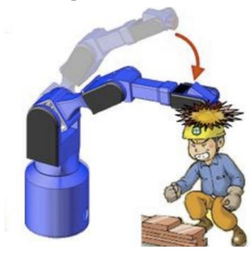
A comparison of human skeleton extractors for real-time human-robot interaction

Wanchen LI[1]
Robin Passama[1]
Vincent Bonnet[2]
Andrea Cherubini[1]

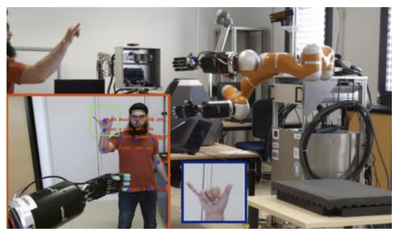
[1] LIRMM UM/ CNRS, Montpellier, France [2] LAAS-CNRS, Toulouse, France

Motivation:

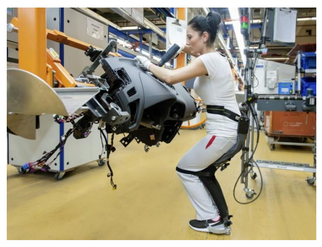
Safety insurance



Context understanding



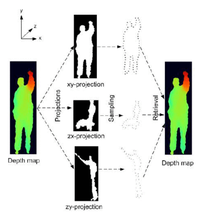
Ergonomic adaptation



Multi-modal data for activity recognition:

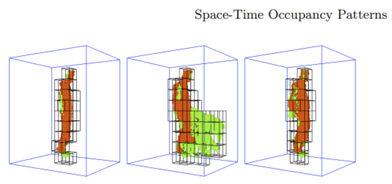
3D silhouette

- Limited for atomic actions recognition
- Occlusion will degrade the extraction.



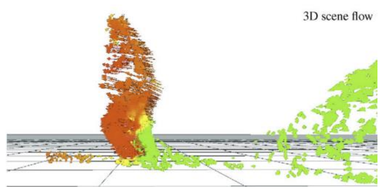
Space-time occupancy pattern

- Pattern can be sparse.
- High computation cost



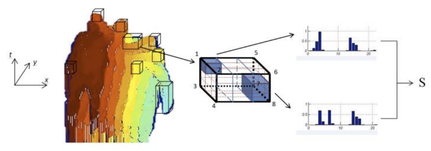
3D optical flow

- Computationally costly
- Not suitable for real time application



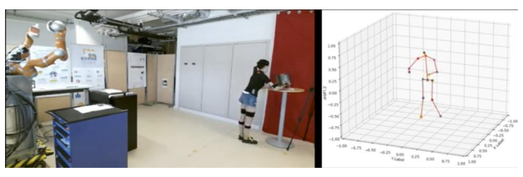
Local spatial temporal features

+ No need for segmentation or tracking
- The feature is view-dependent and local.



Skeletal data

+ Invariant to the camera location, subject appearance and human body size
+ Can be combined with biomechanical model
- Current estimation algorithm is not perfect.



Skeleton extraction frameworks comparison:

General functionality:

Skeleton extractors	Framework	Output	Technique	Specialty
Detectron2	Pytorch	17 key-points	Segmentation on each key-point, one-step estimation	Segmentation
MediaPipe	Tensorflow	33 key-points with 3d inference	Two-step estimation, region of interest detector + joint position tracker	Joint position tracking
YOLOv7	Pytorch	17 key-points	One-step estimation	Occlusion does not influence detection
ALPHA POSE	Pytorch	17/26/136 key-points	Two-step estimation	Pose aware identity mechanism
OpenPose	Caffe	15/18/25/67/137 key-points	Two-step estimation, Using part affinity fields	Direct C++ API is available 3D estimation is possible upon multiple synchronized camera views

Performance evaluation:

Skeleton extractors	Identification	Multi-person detection	Foot keypoints	Hand keypoints	Facial keypoints	Easy C++ interfacing	Robustness with respect to motion	GPU integration	Framerate
Detectron2	×	✓	×	×	ears,eyes,nose	×	✓	✓	3.57 fps
Mediapipe	✓	×	✓	✓	ears,eyes, nose,mouth	×	×	✓on Linux	17 fps
YOLOv7	×	✓	×	×	ears,eyes,nose	×	✓	✓	11.04 fps
Alphapose	✓	✓	×	×	ears,eyes, nose,mouth	×	✓	✓	9.74 fps
Openpose	×	✓	✓	✓	ears,eyes, nose,mouth	✓	✓	✓	9.91 fps

Future work:

- Quantitative comparison of 5 frameworks' outputs
- Using OpenPose library with human biomechanical model to estimate human skeleton on image inputs in real-time
- Human activity classification based on joint space information

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871237.