Supplementary Material: Pedestrian Models for Autonomous Driving Part I: low level models, from sensing to tracking

Fanta Camara[1,2], Nicola Bellotto[2], Serhan Cosar[2], Dimitris Nathanael[3], Matthias Althoff[4], Jingyuan Wu[5], Johannes Ruenz[5], André Dietrich[6] and Charles W. Fox[1,2,7]

I. Quality of citations

These linked papers (Part I and II) review over 450 papers from high quality journals and conferences such as *CVPR, ICRA, PAMI, IROS, ITSC, ECCV, IV*. It is common in Computer Science fields including machine vision and machine learning for conferences to be considered higher quality or similar quality to journals, while psychology and sociology fields typically consider journals to be more authoritative. The following figures give some ideas about the quality of the cited papers.

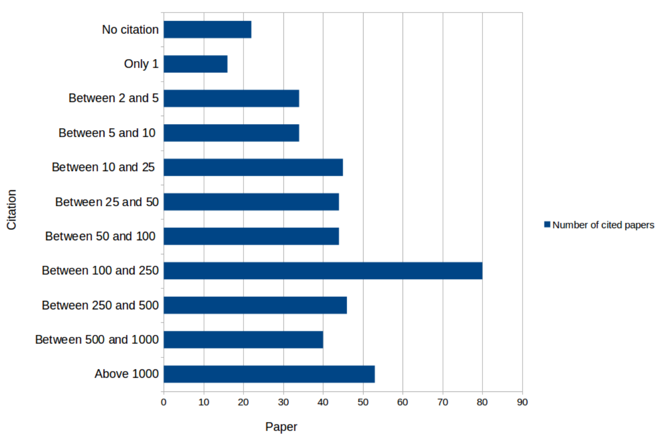

Fig. 1. Number of citations per paper

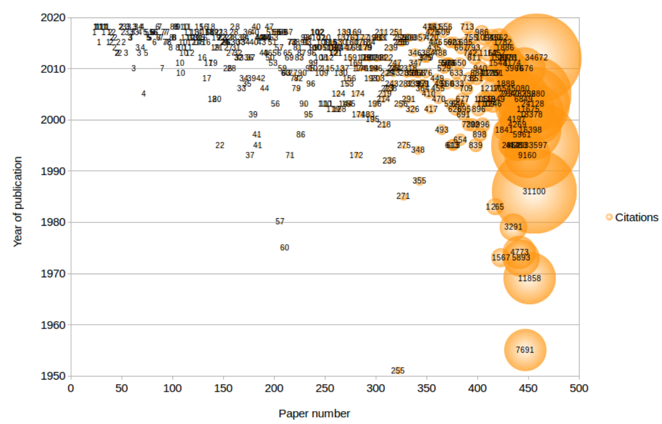

Fig. 2. Number of citations per paper and per year of publication

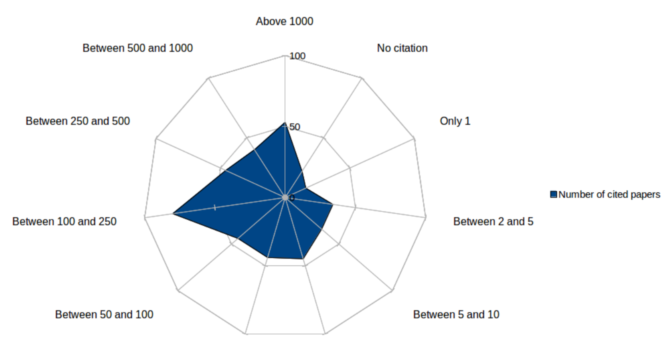

Fig. 3. Number of citations per paper

This project has received funding from EU H2020 interACT: Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments under grant agreement No 723395.

[1] Institute for Transport Studies (ITS), University of Leeds, UK
[2] Lincoln Centre for Autonomous Systems, University of Lincoln, UK
[3] School of Mechanical Engineering, Nat. Tech. University of Athens
[4] Department of Computer Science, Technische Universität München
[5] Robert Bosch GmbH, Germany
[6] Chair of Ergonomics, Technische Universität München (TUM), Germany
[7] Ibex Automation Ltd., UK

Manuscript received March, 11, 2019; revised October, X, 2019

II. Summary of pedestrian recognition models

TABLE I
SUMMARY OF THE RECOGNITION MODELS

Study/Paper	Sensor input	Method	Pedestrian Recognition Type	Additional Info	SAE Level
Cao et al. [1] [2]	Images	CNN model with Part Affinity Fields (PAF)	Pose estimation	OpenPose: open source software	Level 2
Shotton et al. [3]	Motion capture and synthetic data	Body parts representation model	3D human pose estimation		Level 2
Iqbal et al. [4]	Video data	Graphical model	Pose estimation and tracking	Release of PoseTrack a new dataset	Level [2,3]
Tompson et al. [5]	Monocular images	Deep CNN model with Markov Random Field	Pose estimation		Level 2
Ma et al. [6]	Images	CNN model	Body heading	No annotations needed	Level 2
Kohari et al. [7]	Video	CNN model	Body orientation	Service robot	Level 2
Darrell et al. [8]	Images	Statistical model	Head direction	From a mobile robot	Level [2,3]
Schulz et al. [9]	Grayscale images	Multi-classifiers	Head pose		Level 2
Microsoft Azure API [10]	Video		Head pose and facial features	Commercial product	Level 2
Benfold et al. [11]	Video	HOG and colour features	Gaze tracking		Level 2
Baltrusaitis et al. [12]	Video	Deep learning model	Head pose and eye-gaze estimation	Openface: open source software	Level 2
Cornejo et al. [13] [14]	Images	Principal Component Analysis + Gabor wavelets or CENTRIST features	Emotion recognition		Level 2
Cambria et al. [15] [16]	Images	Review paper	Sentiment analysis		
Poria et al. [17]	Videos	CNN model with recurrent multilayer kernel learning	Emotion recognition		Level 2
Horng et al. [18]	Video	Dynamic template matching	Driver fatigue detection		Level 2
Denuyl et al. [19]	Video		Face expression recognition	FaceReader: commercial product	Level 2
Ahmed et al. [20]	Images	Deep neural networks	Re-identification		Level 2
Zheng et al. [21]	Images	Bag of Words model	Re-identification		Level 2
Zheng et al. [22]	Images	CNN model	Re-identification	Unlabeled images	Level 2
Li et al. [23]	Images	Filter airing neural network (FPNN) model	Re-identification	Occlusion handling	Level 2
Chen et al. [24] [25] [26]	Images	Hidden Markov model (HMM)	Gesture recognition		Level 2
Freeman et al. [27]	Images	Orientation histograms	Gesture recognition	10 different hand gestures recognition	Level 2
Ren et al. [28]	Images	Template matching with Finger-earth Mover's Distance (FEMD)	Gesture recognition		Level 2
Quintero et al. [29]	Images	Hidden Markov models (HMM)	Body language Recognition		Level [2,3]
Wang et al. [30]	Images	Background subtraction + PCA	Body language Recognition		Level 2
Chaaraoui et al. [32]	Videos	Contour points	Activity recognition	Real-time method	Level 2
Dollár et al. [33]		Spatio-temporal features	Activity recognition		Level 2
Vail et al. [34]	Videos	Hidden Markov models and Conditional random field	Activity recognition		Level 2
Liu et al. [35]	RGB data	Coupled conditional random field	Activity recognition		Level 2
Coppola et al. [36]	RGB-D data	Dynamic Bayesian mixture model (DBMM)	Activity recognition		Level 2

III. Summary pedestrian tracking models

TABLE II: Summary of pedestrian tracking models

Study/Paper	Input/Evaluation	Method	Tracking Models	Additional Info	SAE Level
Del Pino et al. [37]	Low resolution LiDAR data	Multi-Hypothesis EKF (MHEKF)	Point Tracking		Level 2
Bellotto et al. [38]	Robot with laser and camera	EKF, UKF, SIR Particle filter	Point Tracking	Trade-off between performance and computation cost	Level 2
Arulampalam et al. [39]	Example	Particle filter implementations	Point Tracking		Level 2
Fen et al. [40]	Video data	Color histogram based particle filter	Point Tracking		Level 2
Jurie et al. [41]	Video data	Template matching with SSD	Kernel-based Human Tracking	Real-time method and robustness to occlusions and illuminations	Level 2
Lipton et al. [42]	Video data	Frame differencing + Template matching	Kernel-based Human Tracking	Real-time method	Level 2
Kaneko et al. [43]	Image sequences	Template matching with a feature selection method	Kernel-based Human Tracking		Level 2
Comaniciu et al. [44]	Moving camera data	Mean-shift algorithm with Bhattacharyya coefficient	Kernel-based Human Tracking		Level 2
Collins et al. [45]	Video data	Mean-shift algorithm with 2d blob tracking	Kernel-based Human Tracking		Level 2
Tao et al. [46]	Airborne vehicle tracking system	Dynamic layering method + MAP using EM algorithm	Kernel-based Human Tracking		Level 2
Yalcin et al. [47]	Image sequences	Layering method with optical flow	Kernel-based Human Tracking		Level [2,3]
Geiger et al. [48]	Image sequences	Contour matching method based on Dynamic programming	Tracking pedestrian body state		Level 2
Techmer et al. [49]	Real-world images	Contour tracking with distance transformations of contour images	Tracking pedestrian body state		Level 2
Baumberg [50] Yilmaz [51]	Image sequences	Dynamic Kalman filter with active shape model	Tracking pedestrian body state	Method sensitive to initialization	Level 2
Adam et al. [52]	Image sequences	Region color histogram method	Tracking pedestrian body state	Occlusion and pose changes handling	Level 2
Meyer et al. [53]	Image sequences	Recursive algorithm using image regions information	Tracking pedestrian body state		Level 2
Collins et al. [54]	Gait databases: CMU, MIT, UMD, USH	Silhouette based model	Tracking pedestrian body state	To identify people from their body and gait	Level 2
Schwarz et al. [55]	Kinect data	Graph method with skeleton fitting	Tracking pedestrian body state	Full-body tracker	Level 2
Sinthanayothin et al. [56]	Kinect data	Skeleton tracking	Tracking pedestrian body state	Review paper	Level 2
Konstantinova et al. [57]	5 test matrices	Global Nearest Neighbor with Munkres algorithm	Multi-Target Tracking (MTT)		Level [2, 3, 4]
Azari et al. [58]	IBM, PETS2000 and PETS2001 databases	Kalman filter with Global Nearest Neighbor (GNN)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3,4]
Reid et al. [59]	Monte Carlo simulation	Iterative algorithm with Multi-Hypothesis Tracking(MHT)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]
Luber et al. [60]	Two datasets collected in indoor and outdoor environments	Social force with Multiple Hypothesis Tracking (MHT)	Multi-Target Tracking (MTT)		Level [3,4]
Kim et al. [61]	PETS and MOTChallenge benchmarks	Multiple Hypothesis Tracking (MHT)	Multi-Target Tracking (MTT)		Level [2,3]
Zhou et al. [62]	Computer simulations	Joint probabilistic data association filter (JPDAF) with a depth-search approach	Multi-Target Tracking (MTT)		Level [2,3]
Chen et al. [63]	Video data	Contour based tracker with JPDAF and HMM	Multi-Target Tracking (MTT)	Real-time method	Level [2,3]
Liu et al. [64]	Simulations and real robot	Sample-based JPDAF and multi-sensor fusion	Multi-Target Tracking (MTT)	Real-time method	Level [2,3]
Horridge et al. [65]	Simulations	JPDAF based tracker	Multi-Target Tracking (MTT)	400 tracks in real-time	Level [2,3,4,5]
Rezatofighi et al. [66]	Fluorescence microscopy sequences and surveillance camera data	JPDAF based tracker	Multi-Target Tracking (MTT)		Level [2,3]
Zhang et al. [67]	Simulations	Gaussian Mixture Measurement PHD tracker (GMM-PHD)	Multi-Target Tracking (MTT)	Handle bearing measurements	Level [2,3]
Khazaei et al. [68]	Data from a distributed network of cameras	Probabilistic Hypothesis Density (PHD) filter based tracker	Multi-Target Tracking (MTT)		Level [2,3]
Feng et al. [69]	Simulations with sequences from CAVIAR dataset	Variational Bayesian PHD filter with deep learning updates	Multi-Target Tracking (MTT)		Level [2,3]
Correa et al. [14]	Tested on a real-time crowded environment	PHD filter	Multi-Target Tracking (MTT)		Level [4, 5]

TABLE II: Summary of pedestrian tracking models

Study/Paper	Input/Evaluation	Method	Tracking Models	Additional Info	SAE Level
Yoon et al. [70]	ETH dataset	Sequential Monte Carlo PHD filter (SMC-PHD)	Multi-Target Tracking (MTT)	Can handle missing detections	Level [2,3]
Oh et al. [71] [72]	Simulations	Markov Chain Monte Carlo Data Association (MCMCDA) Metropolis-Hastings method	Multi-Target Tracking (MTT)		Level [2,3]
Yu et al. [73]	Simulations and video data	Data-driven Markov Chain Monte Carlo data association (DD-MCMCDA)	MTMulti-Target Tracking (MTT)T		Level [2,3]
Chen et al. [74]	Video data	Dynamical graph matching	Multi-Target Tracking (MTT)	Tracker can deal with interactions	Level [2,3]
Pirsiavash et al. [75]	Video data	Greedy algorithm based on Dynamic Programming	Multi-Target Tracking (MTT)		Level [2,3]
Zhang et al. [76]	CAVIAR and ETHMS datasets	Min-Cost Flow algorithm with an explicit occlusion model (EOM)	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]
Chari et al. [77]	PETS and TUD datasets	Min-Cost Max-Flow network optimization with pair-wise costs	Multi-Target Tracking (MTT)	Occlusion handling	Level [2,3]
Taycher et al. [78]	Video data	Conditional Random Field (CRF) state estimation and grid filter	Multi-Target Tracking (MTT)	Real-time capability	Level [2,3]
Milan et al. [79]	PETS 2010 Benchmark and TUD-Stadtmitte dataset	CRF-based multiple tracker with HOG-SVM detector	Multi-Target Tracking (MTT)		Level [2,3]
Milan et al. [80]	PETS, TUD, ETHMS datasets	CRF-based multi-target tracker using discrete continuous minimization	Multi-Target Tracking (MTT)	Trajectory estimation of targets	Level [2,3,4]
Brendel et al. [81]	ETHZ Central, TUD Crossing, i-Lids AB, UBC Hockey and ETHZ Soccer datasets	Maximum-weight independent set (MWIS) based tracker	Multi-Target Tracking (MTT)	Long-term occlusion handling	Level [2,3]
He et al. [82]	PETS09, TUD Statmitte, TUD Crossing and ETHMS datasets	Connected component model with MWIS	Multi-Target Tracking (MTT)		Level [2,3]
Gaidon et al. [83]	KITTI Benchmark and PASCAL-to-KITTI dataset	Online Domain Adaptation for Multi-Object Tracking	Multi-Target Tracking (MTT)	Generic detector and video adaptation fro tracking	Level [2,3]
Ondruska et al. [84]	Simulated data	End-to-end recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)	No data association required	Level [2,3]
Dequaire et al. [85]	Real-world environment	End-to-end recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)		Level [2,3]
Milan et al. [86]	MOTChallenge 2015 benchmark	Online recurrent neural network (RNN) tracker	Multi-Target Tracking (MTT)		Level [2,3]
Ristani et al. [87]	Multi-cameras system data	Neural networks	Multi-Target Tracking (MTT)	Features learnt multi-cameras and Re-identification	Level [3, 4]
Liu et al. [88]	Multi-camera systems data	Generalized Maximum Multi-Clique optimization	Multi-Target Tracking (MTT)		Level [2,3]
Park et al. [89]	Monocular camera data	3D object tracking	Multi-Target Tracking (MTT)	3D object Tracking for augmented reality applications	Level [2,3]
Scheidegger et al. [90]	Single camera data	Multi-Bernoulli mixture tracking filter	Multi-Target Tracking (MTT)		Level [2,3,4]
Yan et al. [**?**]	Lidar data	3D LIDAR based tracking with Support Vector Machine (SVM) classifier	Multi-Target Tracking (MTT)	Online classification of humans	Level [2,3]
Leal-taixe et al. [91]	Camera data	Interaction feature strings with Random Forest method	Multi-Target Tracking (MTT)	Scene understanding	Level [2,3]
Choi et al. [92]	Video data	Discriminative model	Multi-Target Tracking (MTT)TT	Group activity recognition	Level [2,3]

REFERENCES

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," in *arXiv preprint arXiv:1812.08008*, 2018.

[2] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," *CoRR*, vol. abs/1611.08050, 2016.

[3] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1297–1304, Ieee, 2011.

[4] U. Iqbal, A. Milan, and J. Gall, "Posetrack: Joint multi-person pose estimation and tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2011–2020, 2017.

[5] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Advances in neural information processing systems*, pp. 1799–1807, 2014.

[6] W. Ma, D. Huang, N. Lee, and K. M. Kitani, "Forecasting interactive dynamics of pedestrians with fictitious play," *CoRR*, vol. abs/1604.01431, 2016.

[7] Y. Kohari, J. Miura, and S. Oishi, "Cnn-based human body orientation estimation for robotic attendant," in *IAS-15 Workshop on Robot Perception of Humans*, 2018.

[8] T. Darrell, A. P. Pentland, A. Azarbayejani, and C. R. Wren, "Pfinder : Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 19, pp. 780–785, 07 1997.

[9] A. T. Schulz and R. Stiefelhagen, "Pedestrian intention recognition using latent-dynamic conditional random fields," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 622–627, June 2015.

[10] Microsoft, "Azure face api."

[11] B. Benfold and I. D. Reid, "Guiding visual surveillance by tracking human attention.," in *BMVC*, vol. 2, p. 7, 2009.

[12] T. Baltruaitis, P. Robinson, and L. Morency, "Openface: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, March 2016.

[13] J. Y. R. Cornejo, H. Pedrini, and F. Flórez-Revuelta, "Facial expression recognition with occlusions based on geometric representation," in *Iberoamerican Congress on Pattern Recognition*, pp. 263–270, Springer, 2015.

[14] J. Correa, J. Liu, and G.-Z. Yang, "Real time people tracking in crowded environments with range measurements," in *International Conference on Social Robotics*, pp. 471–480, Springer, 2013.

[15] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," in *Cognitive behavioural systems*, pp. 144–157, Springer, 2012.

[16] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[17] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional mkl based multimodal emotion recognition and sentiment analysis," in *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pp. 439–448, IEEE, 2016.

[18] W.-B. Horng, C.-Y. Chen, Y. Chang, and C.-H. Fan, "Driver fatigue detection based on eye tracking and dynamk, template matching," in *Networking, Sensing and Control, 2004 IEEE International Conference on*, vol. 1, pp. 7–12, IEEE, 2004.

[19] M. Den Uyl and H. Van Kuilenburg, "The facereader: Online facial expression recognition," in *Proceedings of measuring behavior*, vol. 30, pp. 589–590, Citeseer, 2005.

[20] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3908–3916, 2015.

[21] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Computer Vision, IEEE International Conference on*, 2015.

[22] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3774–3782, Oct 2017.

[23] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, June 2014.

[24] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden markov models," *Image and Vision Computing*, vol. 21, no. 8, pp. 745 – 758, 2003.

[25] A. F. Bobick and A. D. Wilson, "Parametric hidden markov models for gesture recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 21, pp. 884–900, 09 1999.

[26] H.-K. Lee and J. H. Kim, "An hmm-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 961–973, Oct 1999.

[27] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in *International workshop on automatic face and gesture recognition*, vol. 12, pp. 296–301, 1995.

[28] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE transactions on multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013.

[29] R. Quintero, I. Parra, J. Lorenzo, D. Fernndez-Llorca, and M. A. Sotelo, "Pedestrian intention recognition by means of a hidden markov model and body language," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–7, Oct 2017.

[30] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1505–1518, Dec 2003.

[31] N. Bellotto, M. Fernandez-Carmona, and S. Cosar, "Enrichme integration of ambient intelligence and robotics for aal," in *AAAI Spring Symposium – Wellbeing AI: From Machine Learning to Subjectivity Oriented Computing*, (Stanford, CA), 27-29 Mar 2017.

[32] A. A. Chaaraoui, P. Climent-Prez, and F. Flrez-Revuelta, "Silhouette-based human action recognition using sequences of key poses," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1799 – 1807, 2013. Smart Approaches for Human Action Recognition.

[33] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 65–72, Oct 2005.

[34] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '07, (New York, NY, USA), pp. 235:1–235:8, ACM, 2007.

[35] A.-A. Liu, W.-Z. Nie, Y.-T. Su, L. Ma, T. Hao, and Z.-X. Yang, "Coupled hidden conditional random fields for rgb-d human action recognition," *Signal Processing*, vol. 112, pp. 74 – 82, 2015. Signal Processing and Learning Methods for 3D Semantic Analysis.

[36] C. Coppola, D. R. Faria, U. Nunes, and N. Bellotto, "Social activity recognition based on probabilistic merging of skeleton features with proximity priors from rgb-d data," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5055–5061, Oct 2016.

[37] I. del Pino, V. Vaquero, B. Masini, J. Solà, F. Moreno-Noguer, A. Sanfeliu, and J. Andrade-Cetto, "Low resolution lidar-based multi-object tracking for driving applications," in *Iberian Robotics conference*, pp. 287–298, Springer, 2017.

[38] N. Bellotto and H. Hu, "Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of Bayesian filters," *Autonomous Robots*, vol. 28, no. 4, pp. 425–438, 2010.

[39] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Trans. Sig. Proc.*, vol. 50, pp. 174–188, Feb. 2002.

[40] X. Fen and G. Ming, "Pedestrian tracking using particle filter algorithm," in *2010 International Conference on Electrical and Control Engineering*, pp. 1478–1481, June 2010.

[41] F. Jurie and M. Dhome, "Real time robust template matching.," in *BMVC*, pp. 1–10, 2002.

[42] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201)*, pp. 8–14, Oct 1998.

[43] T. Kaneko and O. Hori, "Feature selection for reliable tracking using template matching," in *null*, p. 796, IEEE, 2003.

[44] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 2, pp. 142–149 vol.2, June 2000.

[45] R. T. Collins, "Mean-shift blob tracking through scale space," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, pp. II–234, June 2003.

[46] H. Tao, H. S. Sawhney, and R. Kumar, "Object tracking with bayesian estimation of dynamic layer representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 75–89, Jan 2002.

[47] H. Yalcin, M. J. Black, and R. Fablet, "The dense estimation of motion and appearance in layers," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pp. 165–165, IEEE, 2004.
[48] D. Geiger, A. Gupta, L. A. Costa, and J. Vlontzos, "Dynamic programming for detecting, tracking, and matching deformable contours," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 3, pp. 294–302, 1995.
[49] A. Techmer, "Contour-based motion estimation and object tracking for real-time applications," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 3, pp. 648–651 vol.3, Oct 2001.
[50] A. M. Baumberg and D. C. Hogg, "An efficient method for contour tracking using active shape models," in *Proceedings of 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pp. 194–199, Nov 1994.
[51] A. Yilmaz, X. Li, and M. Shah, "A bayesian approach to object contour tracking using level sets," 2003.
[52] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 798–805, June 2006.
[53] F. Meyer and P. Bouthemy, "Region-based tracking in an image sequence," in *European Conference on Computer Vision*, pp. 476–484, Springer, 1992.
[54] R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 366–371, May 2002.
[55] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow," *Image and Vision Computing*, vol. 30, no. 3, pp. 217–226, 2012.
[56] C. Sinthanayothin, N. Wongwaen, and W. Bholsithi, "Skeleton tracking using kinect sensor & displaying in 3 d virtual scene," *International Journal of Advancements in Computing Technology*, vol. 4, no. 11, 2012.
[57] P. Konstantinova, A. Udvarev, and T. Semerdjiev, "A study of a target tracking algorithm using global nearest neighbor approach," in *Proceedings of the International Conference on Computer Systems and Technologies (CompSysTech03)*, pp. 290–295, 2003.
[58] M. Azari, A. Seyfi, and A. H. Rezaie, "Real time multiple object tracking and occlusion reasoning using adaptive kalman filters," in *Machine Vision and Image Processing (MVIP), 2011 7th Iranian*, pp. 1–5, IEEE, 2011.
[59] D. Reid, "An algorithm for tracking multiple targets," *IEEE transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.
[60] M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras, "People tracking with human motion predictions from social forces," in *Proc. of the Int. Conf. on Robotics & Automation (ICRA)*, (Anchorage, USA), 2010.
[61] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4696–4704, 2015.
[62] B. Zhou and N. Bose, "Multitarget tracking in clutter: Fast algorithms for data association," *IEEE Transactions on aerospace and electronic systems*, vol. 29, no. 2, pp. 352–363, 1993.
[63] Y. Chen, Y. Rui, and T. S. Huang, "Jpdaf based hmm or real-time contour tracking," in *null*, p. 543, IEEE, 2001.
[64] N. Liu, R. Xiong, Q. Li, and Y. Wang, "Human tracking using improved sample-based joint probabilistic data association filter," in *Intelligent Autonomous Systems 12* (S. Lee, H. Cho, K.-J. Yoon, and J. Lee, eds.), (Berlin, Heidelberg), pp. 293–302, Springer Berlin Heidelberg, 2013.
[65] P. Horridge and S. Maskell, "Real-time tracking of hundreds of targets with efficient exact jpdaf implementation," in *Information Fusion, 2006 9th International Conference on*, pp. 1–8, IEEE, 2006.
[66] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, "Joint probabilistic data association revisited," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3047–3055, Dec 2015.
[67] Q. Zhang and T. L. Song, "Improved bearings-only multi-target tracking with gm-phd filtering," *Sensors*, vol. 16, no. 9, p. 1469, 2016.
[68] M. Khazaei and M. Jamzad, "Multiple human tracking using phd filter in distributed camera network," in *2014 4th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 569–574, Oct 2014.
[69] P. Feng, W. Wang, S. M. Naqvi, and J. Chambers, "Variational bayesian phd filter with deep learning network updating for multiple human tracking," in *2015 Sensor Signal Processing for Defence (SSPD)*, pp. 1–5, Sept 2015.
[70] J. H. Yoon, K. J. Yoon, and D. Y. Kim, "Multi-object tracking using hybrid observation in phd filter," in *2013 IEEE International Conference on Image Processing*, pp. 3890–3894, Sept 2013.
[71] S. Oh, S. Russell, and S. Sastry, "Markov chain monte carlo data association for multi-target tracking," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 481–497, 2009.
[72] S. Oh, S. Russell, and S. Sastry, "Markov chain monte carlo data association for general multiple-target tracking problems," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 1, pp. 735–742 Vol.1, Dec 2004.
[73] Q. Yu, G. Medioni, and I. Cohen, "Multiple target tracking using spatio-temporal markov chain monte carlo data association," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
[74] H.-T. Chen, H.-H. Lin, and T.-L. Liu, "Multi-object tracking using dynamical graph matching," in *null*, p. 210, IEEE, 2001.
[75] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *CVPR 2011*, pp. 1201–1208, June 2011.
[76] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.
[77] V. Chari, S. Lacoste-Julien, I. Laptev, and J. Sivic, "On pairwise costs for network flow multi-object tracking," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
[78] L. Taycher, D. Demirdjian, T. Darrell, and G. Shakhnarovich, "Conditional random people: Tracking humans with crfs and grid filters," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 222–229, June 2006.
[79] A. Milan, L. Leal-Taixe, K. Schindler, and I. Reid, "Joint tracking and segmentation of multiple targets," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
[80] A. Milan, K. Schindler, and S. Roth, "Multi-target tracking by discrete-continuous energy minimization," *IEEE TPAMI*, vol. 38, pp. 2054–2068, Oct 2016.
[81] W. Brendel, M. Amer, and S. Todorovic, "Multiobject tracking as maximum weight independent set," in *CVPR 2011*, pp. 1273–1280, June 2011.
[82] Z. He, X. Li, X. You, D. Tao, and Y. Y. Tang, "Connected component model for multi-object tracking," *IEEE Transactions on Image Processing*, vol. 25, pp. 3698–3711, Aug 2016.
[83] A. Gaidon and E. Vig, "Online domain adaptation for multi-object tracking," *CoRR*, vol. abs/1508.00776, 2015.
[84] P. Ondruska and I. Posner, "Deep tracking: Seeing beyond seeing using recurrent neural networks," *CoRR*, vol. abs/1602.00991, 2016.
[85] J. Dequaire, P. Ondrka, D. Rao, D. Wang, and I. Posner, "Deep tracking in the wild: End-to-end tracking using recurrent neural networks," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 492–512, 2018.
[86] A. Milan, S. H. Rezatofighi, A. R. Dick, K. Schindler, and I. D. Reid, "Online multi-target tracking using recurrent neural networks," *CoRR*, vol. abs/1604.03635, 2016.
[87] E. Ristani and C. Tomasi, "Features for multi-target multi-camera tracking and re-identification," *CoRR*, vol. abs/1803.10859, 2018.
[88] W. Liu, O. I. Camps, and M. Sznaier, "Multi-camera multi-object tracking," *CoRR*, vol. abs/1709.07065, 2017.
[89] Y. Park, V. Lepetit, and W. Woo, "Multiple 3d object tracking for augmented reality," in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 117–120, Sept 2008.
[90] S. Scheidegger, J. Benjaminsson, E. Rosenberg, A. Krishnan, and K. Granström, "Mono-camera 3d multi-object tracking using deep learning detections and PMBM filtering," *CoRR*, vol. abs/1802.09975, 2018.
[91] L. Leal-Taixé, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese, "Learning an image-based motion context for multiple people tracking," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
[92] W. Choi and S. Savarese, "A unified framework for multi-target tracking and collective activity recognition," in *European Conference on Computer Vision*, pp. 215–230, Springer, 2012.

Fanta Camara is a PhD student at the Institute for Transport Studies (ITS), University of Leeds, UK, a Visiting Researcher in the School of

Computer Science, University of Lincoln, UK, a member of Lincoln Centre for Autonomous Systems (L-CAS) and a member of Robotics at Leeds. She received her Master degree in Robotics Engineering from Sorbonne Université (former Université Pierre et Marie Curie, UPMC), Paris, France. She was an exchange student at National Taiwan University (NTU), Taipei, Taiwan. She had some work experience in Germany at Airbus Defence and Space, DLR (German Aerospace Center) and Robert Bosch GmbH. Her main research interests are in human-robot interaction for autonomous systems, computer vision, game theory and robotics.

Nicola Bellotto (SMIEEE) is a Reader in the School of Computer Science, University of Lincoln, UK, and a member of the Lincoln Centre for Autonomous Systems. His main research interests are in machine perception, especially for human detection, tracking, identification and activity recognition with autonomous mobile robots. He has a Master in Electronic Engineering from the University of Padua, Italy, and a PhD in Computer Science from the University of Essex, UK. Before joining the University of Lincoln, he was a researcher in the Active Vision Lab at the University of Oxford. Dr Bellotto is the recipient of a Google Faculty Research Award and a PI/Co-I in several EU and UK projects on autonomous mobile robots.

Serhan Cosar is a post-doctoral research fellow at Lincoln Centre for Autonomous Systems Research (L-CAS), University of Lincoln, United Kingdom. He received his M.Sc. and Ph.D. degrees of Electronics Engineering and Computer Science from Sabanc University, Istanbul, Turkey in 2008 and 2013, respectively. He is author of 3 book chapter, 7 international journal papers (CVIU, IMAVIS) and 15 international conference papers (ACPR, ICCV, AVSS) and he has participated in a number of European and National projects. His research interest includes robot perception, human tracking, activity recognition, sparse representation, and distributed estimation.

Dimitris Nathanael is an assistant professor at the National Technical University of Athens, School of Mechanical Engineering, where he teaches ergonomics / human factors, design innovation and safety management courses. His scientific interests span across all facets of the interaction of humans with technology with an emphasis on cognition. Dimitris holds a diploma in Production Engineering & Management science, Technical University of Crete, a D.E.A. in Ergonomics, from the Conservatoire National des Arts et Mtiers (C.N.A.M.) in France, and a PhD in the field of Cognitive Ergonomics, from the National Technical University of Athens. Before joining the academia, Dimitris has worked in for several years in industry as a consultant in ergonomics, human computer interaction and occupational safety. He is the President of the Hellenic Ergonomics Society and the national representative in the International Ergonomics Association.

Matthias Althoff is an assistant professor in computer science at Technische Universität München, Germany. He received his diploma engineering degree in Mechanical Engineering in 2005, and his Ph.D. degree in Electrical Engineering in 2010, both from Technische Universität München, Germany. From 2010 to 2012 he was a postdoctoral researcher at Carnegie Mellon University, Pittsburgh, USA, and from 2012 to 2013 an assistant professor at Technische Universität Ilmenau, Germany. His research interests include formal verification of continuous and hybrid systems, reachability analysis, planning algorithms, nonlinear control, robotics, automated vehicles, and power systems.

Jingyuan Wu is pursuing his PhD degree with Chassis Systems Control, Robert Bosch GmbH in Abstatt, Germany, in cooperation with the Technical University of Munich, Germany. He received his B.Sc. degree in mechanical engineering from University of Science and Technology Beijing, China in 2013 and M.Sc. degree in mechanical engineering from Technical University of Kaiserslautern, Germany in 2017. His work is focused on signal processing and in particular pedestrian motion prediction in urban environments.

Johannes Ruenz is currently working in the advanced development department of Robert Bosch GmbHs Chassis Systems division. He studied Electrical Engineering and Information Technologies at the Karlsruher Institute of Technology (KIT) and received his master degree in 2014. His research topic is focused on active pedestrian protection systems using their smartphones and vehicles environment sensors.

André Dietrich is a research associate and PhD student at the Chair of Ergonomics at the Technical University of Munich. His research focuses on interaction between automated vehicles and pedestrians with an emphasis on the effects of visual communication. He currently leads Work package 2 ”Psychological Models on Human Interaction & Intention Recognition Algorithms” in the European Project interACT. He graduated in Aerospace Engineering at the Technical University of Munich

Charles W. Fox is a Senior Lecturer in Robotics and Automation at the University of Lincoln and a Visiting Lecturer in Vehicle and Road Automation at the Institute for Transport Studies at the University of Leeds. He received his MA in Computer Science from the University of Cambridge, MSc in Cognitive Science from the University of Edinburgh, and DPhil in Engineering from the University of Oxford. He is the author of the Springer textbook ”Data Science for Transport: a self-study guide with computer exercises”.