RICA: Robocentric Indoor Crowd Analysis Dataset

Viktor Schmuck and Oya Celiktutan
{viktor.schmuck; oya.celiktutan}@kcl.ac.uk

Centre for Robotics Research, Department of Engineering
King's College London

Introduction

As robots become increasingly prevalent, a large number of practical applications demand that they autonomously navigate in indoor spaces, recognise and approach groups or individuals, and through human-robot interaction assist them. However, available datasets for indoor crowd analysis from a robot's perspective are really scarce and limited. This paper introduces the first multisensory, egocentric dataset from a robot's point of view (robocentric) for group detection and F-formation recognition in crowded spaces.

Data Collection

Recorded at a reception-style semi-public event in an indoor environment with Toyota's Human Support Robot (HSR) with:

- RGB-D camera – ASUS Xtion PRO LIVE
- Wide angle camera – Nippon Chemi-Con NCM13-J-02
- LIDAR sensor – Laser measuring range sensor (UST-20LX)
- IMU data
- Joint position data

Data

- LIDAR data: 963 samples from -2.098 to 2.098 radians per sample
- Camera-to-subject distance: 0.1-25m
- In each frame: 1 to 8 people with an average of 3.92
- Annotated with the Actanno annotation tool [3] producing group-level (40366 images) and person-level (8148 images) annotations.

Comparison to JRDB

Sensor Type	Num. of Samples		Average Framerate	
	RICA	*JRDB*	*RICA*	*JRDB*
RGB camera	43,060	57,713	10.542	15.116
Depth camera	39,909	57,714	9.771	15.116
Wide-angle camera	17,877	58,313	4.377	15.273
Joint position	63,569	38,476	15.563	10.078
IMU	127,324	74,234	31.172	19.443
LIDAR	50,926	56,844	12.468	14.888

Fig. 1: Summary of the collected - unfiltered - data using robot's on-board sensors compared to the relevant recordings of JRDB [2].

Acknowledgement

We thank Toyota Motor Europe for providing the Toyota HSR robot as a development platform.

Toyota HSR

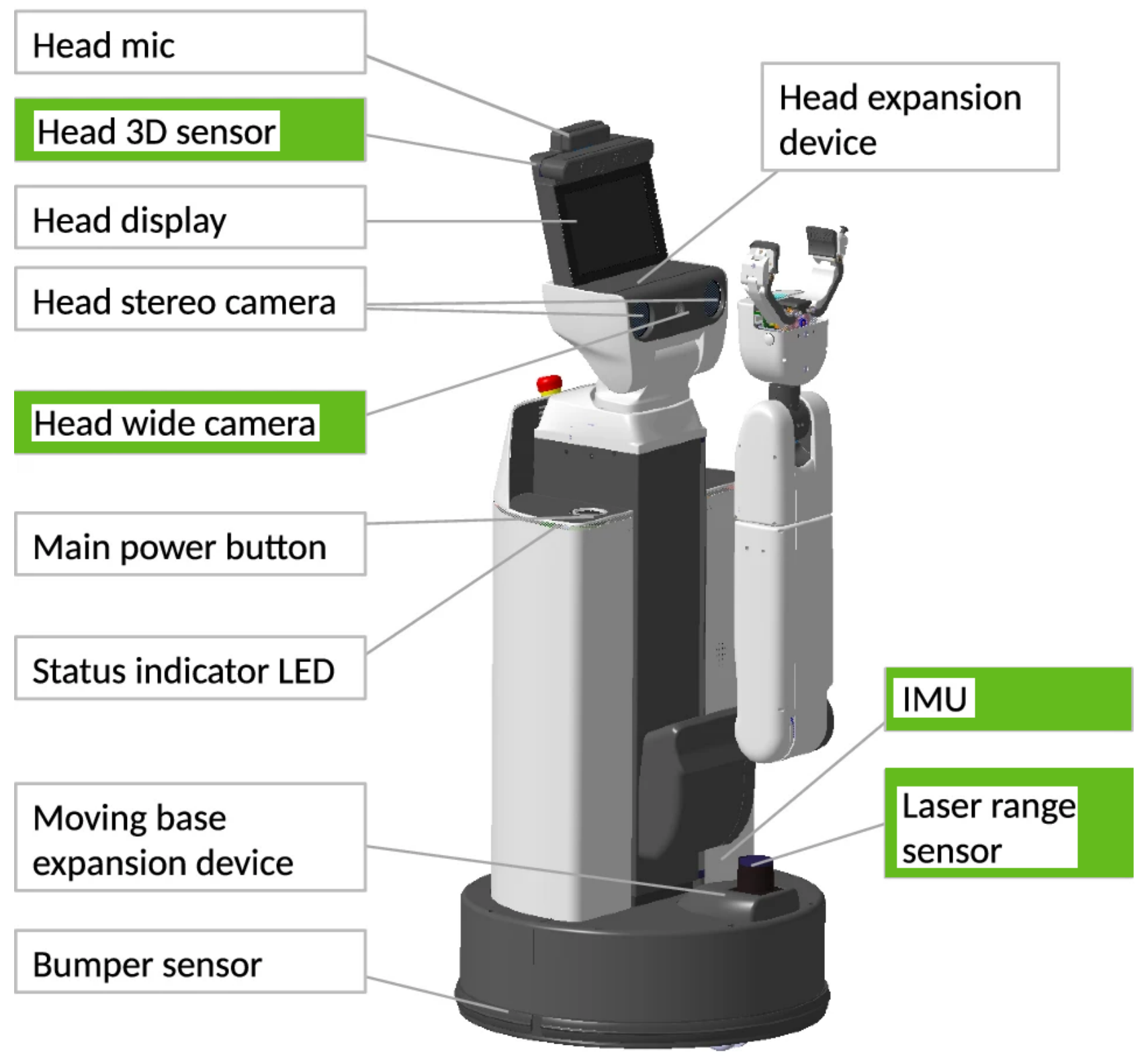


Fig. 2: Sensors of the Toyota HSR robot [4].

RGB and Wide-Angle Samples

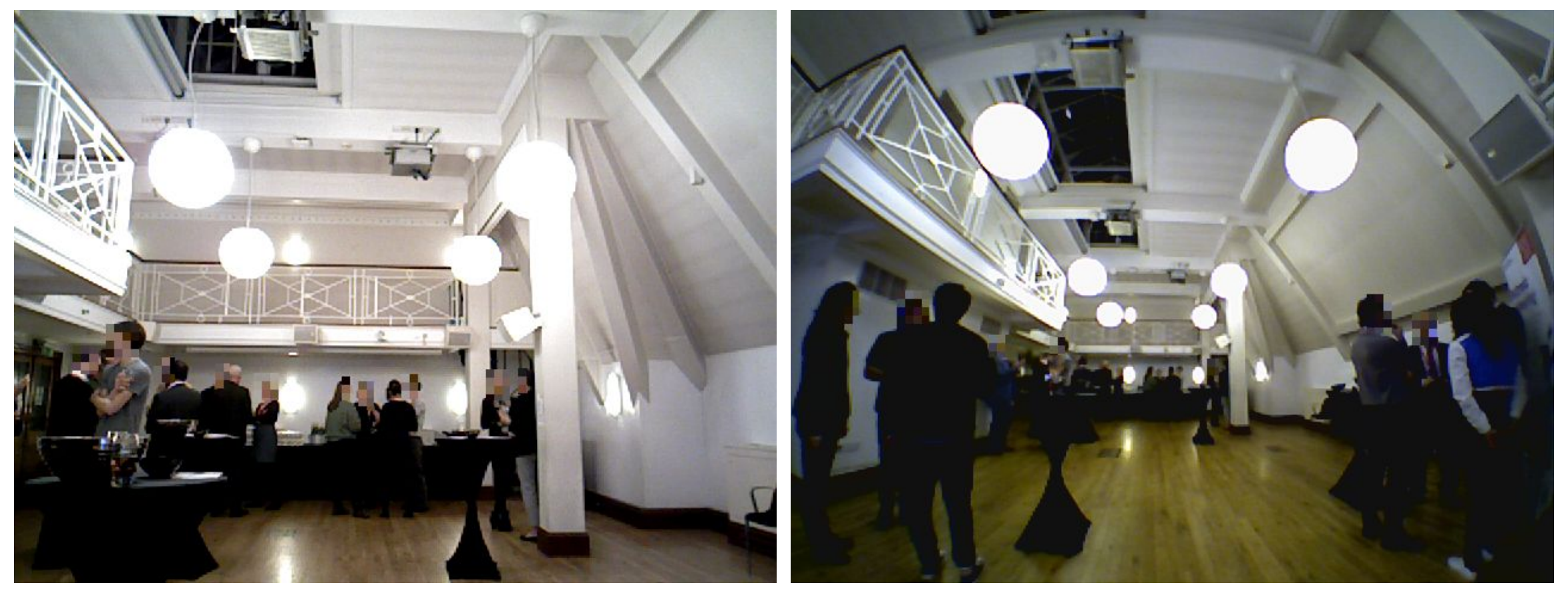


Fig. 3: RGB (left) and Wide-angle camera (right) samples from the RICA dataset.

F-formations

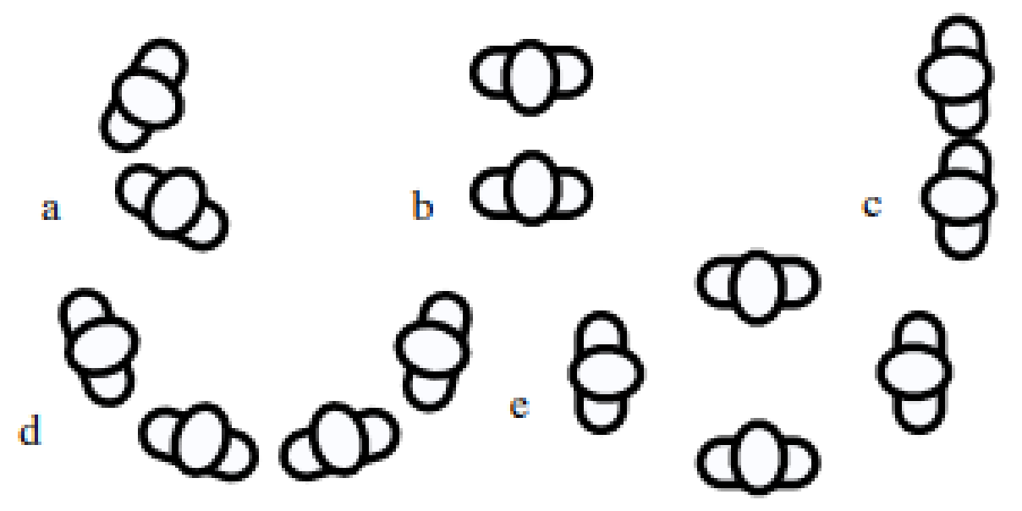


Fig. 4: Group-level annotations also include group formations i.e. (a) L-arrangement, (b) face-to-face, (c) side-by-side, (d) semi-circular, (e) rectangular [1]

Annotated sample

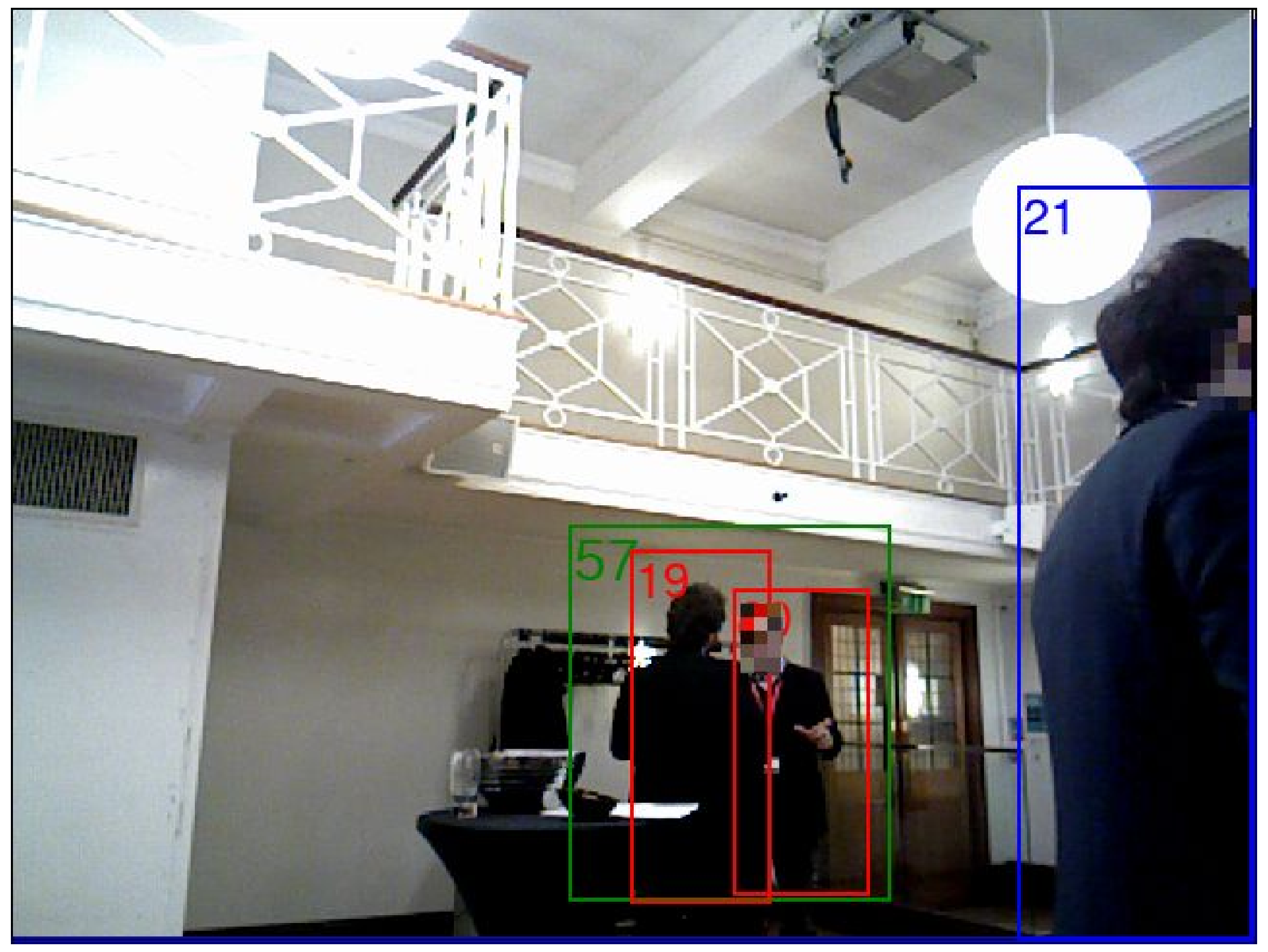


Fig. 5: An annotated image recorded with the RGB camera, showing a person (ID 21) not belonging to any group, and two individuals (IDs 19-20) belonging to group ID 57, where the group formation of group ID 57 is annotated as *face-to-face*.

Evaluation

Benchmarking three methods on our RICA dataset, without fine-tuning:

- Histogram of Oriented Gradients (HOG) combined with non-maxima suppression (NMS)
- MobileNet-SSD (SSD) – trained on MS-COCO, and then fine-tuned on VOC0712 – with centroid tracking
- YOLO – trained on MS-COCO

Measured the intersection over union (IOU) values of bounding boxes against ground truth (GT).

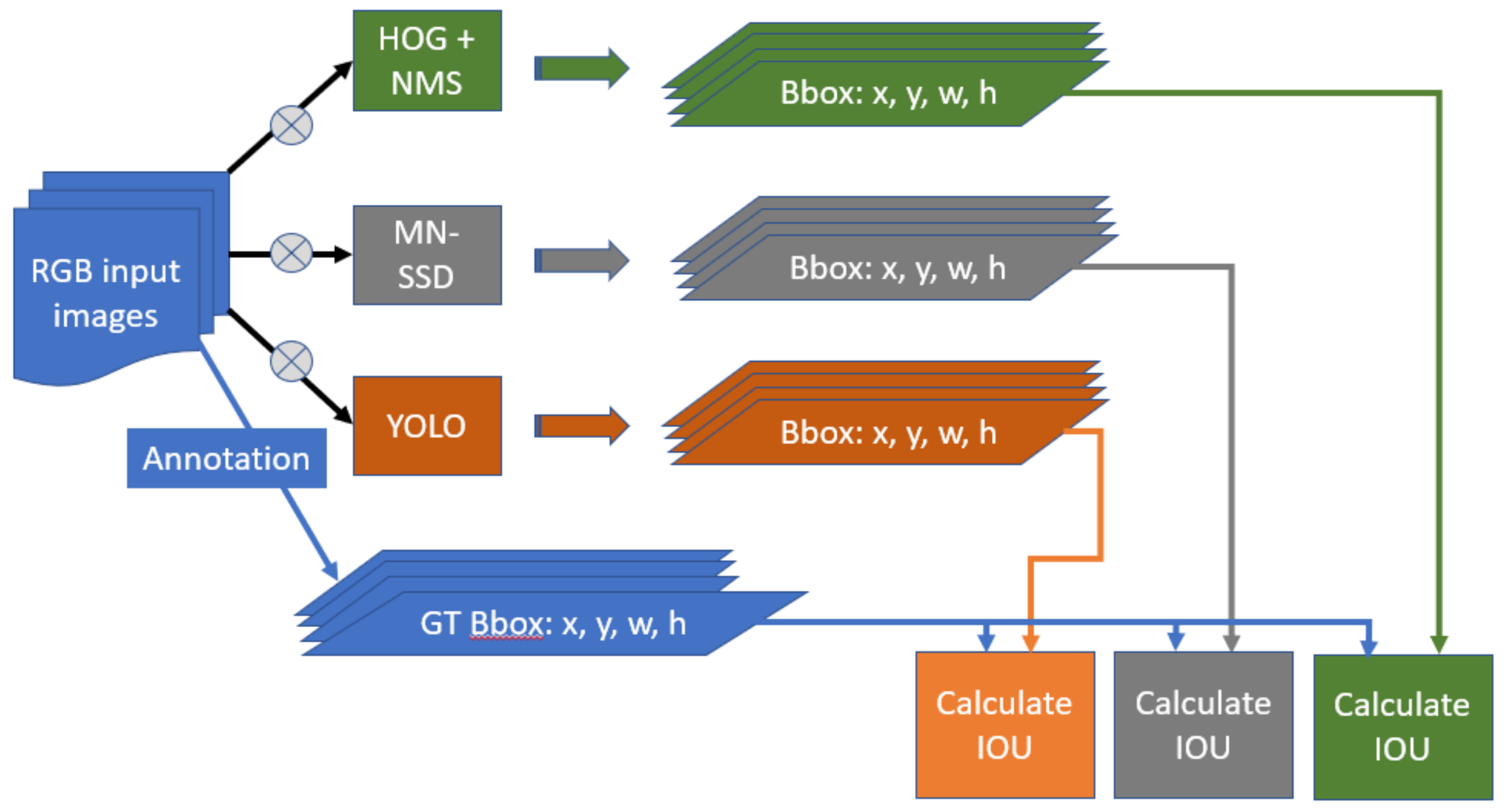
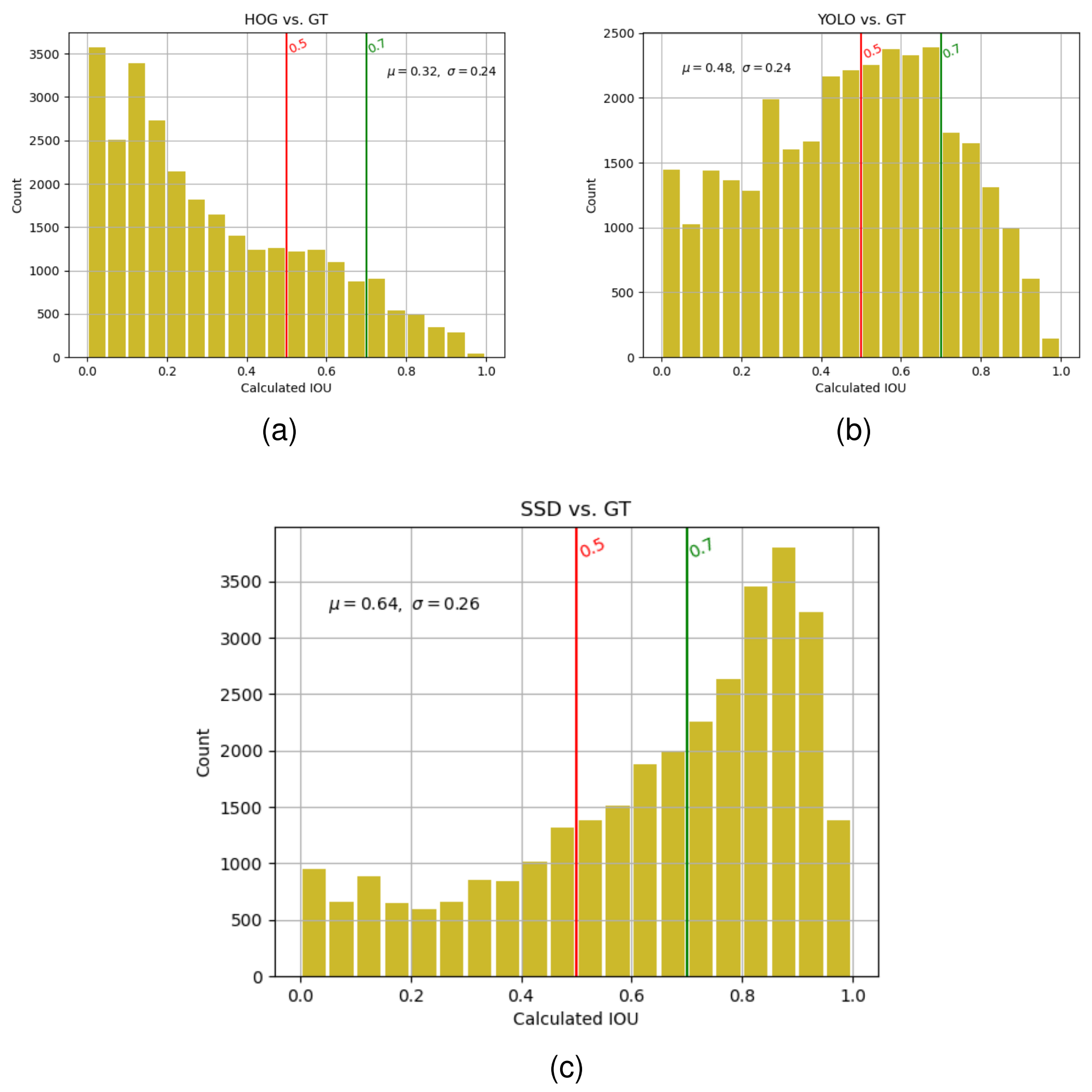


Fig. 6: Flowchart of the evaluation method.

Results



(a) (b) (c)

Fig. 7: Histograms of IOU values for GT vs. (a) HOG; (b) YOLO; and (c) SSD. Red lines – min. IOU for True Positive detection. Green lines – IOU for desired detection.

Conclusion

- The best mean IOU score (0.64) was obtained with the SSD detector.
- Human detectors fall short and sometimes are unable to detect any humans. – 11% of all images in case of HOG+NMS and 0.7% for both SSD and YOLO.

Dataset access

The dataset will be made available at https://sairlab.github.io/rica/

Future Work

Design an unsupervised approach to group detection in indoor crowded scenes.

References

[1] P. Marshall, Y. Rogers, and N. Pantidi. Using F-formations to analyse spatial patterns of interaction in physical environments. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 445–454, 2011.

[2] R. Martín-Martín, H. Rezatofighi, A. Shenoi, M. Patel, J. Gwak, N. Dass, A. Federman, P. Goebel, and S. Savarese. *JRDB: A Dataset and Benchmark for Visual Perception for Navigation in Human Environments*. 2019.

[3] C. Wolf, E. Lombardi, J. Mille, O. Celiktutan, M. Jiu, E. Dogan, G. Eren, M. Baccouche, E. Dellandréa, C.-E. Bichot, C. Garcia, and B. Sankur. Evaluation of video activity localizations integrating quality and quantity measurements. *Computer Vision and Image Understanding*, 127:14–30, Oct. 2014.

[4] T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase. Development of Human Support Robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(1):4, Apr. 2019.