

Motivation & Contribution

- The IPT dataset aims to facilitate the application of modern computer vision methods to privacy-critical applications such as:
 - Ambient Assisted Living (AAL)
 - Security
- Privacy is protected using a depth sensor as an anonymizing imaging method. In many cases behavior analysis does not need to know who is acting but rather what they are doing.
- We present:
 - A new public dataset for identity preserved human detection and tracking in 2d or 3d.
 - Baseline results for person detection and tracking.

Dataset Overview

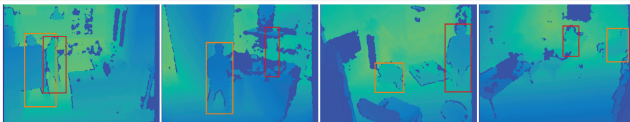
- 10 sequences featuring:
 - Indoor environments
 - Both scripted and natural behavior
 - Frequent person-person and person-object occlusions
- 72k frames, static depth sensor at 640×480 resolution, 30 fps.
- Split by sequence into training (70%), validation (20%) and test sets (10%).



Representative sample frames for each recorded sequence

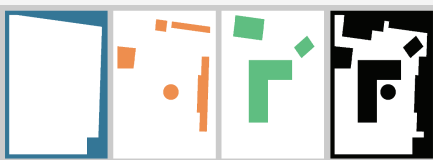
Labeling

- 3d location of actors
 - Consistent actor ID across all sequences
- 2d bounding boxes
 - Inferred from 3d locations



Inferred 2d bounding boxes

- Semantic room layout layers
 - Exterior space (walls, outside space)
 - Obstructed space (furniture, appliances, ...)
 - Semi-obstructed space (chairs, beds, ...)



Semantic room layout layers: Decomposed into exterior space (blue), obstructed space (orange), semi-obstructed space (green) and combined (black)

- Sensor pose
- Tracking ground truth in the MOT Challenge format

Results

- 2d detection baseline using a YOLOv3 model.
- Improved model performance using an efficient background model.

Model	Average Precision	
	Validation	Test
YOLOv3	77.2%	89.2%
YOLOv3 + BG Model	85.2%	89.6%
YOLOv3 Tiny	71.9%	82.0%

Baseline person detection scores on validation and test set

- 2d tracking baseline using the SORT and DeepSORT tracking algorithm.
- Evaluation using full MOT Challenge metrics

Mode	MOT Accuracy	Identity F1
SORT	76.6%	14.5%
DeepSORT	77.2%	17.5%
DeepSORT Short Tracks	77.3%	58.8%

Baseline tracking results using MOT Challenge metrics

Conclusion

- We presented the IPT dataset and demonstrated baseline human detection and tracking results.
- The depth modality can serve as an alternate imaging mode in privacy-sensitive applications.

Future work

- Future versions of IPT will increase in size, improve instance labeling and add annotations for human interactions.
- Multimodal behavior analysis based on depth/thermal fusion.