Human Body Pose Estimation

C. Papaioannidis, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 4.4

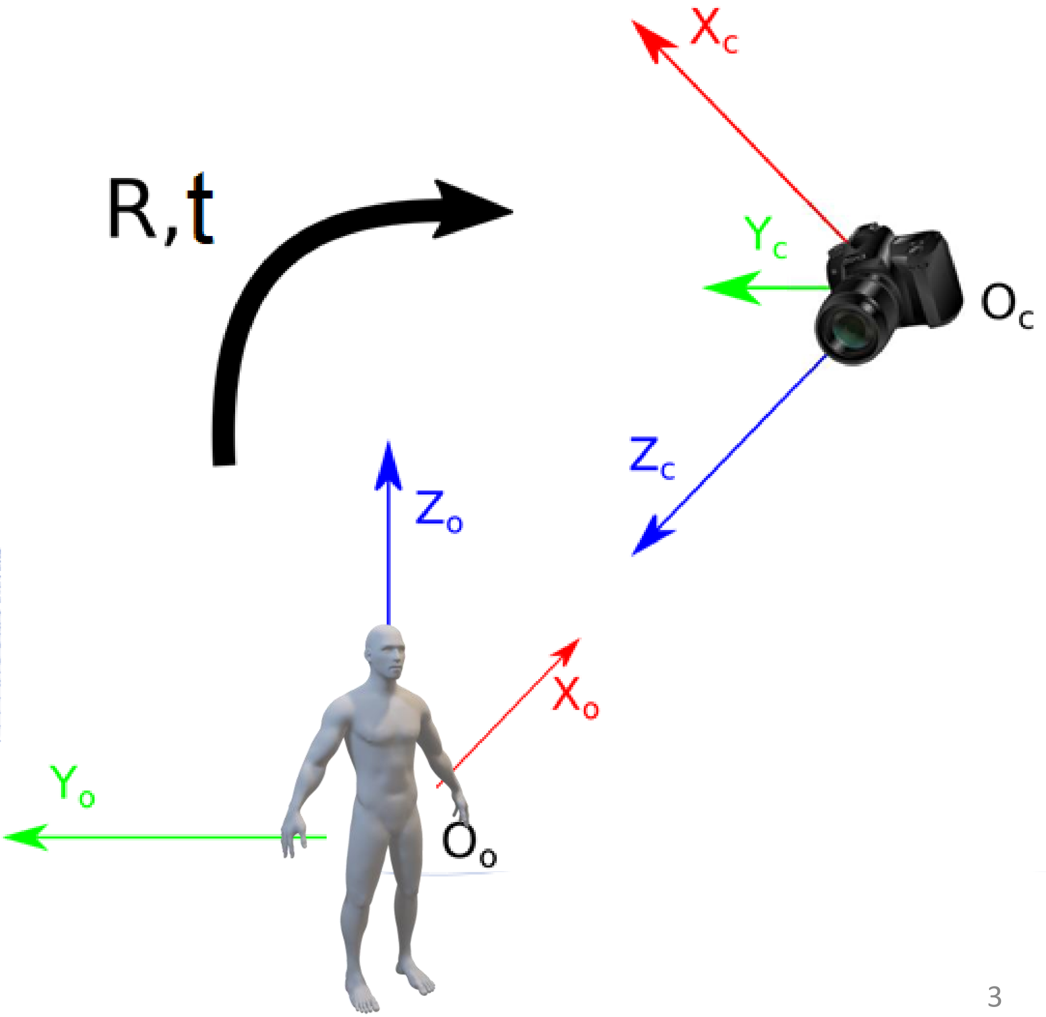
Artificial Intelligence & Information Analysis Lab

Human Body Pose Estimation

- **Introduction**
- Human body modeling
- Visual 2D human pose estimation
- Visual 3D human pose estimation
- 3D HPE from other sensors
- HPE data sets

Introduction

- ***Camera pose estimation*** involves estimating the 3D orientation and 3D translation of the camera relative to an object/human or vice-versa.



Introduction

- The human body is an ***articulated object***.
- ***Human body pose estimation*** entails estimating the locations of specific human body joints.
- It should not be confused with:
  - Either camera pose estimation or
  - human posture recognition.

Human body posture recognition

- ***Human body posture*** is a specific configuration of the body joints and is bound to a specific state, e.g., standing, sitting, lying, etc. .
- Human postures are different from human actions:
  - postures are static, while actions are dynamic.
- Human body posture recognition applications:
  - Physical training,
  - Rehabilitation training,
  - Sign language communication,
  - Human-computer interaction (HCI).

Human pose estimation

Human Pose Estimation (HPE) estimates the configuration of human body parts from input data captured by sensors (usually images and videos).

- It provides geometric and motion information of the human body.
- It can regress human body configuration parameters.
- Wide range of applications:
  - Human-computer interaction (HCI),
  - Motion analysis,
  - AR/VR,
  - Healthcare.

Human pose estimation

- Deep Neural Networks (DNNs) have achieved remarkable results in HPE.
- DNN-based approaches have outperformed classical computer vision methods.
- HPE challenges:
  - Human body part occlusion,
  - Training data availability,
  - Depth information ambiguity.

Human Body Pose Estimation

- Introduction
- **Human body modeling**
- Visual 2D human pose estimation
- Visual 3D human pose estimation
- 3D HPE from other sensors
- HPE data sets

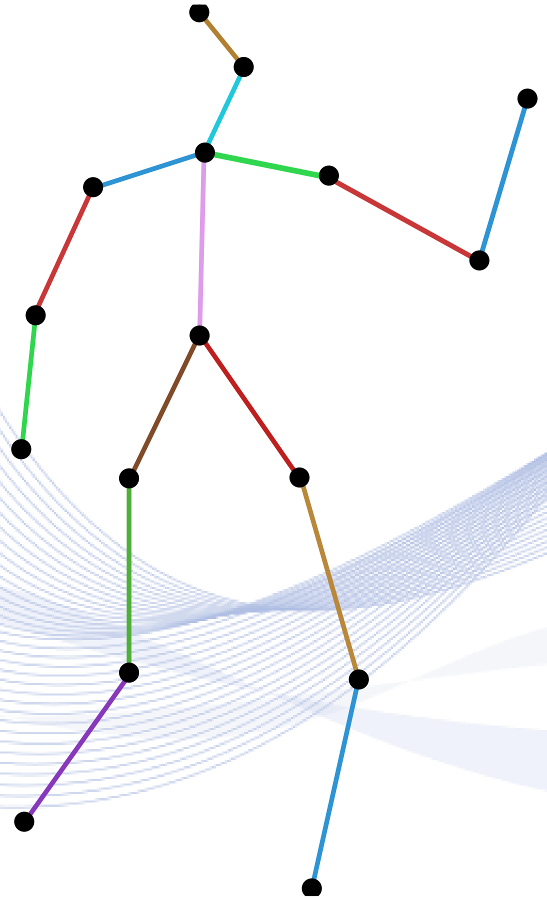
Human body modeling

- Human body modeling is an important aspect of HPE.
- Human body is a ***deformable articulated*** solid object:
  - It consists of joints and limbs,
  - It has a kinematic structure,
  - Body shape information is important.
- Body model types:
  - Kinematic model (2D/3D HPE),
  - Planar model (2D HPE),
  - 3D surface body model (3D HPE)
  - Volumetric model (3D HPE).

Human body modeling

Kinematic human body model

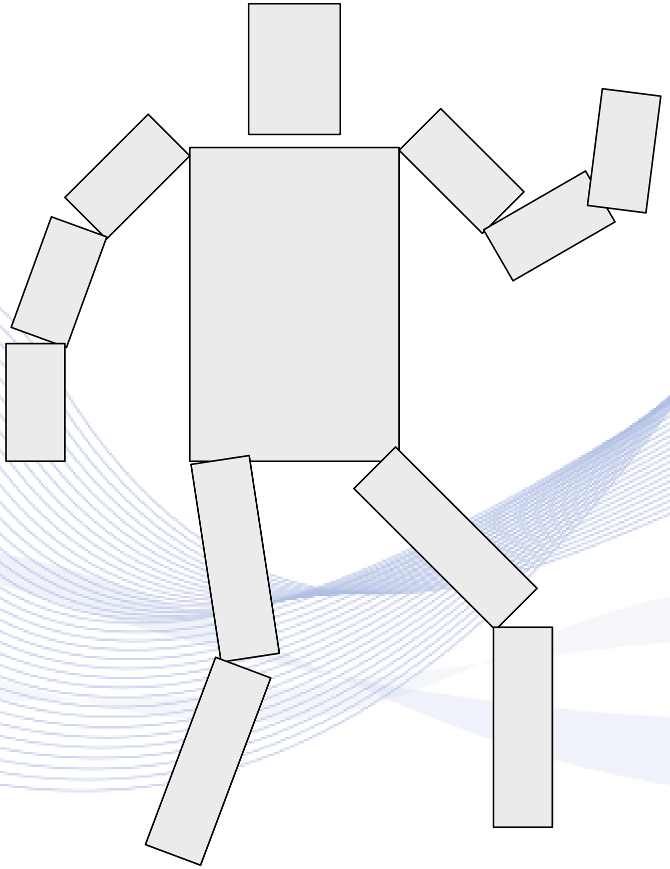
- Human body structure is represented by a set of 2D/3D joint positions (and limb position/orientations).
- ***Pictorial structure model*** (PSM) [ZUF2012] a.k.a. tree-structured model.
- Flexible and intuitive.
- Cannot represent texture and shape information.



Human body modeling

Planar human body model

- Body parts are represented by rectangles.
- ***Cardboard model*** [JU1996].
- Represents shape and appearance of the human body.



Human body modeling

3D surface human body model

- It describes the 3D body surface.
  - Triangular or polygonal mesh.
- ***Skinned Multi-Person Linear*** (SMPL) model [LOP2015].
- Modeled with natural pose-dependent deformations.
- Joint locations are calculated from the model vertices.

Human body modeling

Volumetric human body model

- Voxel-based human body models.
- Octree representations

Human Body Pose Estimation

- Introduction
- Human body modeling
- **Visual 2D human pose estimation**
- Visual 3D human pose estimation
- 3D HPE from other sensors
- HPE data sets

2D human pose estimation

- It involves the prediction of the 2D position or spatial location of human body key-points/joints from images or videos.
- Deep learning-based approaches have achieved remarkable results.
- Single-person 2D HPE:
  - Direct regression methods,
  - Heatmap-based methods.
- Multi-person 2D HPE:
  - Top-down pipeline,
  - Bottom-up pipeline.

2D human pose estimation

Single-person 2D HPE

- Localize human body joints when the input is a single-person image.

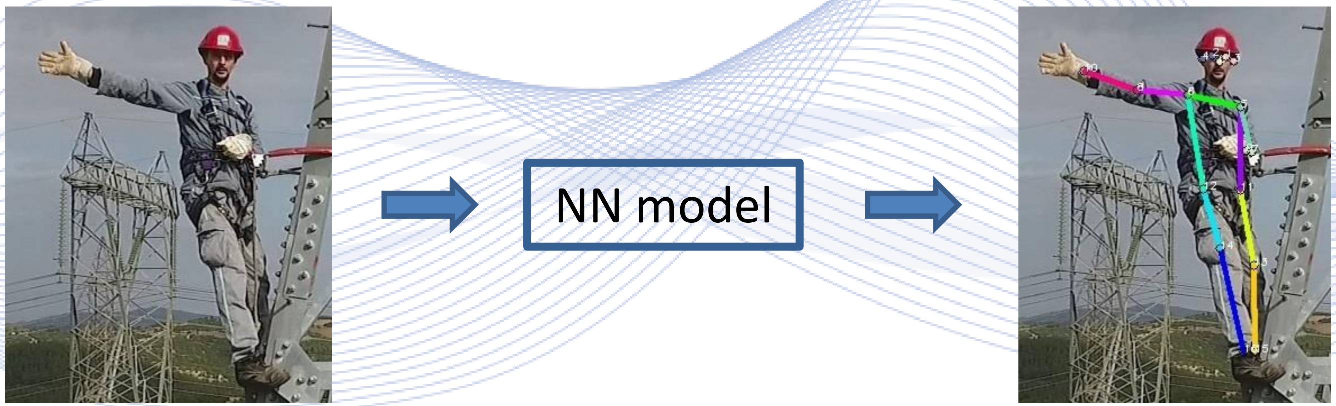
Artificial Intelligence & Information Analysis Lab

2D human pose estimation

Single-person 2D HPE

Direct regression methods

- End-to-end framework.
- Learn a mapping from the input image to body joints or parameters of human body models.



2D human pose estimation

Single-person 2D HPE

Direct regression methods

- If $\mathbf{I}$ is an input RGB image of resolution $M \times N$ and $\boldsymbol{f}$ is the 2D HPE DNN, direct regression methods aim to directly predict (estimate):

$$\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\} = \boldsymbol{f}(\mathbf{I}),$$

- $\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\}$: pre-defined set of body joints that constitute the 2D human pose,
- K is the number of the body joints
- $\mathbf{j}_k = [x_k, y_k]^T \in \mathbb{N}^2, k = 1, \dots, K$ human skeleton body joint representation using the pixel coordinates on the image plane.

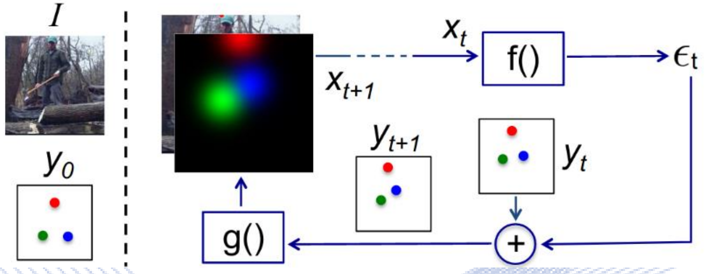
2D human pose estimation

Single-person 2D HPE

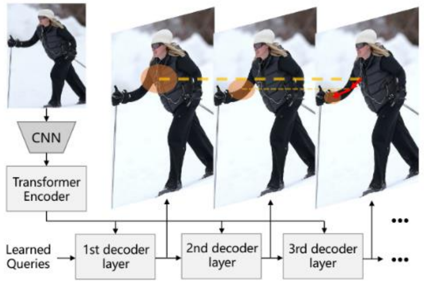
Direct regression methods

- Popular approaches:
  - DeepPose [TOS2014],
  - Iterative Error Feedback (IEF) network [CAR2016],
  - Compositional pose regression [SUN2017],
  - Cascaded transformer-based model (PRTR) [LI2021].

[TOS2014]



[CAR2016]



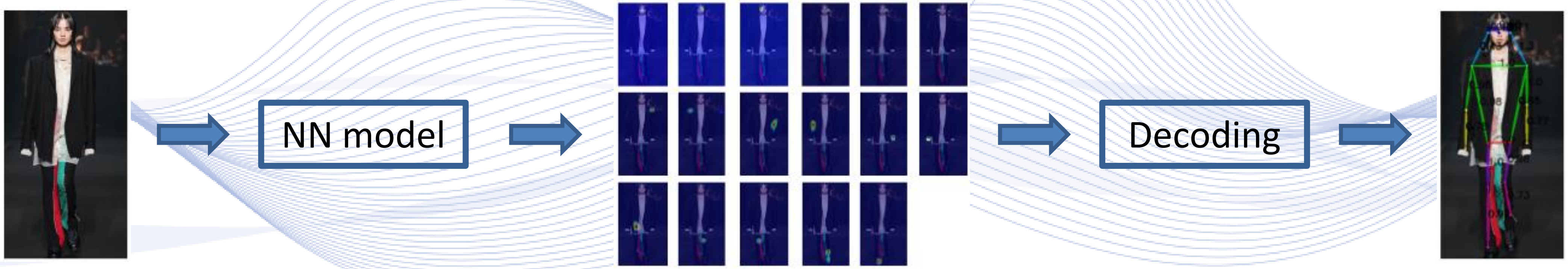
[LI2021]

2D human pose estimation

Single-person 2D HPE

Heatmap-based methods

- Train a body part detector to predict the position of body joints.
- Estimate ***joint heatmap images*** that represent the joint locations.



[DAN2019]

2D human pose estimation

Single-person 2D HPE

Heatmap-based methods

- Instead of directly predicting $\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\}$, $\boldsymbol{f}$ predicts 2D body joint heatmaps $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K\}$ of resolution $M \times N$ (one for each joint):
$$\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K\} = \boldsymbol{f}(\mathbf{I}).$$
- Each heatmap $\mathbf{H}_k \in \mathbb{R}^{M \times N}$ encodes the 2D location of the corresponding body joint by using a 2D Gaussian function centered at the 2D position of the body joint in the input image.
- 2D pixel coordinates of each body joint can be obtained by choosing the (x_k, y_k) pairs with the highest heat value.

2D human pose estimation

Single-person 2D HPE

Heatmap-based methods

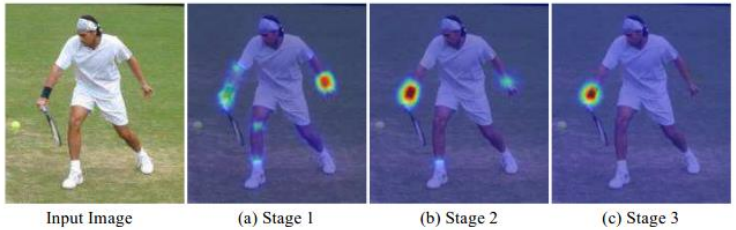
- Heatmaps provide richer supervision information, by preserving the spatial location information.
- Allow using the powerful ***Convolutional Neural Networks*** (CNNs).
- Facilitate DNN/CNN training.
- Used in state-of-the-art 2D HPE approaches.

2D human pose estimation

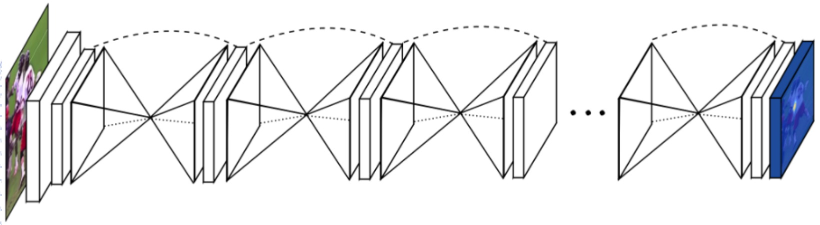
Single-person 2D HPE

Heatmap-based methods

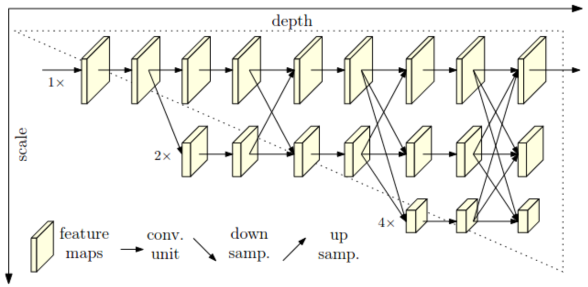
- Typical CNN-based approaches:
  - Convolutional Pose Machines (CPM) [WEI2016],
  - Stacked Hourglass [NEW2016],
  - High-Resolution Network (HRNet) [SUN2019].



[WEI2016]



[NEW2016]



[SUN2019]

2D human pose estimation

Single-person 2D HPE

Heatmap-based methods

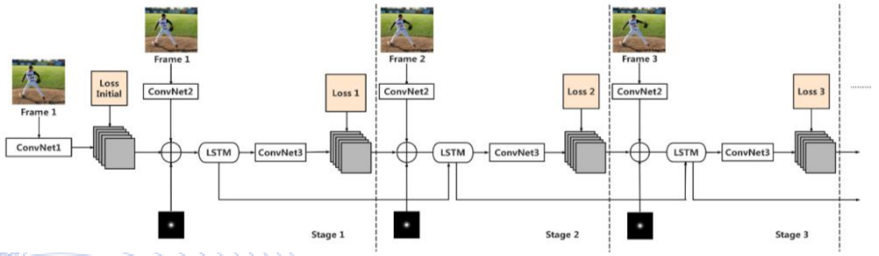
- The emergence of Generative Adversarial Networks (GANs) gave rise to GAN-based 2D HPE methods.
- GANs can discriminate between real human and predicted ones.
- GANs were used to force the 2D HPE model to predict plausible pose configurations.
- They provide increased performance in difficult cases (e.g., body occlusion).
- GAN-based approaches:
  - Adversarial PoseNet [CHE2017],
  - Adversarial HPE [PEN2018].

2D human pose estimation

Single-person 2D HP

2D HPE in video sequences

- Video sequences are spatio-temporal (3D) signals.
- The temporal information of a video can be exploited by a model capable of handling sequential data, such as:
  - ***Recurrent Neural Networks*** (RNN) or
  - ***Long Shot-Term Memory*** (LSTM) networks.



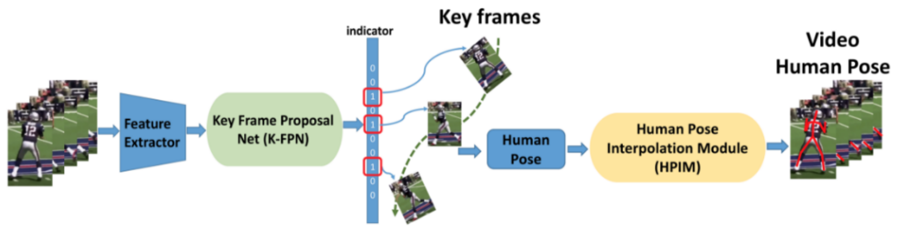
[LUO2018]

2D human pose estimation

Single-person 2D HP

2D HPE in video sequences

- Video-based 2D HPE approaches aim to model the spatio-temporal human body pose information.
  - Long Shot-Term Memory (LSTM) Machines [LUO2018],
  - Key Frame Proposal Network (K-FPN) [ZHA2020].



[ZHA2020]

2D human pose estimation

Multi-person 2D HPE

- Estimate the 2D skeletons of multiple persons that appear in the input image.
  - All persons must be localized,
  - Detected body keypoints must be grouped for different persons.

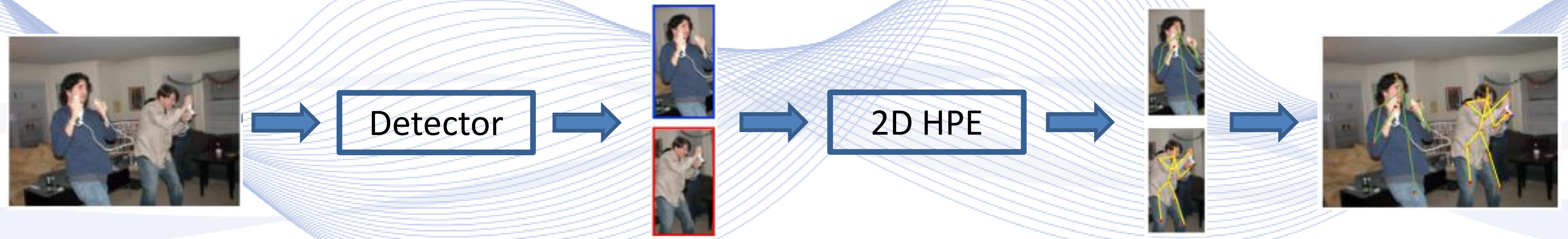
[CAO2017]

2D human pose estimation

Multi-person 2D HPE

Top-down pipeline

- Each person is detected on the input image (2D bounding boxes) using off-the-shelf person detectors [REN2015].
- Single-person HPE is performed to each person bounding box.



[DAN2019]

2D human pose estimation

Multi-person 2D HPE

Top-down pipeline

- Inference speed increases linearly with the number of persons.
- Research focuses on:
    - Designing and improving the person detection and 2D HPE components, as well as the cooperation between them [MOO2019].
    - Successfully handling cases with occlusion and/or truncation [FAN2017].
    - Exploiting the power of ***Transformers*** and their ability to encode long-range dependencies [LI2021].

2D human pose estimation

Multi-person 2D HPE

Bottom-up pipeline

- Localize all the body joints in the input image.
- Group the detected body joints to the corresponding persons.

[DAN2019]

2D human pose estimation

Multi-person 2D HPE

Bottom-up pipeline

- Inference speed is usually increased, compared to top-down approaches, since there is no need to detect the body joints for each person separately.
- Research mainly focuses on:
  - Improving ***body joint grouping*** and association to each person [INS2016], [JIN2020].
  - Improving multi-person 2D HPE in low-resolution images [KRE2019].
  - Unifying the body joint detection and grouping stages with single-stage DNNs [NEW2017].

Human Body Pose Estimation

- Introduction
- Human body modeling
- Visual 2D human pose estimation
- **Visual 3D human pose estimation**
- 3D HPE from other sensors
- HPE data sets

3D human pose estimation

- It predict the body joint locations in 3D space.
- It provides 3D structure information related to human body.

- It remains a challenging task.
- ***3D pose annotation*** is costly and time-consuming.
- ***Limited availability of datasets***:
  - Generalization issues,
  - Problems in real-world applications.

Artificial Intelligence & Information Analysis Lab

3D human pose estimation

- Image/video-based 3D HPE:
  - Monocular, single-person.
    - 3D skeleton estimation,
    - Human mesh reconstruction.
  - Monocular, multi-person.
    - Top-down pipeline,
    - Bottom-up pipeline.
  - Multi-view.
- 3D HPE from other sources.

3D human pose estimation

3D HPE from monocular images/videos

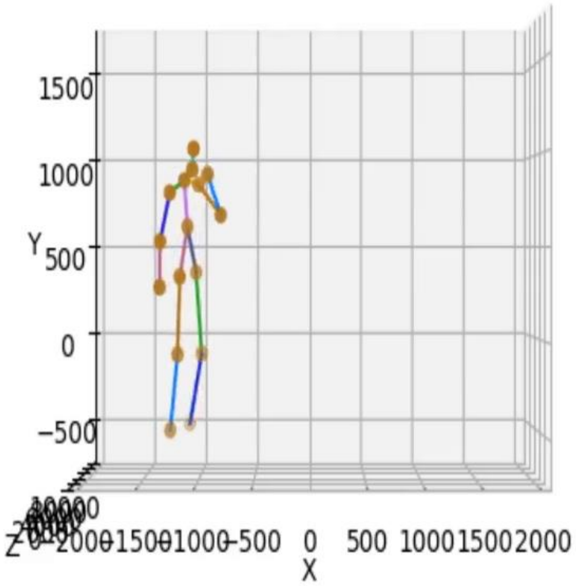
- 3D HPE from monocular images/videos is the most popular approach.
- One monocular RGB camera is required.
- Predicting 3D human poses in this is very challenging:
  - Occlusions,
  - Depth ambiguities,
  - Insufficient data,
  - Different 3D human poses can be projected to similar 2D poses.

3D human pose estimation

3D HPE from monocular images

Single-person

- 3D skeleton estimation (kinematic model): Predict the body joint locations in the 3D space.

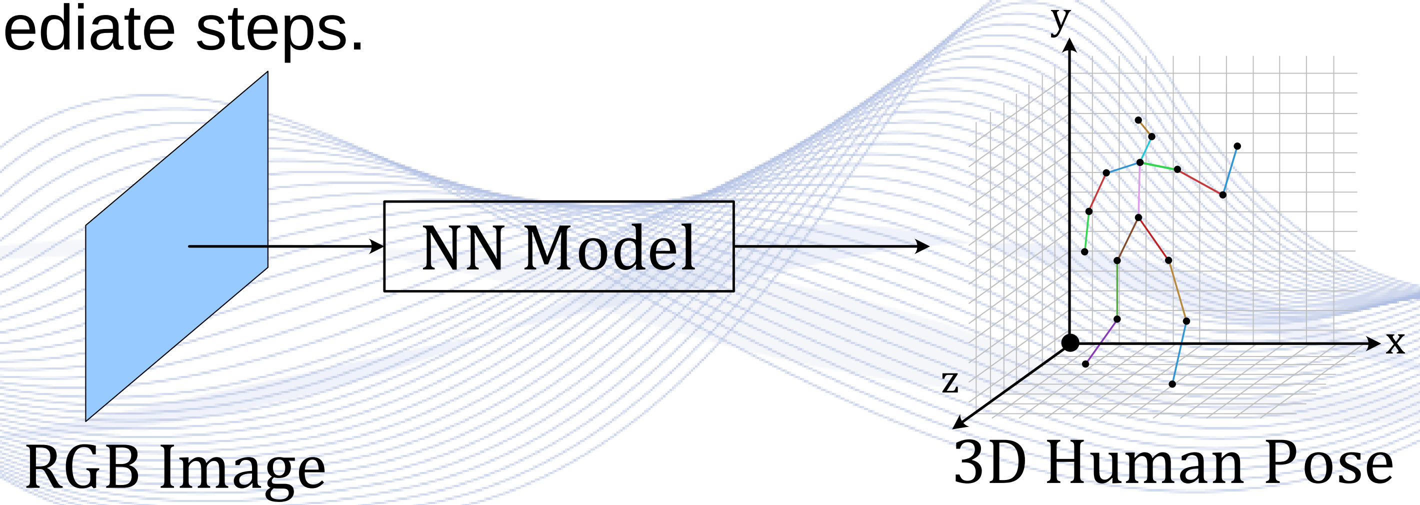


3D human pose estimation

3D HPE from monocular images

Single-person

- Direct 3D skeleton estimation from an RGB image: The 3D human pose is obtained directly from the input image without any intermediate steps.



3D human pose estimation

3D HPE from monocular images

Single-person

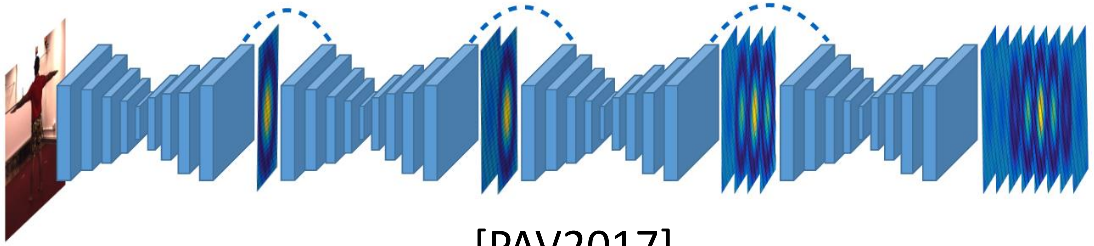
- Methods based on CNNs.
- If $\mathbf{I}$ is an input RGB image of resolution $M \times N$ and f is the 3D HPE CNN, direct 3D skeleton estimation methods aim to predict:
$$\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\} = \boldsymbol{f}(\mathbf{I}),$$
- $\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\}$ is the set of 3D skeleton body joints,
- K is the number of the body joints
- $\mathbf{j}_k = [X_k, Y_k, Z_k]^T \in \mathbb{R}^3, k = 1, \dots, K$ represents the 3D coordinates of each 3D human body.

3D human pose estimation

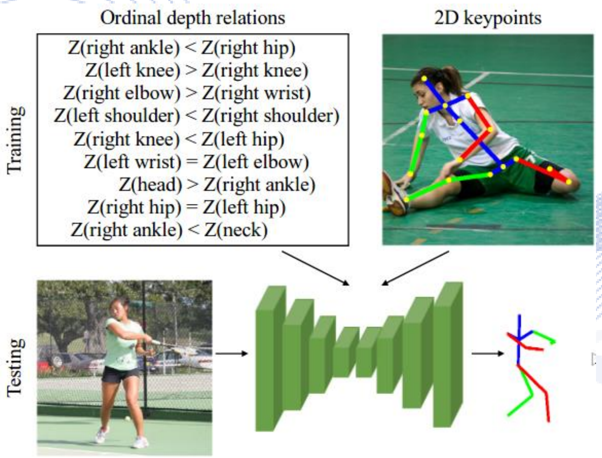
3D HPE from monocular images

Single-person

- Typical direct 3D skeleton estimation approaches:
  - DconvMP [LI2014],
  - Coarse-to-Fine 3D HPE [PAV2017],
  - Ordinal Depth Supervision for 3D HPE [PAV2018].



[PAV2017]



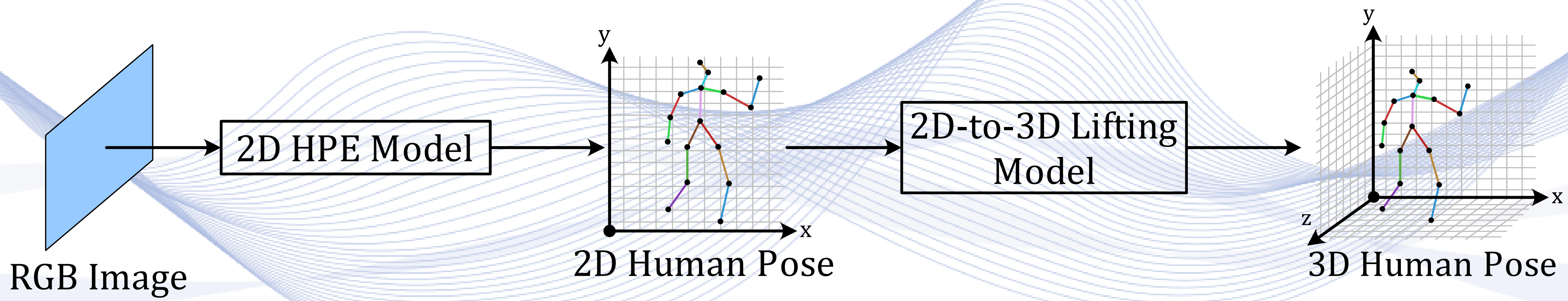
[PAV2018]

3D human pose estimation

3D HPE from monocular images

Single-person

- ***2D-to-3D lifting***: A 2D skeleton is first extracted from the input RGB image, which is then lifted to the corresponding 3D skeleton.



3D human pose estimation

3D HPE from monocular images

Single-person

- 2D-to-3D lifting methods were motivated by the success of 2D HPE methods.
- The 2D skeleton extraction stage can be implemented using off-the-shelf 2D HPE methods [SUN2019].
- If $\mathbf{I}$ is an input RGB image of resolution $M \times N$, $\boldsymbol{f}$ is the 2D HPE CNN and $\boldsymbol{g}$ is the 2D-to-3D lifting DNN, then the corresponding 3D skeleton is predicted as follows:

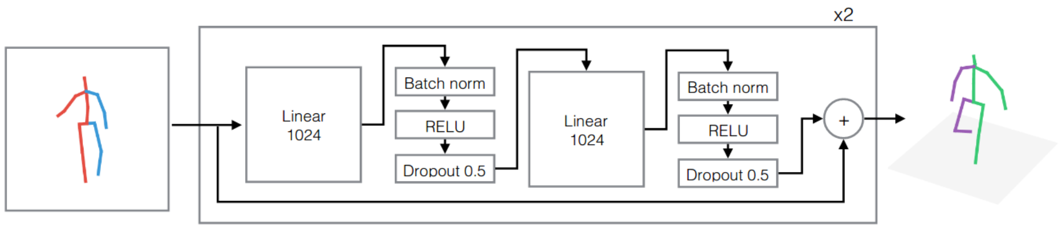
$$\{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_K\} = \boldsymbol{g}(\boldsymbol{f}(\mathbf{I})).$$

3D human pose estimation

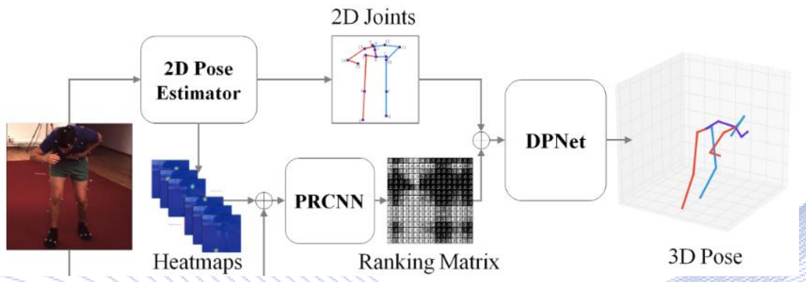
3D HPE from monocular images

Single-person

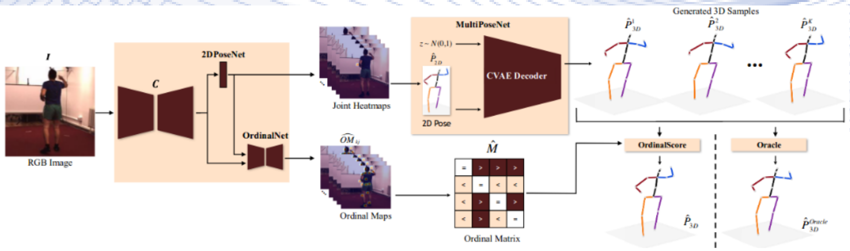
- Typical 2D-to-3D lifting approaches with CNNs/DNNs:
  - Simple yet effective 3D HPE [MAR2017],
  - DRPose3D [WAN2018],
  - MultiPoseNet [SHA2019].



[MAR2017]



[WAN2018]



[SHA2019]

3D human pose estimation

3D HPE from monocular images

Single-person

- The human body kinematic model allows the representation of 2D and 3D human poses as graphs.
- The body joints and bones are the graph nodes and the edges.
- Human graph: $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of K body joints/nodes and $\mathcal{E}$ is a set of B bones/edges.
- This allowed 2D-to-3D lifting to be performed using ***Graph Convolutional Networks*** (GCNs).

3D human pose estimation

3D HPE from monocular images

Single-person

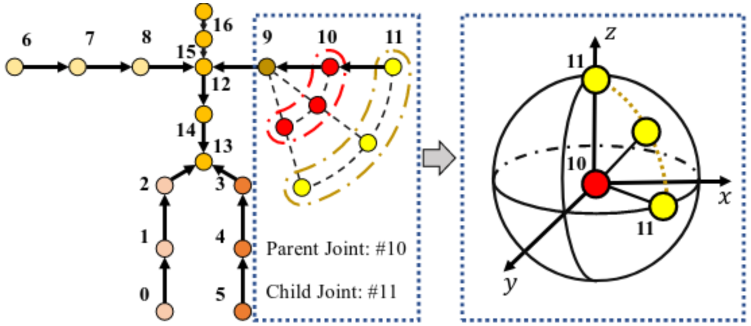
- 2D-to-3D lifting with GCNs allows:
  - Modeling local and global body joint and bone relations by utilizing an adjacency matrix $\mathbf{A} \in \mathbb{R}^{K \times K}$ in each GCN layer, which encodes the human graph structure.
- GCN-based 2D-to-3D lifting approaches:
  - Locally Connected Network (LCN) [CI2019],
  - SemGCN [ZHA2019].

3D human pose estimation

3D HPE from monocular images

Single-person

- The kinematic model also allows the exploitation of the kinematic constraints of the human body.
    - Body joints connectivity information,
    - Joints rotation properties,
    - Fixed bone length ratios.
- Constraints can be enforced on the 3D HPE model outputs.



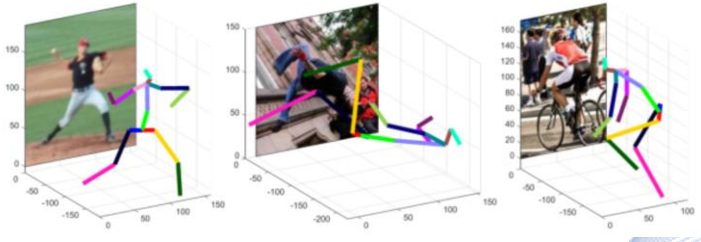
[XU2020]

3D human pose estimation

3D HPE from monocular images

Single-person

- ***3D HPE in-the-wild*** involves predicting 3D human poses in more challenging scenarios, such as outdoor sports.
  - Limited or no availability of annotated datasets.
- Approaches:
  - Enforce kinematic constraints.
  - Weakly-supervised training through 3D-to-2D reprojection [WAN2019].



[WAN2019]

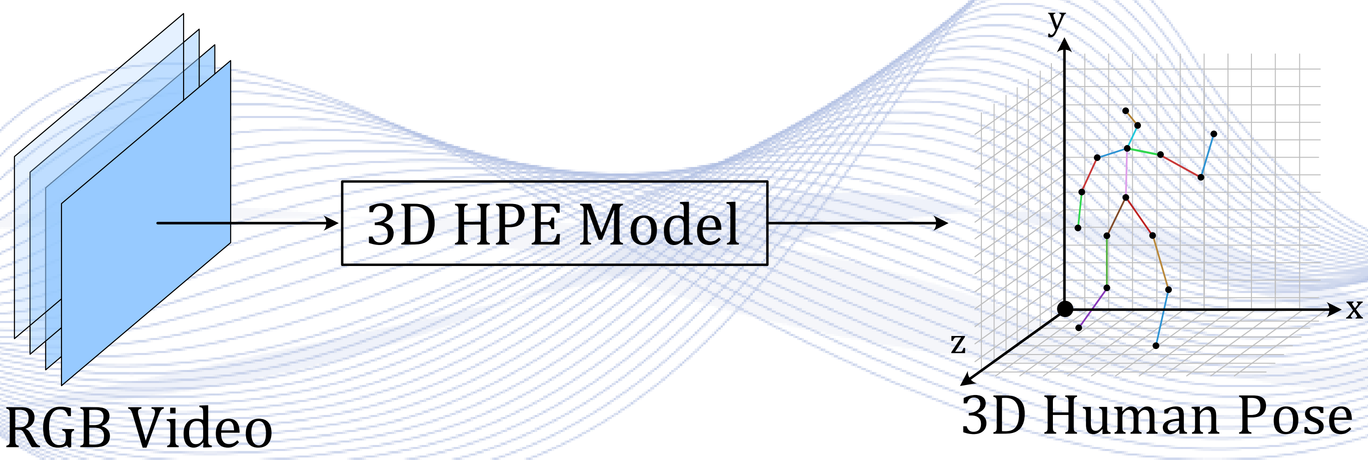
Artificial Intelligence & Information Analysis Lab

3D human pose estimation

3D HPE from monocular videos

Single-person

- Videos provide temporal information, which can improve the accuracy and the robustness of 3D HPE.



3D human pose estimation

3D HPE from monocular videos

Single-person

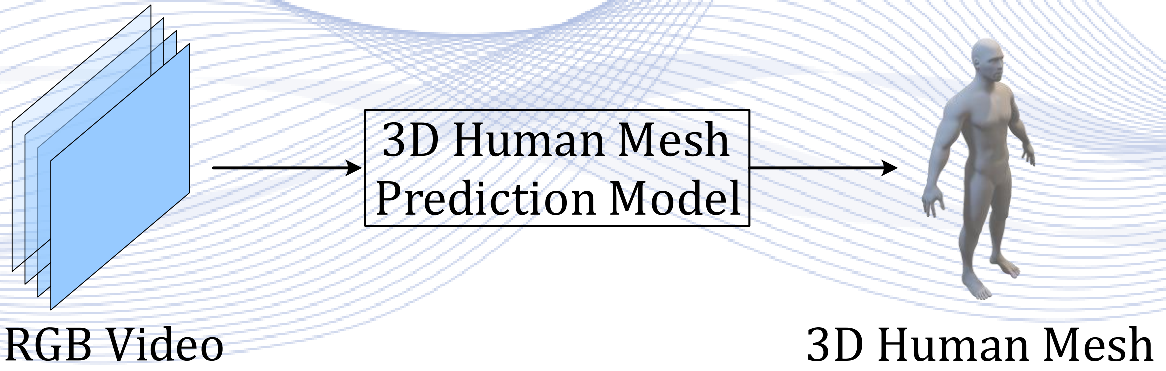
- The temporal information of a video can be exploited by a model capable of handling sequential data, such as ***RNNs*** or ***LSTM network***.
- Occlusions or ambiguities on a single frame can be alleviated by additional information provided by neighbouring frames.
- Video-based approaches:
  - LSTM-based [HOS2018],
  - GCN-based [CAI2019],
  - Transformer-based [LI2022].

3D human pose estimation

3D HPE from monocular images/videos

Single-person

- Human body surface mesh reconstruction methods incorporate parametric body models (Human body surface model).
- The 3D skeleton can also be obtained using a model-defined joint regression matrix.



3D human pose estimation

3D HPE from monocular images/videos

Single-person

- Human surface meshes provide rich information about body shape and texture, as well as a more accurate representation of the 3D human pose.
- The ***SMPL*** is the most popular human body model.
  - Predefined representation of a human mesh.
  - Simple to use,
  - Compatible with existing rendering engines,
  - Computationally intensive.

3D human pose estimation

3D HPE from monocular images/videos

Single-person

- Human mesh reconstruction approaches:
  - Regression of SMPL parameters [OMR2018],
  - Regression of vertex locations [KOL2019a].
- SMPL-based models:
  - SMPLify [LAS2017],
  - SMPL-X [PAV2019],
  - SPIN [KOL2019b]
  - STAR [OSM2020].

3D human pose estimation

3D HPE from monocular images

Multi-person

- Estimate the 3D skeletons of multiple persons in an input image.
- ***Top-down pipeline***: Similar to the 2D HPE case,
  - each person is first detected on the input image and
  - individual 3D skeletons are then estimated.
- ***Bottom-up pipeline***:
  - First predict all body joints and depth maps and then
  - group and associate all detected body parts to each person.

3D human pose estimation

3D HPE from monocular images

Multi-person

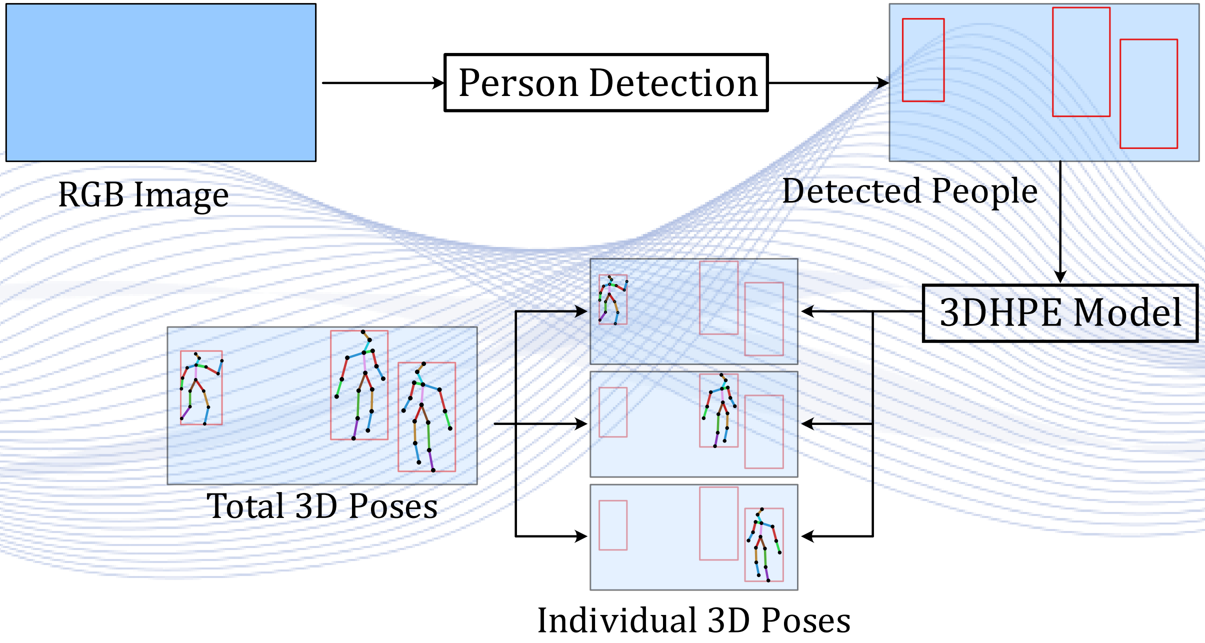
- ***Top-down pipeline***:
  - Utilize off-the-shelf person detectors to predict a 2D bounding box for each person in the image.
  - For each predicted person 2D bounding box, predict 3D human poses using single-person 3D HPE approaches.
  - The estimated 3D human poses are ***aligned*** to the 3D world coordinates system by also predicting an absolute 3D coordinate for each detected person.

3D human pose estimation

3D HPE from monocular images

Multi-person

- Top-down pipeline.



3D human pose estimation

3D HPE from monocular images

Multi-person

- ***Top-down pipeline***:
  - It achieves promising results.
  - Human mesh reconstruction is straightforward.
  - Computations increase linearly with the person number.
  - Global information for the scene is lost since a detection step is first applied.
  - Popular approaches:
    - LCR-Net [ROG2017], LCR-Net++ [ROG2019], PandaNet [BEN2020].

3D human pose estimation

3D HPE from monocular images

Multi-person

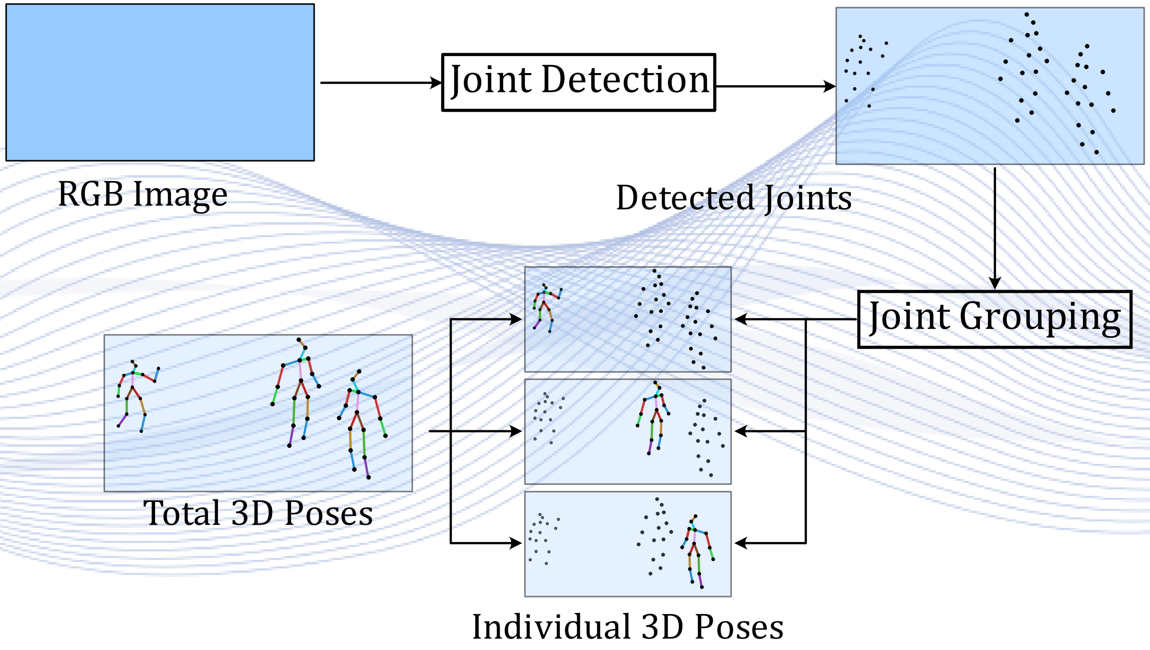
- ***Bottom-up pipeline***:
  - All visible body joints are detected on the 2D image along with the corresponding depth maps.
  - Detected body parts are associated to each person, according to a predicted ***global depth*** and ***part relative depth***.

3D human pose estimation

3D HPE from monocular images

Multi-person

- Bottom-up pipeline.



3D human pose estimation

3D HPE from monocular images

Multi-person

- ***Bottom-up pipeline***:
  - Faster execution speed.
  - Human mesh reconstruction is not straightforward.
  - ***Body joint grouping is challenging.***
  - Occlusions can cause inaccurate predictions.
  - Popular approaches:
    - Single-stage multi-person Pose Machine [NIE2019],
    - Occlusion-Robust Pose-Maps (ORPM) [MEH2018].

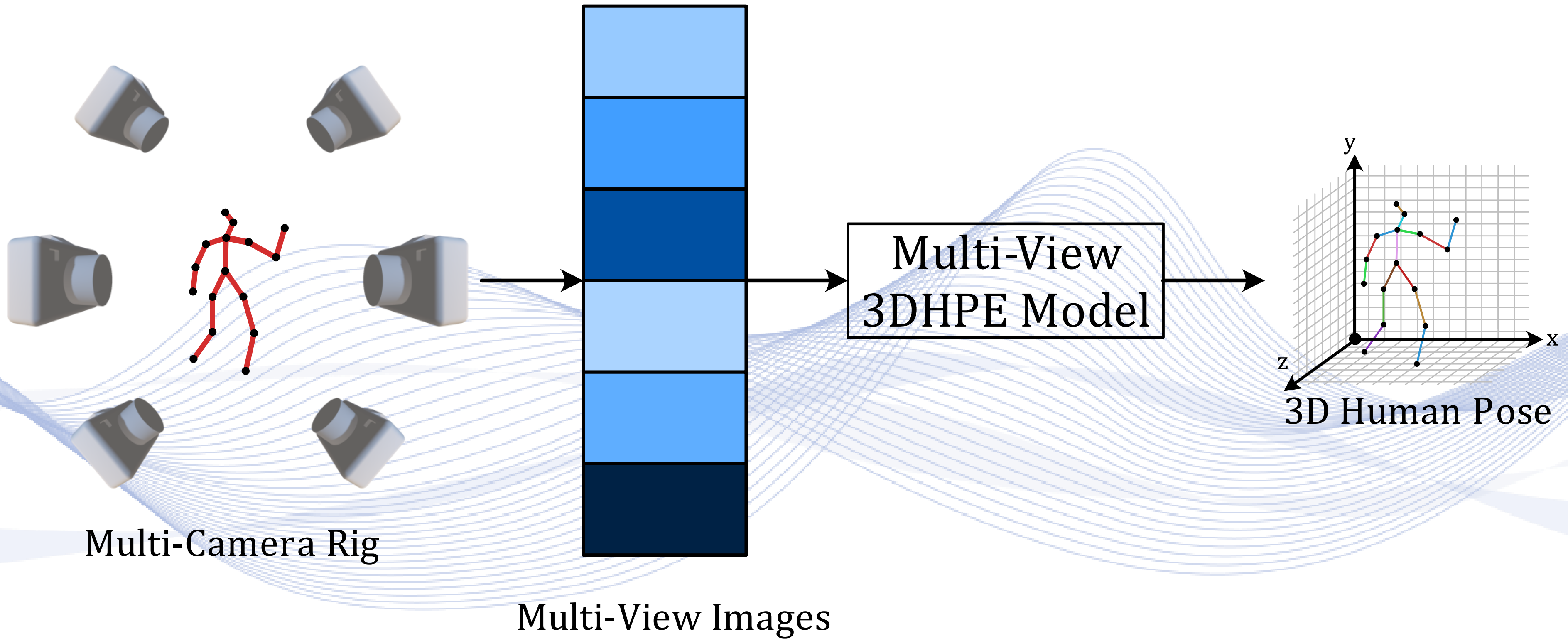
3D human pose estimation

Multi-view 3D HPE

- It can provide a solution in the case of ***partial human body occlusion***.
  - Since the 3D human pose is estimated from multiple views, the occluded part in one view may become visible in other views.
- 3D human pose reconstruction from multiple views requires the ***association of corresponding joint locations***, as images by different cameras.
- Mainly used for multi-person 3D HPE.

3D human pose estimation

Multi-view 3D HPE



3D human pose estimation

Multi-view 3D HPE

- There are various Multiview 3D HPE approaches:
  - Based on body models (3D pictorial model [BUR2013]) [DON2021].
    - Increased computational cost.
    - Memory-demanding
  - Multi-view matching frameworks [HUA2020].
  - Direct 3D human pose regression from multi-view images [ZHA2021].
- ***Lightweight architectures***, increased inference speed and efficient adaptation to different multi-view settings are also important features.

Human Body Pose Estimation

- Introduction
- Human body modeling
- Visual 2D human pose estimation
- Visual 3D human pose estimation
- **3D HPE from other sensors**
- HPE data sets

3D human pose estimation

3D HPE from other sensors

- Besides RGB images/videos, data from other sensors can also be used for 3D HPE.
  - Depth sensors,
  - Inertial Measurement Units (IMUs),
  - Radio frequency devices,
  - Non-line-of-sight (NLOS) imaging system, etc. .
- Data from these sensors can be used individually or alongside RGB data.

3D human pose estimation

3D HPE from other sensors

Depth sensors

- Popular in 3D vision tasks.
- Low-cost.
- Easy-to-use.
- Tackle depth ambiguity problem.
- 3D HPE approaches that utilize depth sensors:
  - [KAD2017],
  - [YU2018],
  - [ZHI2020].

[TER]

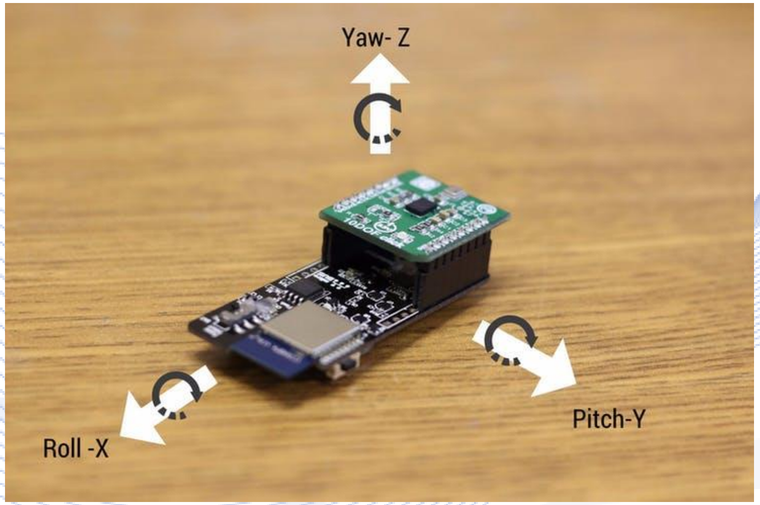
Artificial Intelligence & Information Analysis Lab

3D human pose estimation

3D HPE from other sensors

Inertial Measurement Units (IMUs)

- ***Wearable devices*** that can track the movement of specific body parts.
- Can be used to infer body motion and structure.
- Do not suffer from occlusion or clothes obstruction problems.
- They can be inaccurate due to ***drifting***.
- IMU-based 3D HPE approaches:
  - [HUA2020], [VON2018], [ZHA2020b].



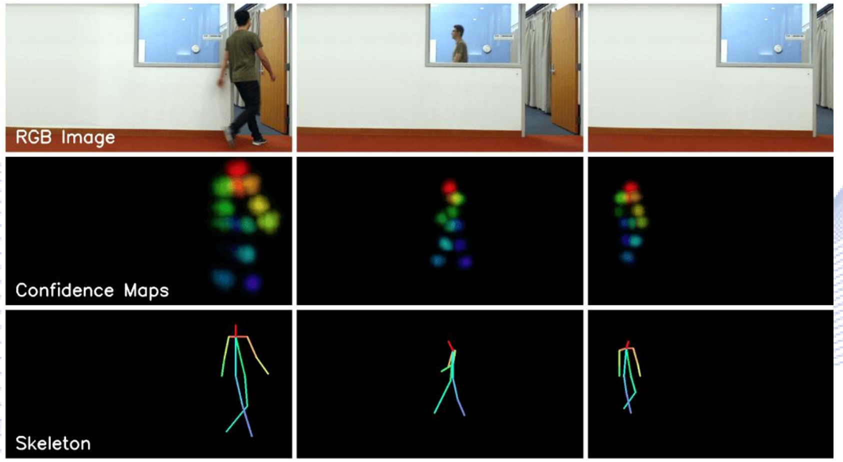
[HAC]

3D human pose estimation

3D HPE from other sensors

Radio Frequency** (**RF**) **devices

- ***Ability to traverse walls***.
- Humans do not need to carry any device.
- Privacy-preserving.
- Low spatial resolution.
- 3D HPE approaches based on RF signals:
  - [ZHA2018], [ZHA2019b].



[ZHA2018b]

Artificial Intelligence & Information Analysis Lab

Human Body Pose Estimation

- Introduction
- Human body modeling
- Visual 2D human pose estimation
- Visual 3D human pose estimation
- 3D HPE from other sensors
- **HPE data sets**

Human pose estimation datasets

- Annotated data are really important for producing reliable human pose estimation algorithms.
- Ideally, a dataset for human pose estimation should contain a large number of data samples, obtained using different persons, different scenes and a great variety of postures/actions.
- Each dataset may correspond to a specific real-world application scenario.

Human pose estimation datasets

2D HPE datasets

Image-based

- Single-person:
  - LSP [JOH2010], LSP-extended [JOH2011],
  - FLIC - FLIC-full [SAP2013], FLIC-plus [TOM2014],
  - MPII [AND2014].
- Multi-person:
  - MPII [AND2014],
  - COCO2016 - COCO2017[LIN2014],
  - CrowdPose [LI2019].

Video-based

- Single-person:
  - Penn Action [ZHA2013],
  - J-HMDB [JHU2013].
- Multi-person:
  - PoseTrack [AND2018].

Human pose estimation datasets

3D HPE datasets

Monocular and multi-view

- Single-person:
  - HumanEva [SIG2010],
  - Human3.6M [ION2013],
  - CMU Panoptic [JOO2015],
  - MPI-INF-3DHP [MEH2017],
  - 3DPW [VON2018] (no multi-view),
  - MuPoTS-3D [MEH2018].
- Multi-person:
  - CMU Panoptic [JOO2015],
  - 3DPW [VON2018] (no multi-view),
  - MuPoTS-3D [MEH2018].

Bibliography

[ZUF2012] Zuffi, Silvia, Oren Freifeld, and Michael J. Black. "From pictorial structures to deformable structures." *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[JU1996] Ju, Shanon X., Michael J. Black, and Yaser Yacoob. "Cardboard people: A parameterized model of articulated image motion." *2nd International Conference on Automatic Face and Gesture Recognition*, 1996.

[LOP2015] Loper, Matthew, et al. "SMPL: A skinned multi-person linear model." *ACM Transactions on Graphics,* vol 34, no. 6, pp.1-16, 2015.

[TOS2014] Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." *IEEE Conference on Computer Vision and Pattern Recognition,* 2014.

[CAR2016] Carreira, Joao, et al. "Human pose estimation with iterative error feedback." *IEEE Conference on Computer Vision and Pattern Recognition,* 2016.

[SUN2017] Sun, Xiao, et al. "Compositional human pose regression." *IEEE International Conference on Computer Vision,* 2017.

[LI2021] Li, Ke, et al. "Pose recognition with cascade transformers." *IEEE Conference on Computer Vision and Pattern Recognition,* 2021.

[DAN2019] Dang, Qi, et al. "Deep learning based 2d human pose estimation: A survey." *Tsinghua Science and Technology* vol 24, no. 6, pp. 663-676, 2019.

[WEI2016] Wei, Shih-En, et al. "Convolutional pose machines." *IEEE Conference on Computer Vision and Pattern Recognition,* 2016.

[NEW2016] Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." *European Conference on Computer Vision,* 2016.

[SUN2019] Sun, Ke, et al. "Deep high-resolution representation learning for human pose estimation" *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.

[CHE2017] Chen, Yu, et al. "Adversarial posenet: A structure-aware convolutional network for human pose estimation." *IEEE International Conference on Computer Vision,* 2017.

[PEN2018] Peng, Xi, et al. "Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.

Bibliography

[LUO2018]Luo, Yue, et al. "Lstm pose machines." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.
[ZHA2020] Zhang, Yuexi, et al. "Key frame proposal network for efficient pose estimation in videos." *European Conference on Computer Vision,* 2020.
[CAO2017] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *IEEE Conference on Computer Vision and Pattern Recognition,* 2017.
[REN2015] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in Neural Information Processing Systems,* 2015.
[MOO2019] Moon, Gyeongsik, Ju Yong Chang, and Kyoung Mu Lee. "Posefix: Model-agnostic general human pose refinement network."*IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[FAN2017] Fang, Hao-Shu, et al. "Rmpe: Regional multi-person pose estimation." *IEEE International Conference on Computer Vision,* 2017.
[LI2021] Li, Yanjie, et al. "Tokenpose: Learning keypoint tokens for human pose estimation." *IEEE International Conference on Computer Vision,* 2021.
[INS2016] Insafutdinov, Eldar, et al. "Deepercut: A deeper, stronger, and faster multi-person pose estimation model." *European Conference on Computer Vision,* 2016.
[JIN2020] Jin, Sheng, et al. "Differentiable hierarchical graph grouping for multi-person pose estimation." *European Conference on Computer Vision,* 2020.
[KRE2019] Kreiss, Sven, Lorenzo Bertoni, and Alexandre Alahi. "Pifpaf: Composite fields for human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[NEW2017] Newell, Alejandro, Zhiao Huang, and Jia Deng. "Associative embedding: End-to-end learning for joint detection and grouping." *Advances in Neural Information Processing Systems,* 2017.
[PAV2018] Pavlakos, Georgios, Xiaowei Zhou, and Kostas Daniilidis. "Ordinal depth supervision for 3d human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.
[PAV2017] Pavlakos, Georgios, et al. "Coarse-to-fine volumetric prediction for single-image 3D human pose."*IEEE Conference on Computer Vision and Pattern Recognition,* 2017.
[LI2014] Li, Sijin, and Antoni B. Chan. "3d human pose estimation from monocular images with deep convolutional neural network." *Asian Conference on Computer Vision*. Springer, 2014.

Bibliography

[MAR2017] Martinez, Julieta, et al. "A simple yet effective baseline for 3d human pose estimation." *IEEE International Conference on Computer Vision,* 2017.
[WAN2018] Wang, Min, et al. "Drpose3d: Depth ranking in 3d human pose estimation." *arXiv preprint arXiv:1805.08973,* 2018.
[SHA2019] Sharma, Saurabh, et al. "Monocular 3d human pose estimation by generation and ordinal ranking." *IEEE International Conference on Computer Vision,* 2019.
[CI2019] Ci, Hai, et al. "Optimizing network structure for 3d human pose estimation." *IEEE International Conference on Computer Vision,* 2019.
[ZHA2019] Zhao, Long, et al. "Semantic graph convolutional networks for 3d human pose regression." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[XU2020] Xu, Jingwei, et al. "Deep kinematics analysis for monocular 3d human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2020.
[WAN2019] Wandt, Bastian, and Bodo Rosenhahn. "Repnet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[HOS2018] Hossain, Mir Rayat Imtiaz, and James J. Little. "Exploiting temporal information for 3d human pose estimation." *European Conference on Computer Vision,* 2018.
[CAI2019] Cai, Yujun, et al. "Exploiting spatial-temporal relationships for 3d pose estimation via graph convolutional networks." *IEEE International Conference on Computer Vision,* 2019.
[LI2022] Li, Wenhao, et al. "Exploiting temporal contexts with strided transformer for 3d human pose estimation." *IEEE Transactions on Multimedia,* 2022.
[OMR2018] Omran, Mohamed, et al. "Neural body fitting: Unifying deep learning and model based human pose and shape estimation." *IEEE International Conference on 3D Vision*, 2018.
[KOL2019a] Kolotouros, Nikos, Georgios Pavlakos, and Kostas Daniilidis. "Convolutional mesh regression for single-image human shape reconstruction." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[LAS2017] Lassner, Christoph, et al. "Unite the people: Closing the loop between 3d and 2d human representations." *IEEE Conference on Computer Vision and Pattern Recognition,* 2017.
[PAV2019] Pavlakos, Georgios, et al. "Expressive body capture: 3d hands, face, and body from a single image." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.

Bibliography

[KOL2019b] Kolotouros, Nikos, et al. "Learning to reconstruct 3D human pose and shape via model-fitting in the loop." *IEEE International Conference on Computer Vision,* 2019.

[OSM2020] Osman, Ahmed AA, Timo Bolkart, and Michael J. Black. "Star: Sparse trained articulated human body regressor." *European Conference on Computer Vision*, 2020.

[ROG2017] Rogez, Gregory, Philippe Weinzaepfel, and Cordelia Schmid. "Lcr-net: Localization-classification-regression for human pose." *IEEE Conference on Computer Vision and Pattern Recognition,* 2017.

[ROG2019] Rogez, Gregory, Philippe Weinzaepfel, and Cordelia Schmid. "Lcr-net++: Multi-person 2d and 3d pose detection in natural images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol.42 no. 5, pp. 1146-1161, 2019.

[BEN2020] Benzine, Abdallah, et al. "Pandanet: Anchor-based single-shot multi-person 3d pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2020.

[NIE2019] Nie, Xuecheng, et al. "Single-stage multi-person pose machines." *IEEE International Conference on Computer Vision,* 2019.

[MEH2018] Mehta, Dushyant, et al. "Single-shot multi-person 3d pose estimation from monocular rgb." *IEEE International Conference on 3D Vision*, 2018.

[BUR2013] Burenius, Magnus, Josephine Sullivan, and Stefan Carlsson. "3d pictorial structures for multiple view articulated pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2013.

[DON2021] Dong, Zijian, et al. "Shape-aware multi-person pose estimation from multi-view images." *IEEE International Conference on Computer Vision,* 2021.

[HUA2020] Huang, Congzhentao, et al. "End-to-end dynamic matching network for multi-view multi-person 3d pose estimation." *European Conference on Computer Vision*, 2020.

[ZHA2021] Zhang, Jianfeng, et al. "Direct multi-view multi-person 3d pose estimation." *Advances in Neural Information Processing Systems,* 2021.

[KAD2017] Kadkhodamohammadi, Abdolrahim, et al. "A multi-view RGB-D approach for human pose estimation in operating rooms." *IEEE Winter Conference on Applications of Computer Vision*, 2017.

[YU2018] Yu, Tao, et al. "Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.

[ZHI2020] Zhi, Tiancheng, et al. "Texmesh: Reconstructing detailed human texture and geometry from rgb-d video." *European Conference on Computer Vision*, 2020.

Bibliography

[HUA2020] Huang, Fuyang, et al. "Deepfuse: An imu-aware network for real-time 3d human pose estimation from multi-view image." *IEEE Winter Conference on Applications of Computer Vision,* 2020.
[VON2018] Von Marcard, Timo, et al. "Recovering accurate 3d human pose in the wild using imus and a moving camera." *European Conference on Computer Vision,* 2018.
[ZHA2020b] Zhang, Zhe, et al. "Fusing wearable imus with multi-view images for human pose estimation: A geometric approach." *IEEE Conference on Computer Vision and Pattern Recognition,* 2020.
[ZHA2019b] Zhao, Mingmin, et al. "Through-wall human mesh recovery using radio signals." *IEEE International Conference on Computer Vision,* 2019.
[ZHA2018] Zhao, Mingmin, et al. "RF-based 3D skeletons." *Conference of the ACM Special Interest Group on Data Communication,* 2018.
[ZHA2018b]Zhao, Mingmin, et al. "Through-wall human pose estimation using radio signals." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.
[JOH2010] Johnson, Sam, and Mark Everingham. "Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation." *BMVC*. Vol. 2. No. 4, 2010.
[JOH2011] Johnson, Sam, and Mark Everingham. "Learning effective human pose estimation from inaccurate annotation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2011.
[SAP2013] Sapp, Ben, and Ben Taskar. "Modec: Multimodal decomposable models for human pose estimation." *IEEE Conference on Computer Vision and Pattern Recognition,* 2013.
[TOM2014] Tompson, Jonathan J., et al. "Joint training of a convolutional network and a graphical model for human pose estimation." *Advances in Neural Information Processing Systems,* 2014.
[AND2014] Andriluka, Mykhaylo, et al. "2d human pose estimation: New benchmark and state of the art analysis.“ *IEEE Conference on Computer Vision and Pattern Recognition,* 2014.
[LIN2014] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European Conference on Computer Vision*, 2014.
[LI2019] Li, Jiefeng, et al. "Crowdpose: Efficient crowded scenes pose estimation and a new benchmark." *IEEE Conference on Computer Vision and Pattern Recognition,* 2019.
[ZHA2013] Zhang, Weiyu, Menglong Zhu, and Konstantinos G. Derpanis. "From actemes to action: A strongly-supervised representation for detailed action understanding." *IEEE International Conference on Computer Vision,* 2013.

Bibliography

[JHU2013] Jhuang, Hueihan, et al. "Towards understanding action recognition." *IEEE International Conference on Computer Vision,* 2013.
[AND2018] Andriluka, Mykhaylo, et al. "Posetrack: A benchmark for human pose estimation and tracking." *IEEE Conference on Computer Vision and Pattern Recognition,* 2018.
[SIG2010] Sigal, Leonid, Alexandru O. Balan, and Michael J. Black. "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion." *International Journal of Computer Vision* vol.87 no. 1, pp. 4-27, 2010.
[ION2013] Ionescu, Catalin, et al. "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments." *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 36, no. 7, pp. 1325-1339, 2013.
[JOO2015] Joo, Hanbyul, et al. "Panoptic studio: A massively multiview system for social motion capture." *IEEE International Conference on Computer Vision,* 2015.
[MEH2017] Mehta, Dushyant, et al. "Monocular 3d human pose estimation in the wild using improved cnn supervision." *IEEE International Conference on 3D Vision*, 2017.
[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.
[PIT2017] I. Pitas, “Digital video processing and analysis” , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, “Digital Video and Television” , Createspace/Amazon, 2013.
[NIK2000] N. Nikolaidis and I. Pitas, “3D Image Processing Algorithms”, J. Wiley, 2000.
[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.
[TER] https://www.terabee.com
[HAC] https://www.hackster.io

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr