TABLE VII
PROCESSING TIME OF MODULES

Module	Avg. processing time [ms]	
	800x600 px	640x480 px
Face detector	172.4	99.7
Upper-body / HOG detector	408.8/423.0	242.3/225.4
Motion / Leg detector	3.1/1.0	1.6/1.0
FPDW (offline)	535.4	359.0
PartHOG (offline)	4975.7	2864.7
People Tracker	0.2	0.2

to bounding boxes is error prone, especially in distance estimation, because the camera is looking horizontally. These reasons need further investigations in future work. On the other hand, the combination of FPDW+laser and our real-time tracker achieve higher performances than the single FPDW detector, especially when people sit. The partHOG detector (Tab. VI(b)) achieved similar performance as the partHOG tracker, because the detector processed every frame in an offline evaluation.

All presented results were produced using the same set of parameter of the tracker and the detection modules. We scaled down the original image resolution of the data sets to 640x480 to increase the computational performance. A performance evaluation of the detection modules of the people tracker can be found in Tab. VII. From there it becomes obvious that the face and HOG modules do not process every frame but are set to run every 500 ms. The complete tracking system runs in real-time and is configured to consume 60% of the robot's on-board CPU (Intel i7-620M quad core processor) leaving enough space for the other required modules of the robot [1].

V. CONCLUSION

We presented a real-time, multi-modal people tracking system for mobile companion robots, that tracks walking people and is able to track people in sitting poses, if there are enough detector inputs. The system is evaluated on different data sets with increasing difficulty. Furthermore, we compared the performance to offline state-of-the-art people detectors like FPDW and partHOG and trackers based on these detectors. Our real-time version of the people tracker achieves better results than a tracker based on the FPDW detector and the pure detector, particularly when people sit. Best results are achieved when using the partHOG detector, which, unfortunately, is far from being real-time capable at the moment. Yet, the only moderate performances of all tested trackers show that more research is necessary to track people in home environments - especially for non-upright poses. To achieve the long-term goal of autonomous companion robots that support the elderly, we need to enhance current person detection algorithms. The modules for face and upper body detection are not robust enough to detect people in sitting postures or given occlusion. Using the real-time capable FPDW detector in the combined tracker could help to raise up-right posture performance. Real-time implementations of part based detection concepts like partHOG or poselets [7] that handle occlusion and multiple postures would greatly improve detection and tracking performance. Therefore, a major challenge lies in the development of real-time capable methods for detection of people in different poses, like sitting and lying. The Kinect sensor could help to achieve this goal.

REFERENCES

[1] H.-M. Gross et al., "Further progress towards a home robot companion for people with mild cognitive impairment," in *IEEE Trans. Syst., Man, Cybern.* Seoul, South Korea: IEEE, 2012, pp. 637–644.

[2] C. Pantofaru, "The Moving People, Moving Platform Dataset," http://bags.willowgarage.com/downloads/people_dataset/.

[3] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: an evaluation of the state of the art." *Transactions on pattern analysis and machine intelligence*, vol. 34, no. 4, pp. 743–761, 2012.

[4] P. Dollár, S. Belongie, and P. Perona, "The Fastest Pedestrian Detector in the West," in *British Machine Vision Conference*, 2010.

[5] R. Benenson and M. Mathias, "Pedestrian detection at 100 frames per second," *Computer Vision and Pattern Recognition*, 2012.

[6] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2241–2248.

[7] L. Bourdev, S. Maji, T. Brox, and J. Malik, "Detecting people using mutually consistent poselet activations," in *ECMR*, 2010, pp. 168–181.

[8] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2002.

[9] N. Bellotto and H. Hu, "A Bank of Unscented Kalman Filters for Multimodal Human Perception with Mobile Service Robots," *International Journal of Social Robotics*, vol. 2, no. 2, pp. 121–136, 2010.

[10] G. Cielniak, T. Duckett, and A. J. Lilienthal, "Data association and occlusion handling for vision-based people tracking by mobile robots," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 435–443, 2010.

[11] K. O. Arras, O. M. Mozos, and W. Burgard, "Using Boosted Features for the Detection of People in 2D Range Data," in *International Conference on Robotics and Automation*, 2007, pp. 3402–3407.

[12] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, "A mobile vision system for robust multi-person tracking," in *Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[13] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool, "Coupled object detection and tracking from static cameras and moving vehicles." *TPAMI*, vol. 30, no. 10, pp. 1683–98, 2008.

[14] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *TPAMI*, vol. 33, no. 9, pp. 1820–1833, 2011.

[15] D. Mitzel, P. Sudowe, and B. Leibe, "Real-Time Multi-Person Tracking with Time-Constrained Detection," in *Proceedings of the British Machine Vision Conference*, 2011, pp. 104.1–104.11.

[16] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[17] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *CVPR*, 2008, pp. 1–8.

[18] M. Everingham et al., "The PASCAL Visual Object Classes Challenge 2009 Results," http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html.

[19] A. L. Simpson et al., "Uncertainty propagation and analysis of image-guided surgery," in *SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling Conference*, vol. 7964, 2011.

[20] E. Einhorn, T. Langner, R. Stricker, C. Martin, and H.-M. Gross, "Mira - middleware for robotic applications," in *In IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 2591–2598.

[21] L. Chen, P. Armabel, and R. Mehra, "Estimation Under Unknown Correlation: Covariance Intersection Revisited," *IEEE Transactions on Automatic Control*, vol. 47, pp. 1879–1882, 2002.

[22] M. Volkhardt, S. Müller, C. Schröter, and H.-M. Gross, "Playing Hide and Seek with a Mobile Companion Robot," in *Proc. 11th IEEE-RAS Int. Conf. on Humanoid Robots*, 2011, pp. 40–46.

[23] D. F. Sebastian Thrun, Wolfram Burgard, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The Mit Press, 2005.

[24] D. R. Carl Vondrick, Donald Patterson, "Efficiently scaling up crowd-sourced video annotation," in *IJCV*, 2012.

[25] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics," in *EURASIP Journal on Image and Video Processing*, 2008, pp. 1–10.