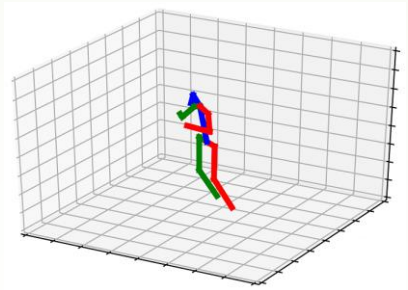
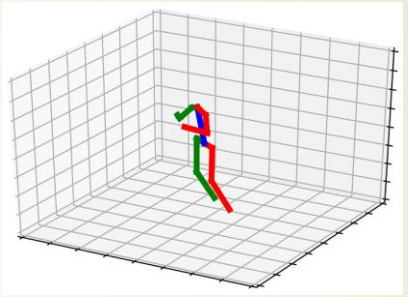
Subjective results under partial body presence:

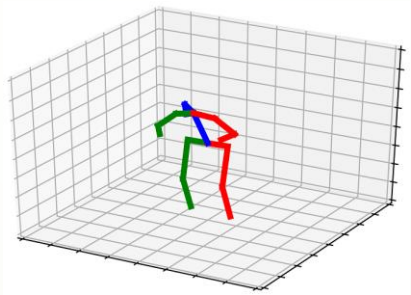
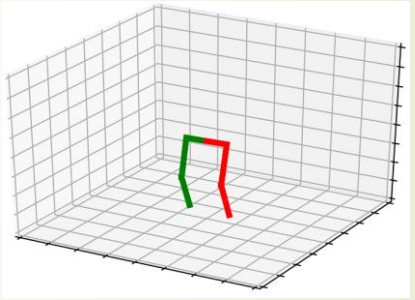
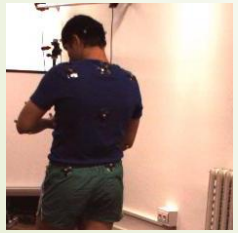
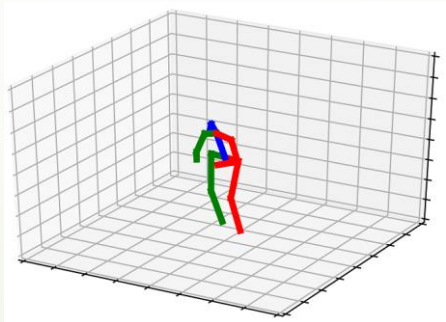
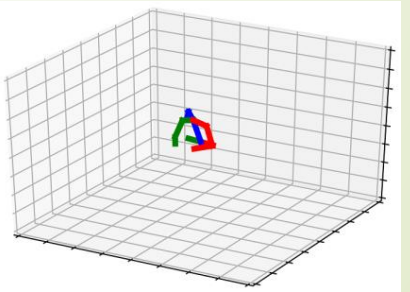




Results: Subjective evaluation

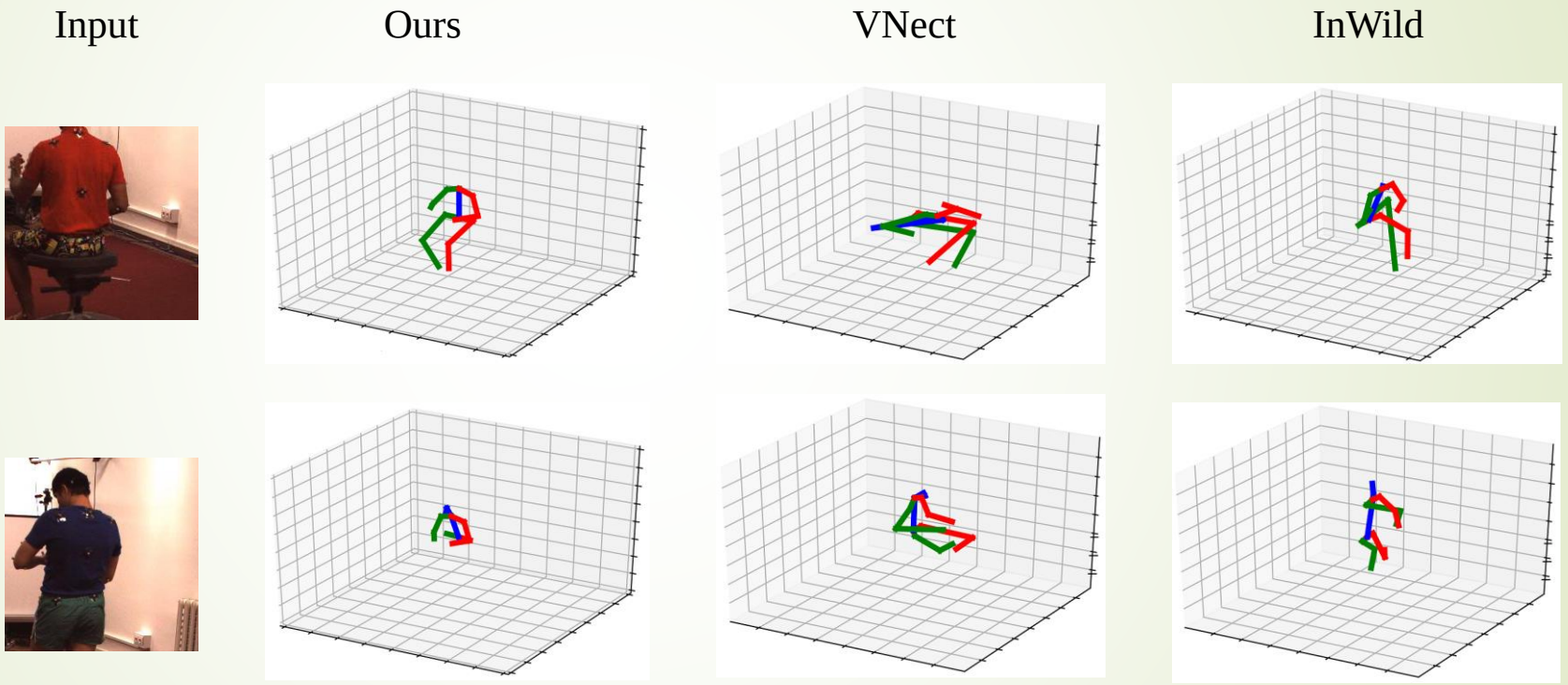
Subjective results under partial body presence:

Results: Subjective evaluation

Our method compared to VNect and InWild under partial body presence:



Conclusion

- ✓ A deep learning based method to handle 3D human pose estimation
- ✓ Handling partial presence in the input 2D image
- ✓ A CNN based detection network to classify the presence
- ✓ A deep CNN to regress the human pose from images containing partial body
- ✓ Empirical evaluations yield promising results on Human3.6M dataset.

Thank you for your time and attention.