

## Human Movement Datasets: An Interdisciplinary Scoping Review (Supplementary Material 2)

TEMITAYO OLUGBADE, University College London, United Kingdom

MARTA BIEŃKIEWICZ, EuroMov Digital Health in Motion, Univ. Montpellier IMT Mines Ales, France

GIULIA BARBARESCHI, University College London, United Kingdom

VINCENZO D'AMATO, Università di Genova, Italy

LUCA ONETO, Università di Genova, Italy

ANTONIO CAMURRI, Università di Genova, Italy

CATHERINE HOLLOWAY, University College London, United Kingdom

MÅRTEN BJÖRKMAN, KTH Royal Institute of Technology, Sweden

PETER KELLER, Western Sydney University, Australia

MARTIN CLAYTON, Durham University, United Kingdom

AMANDA C DE C WILLIAMS, University College London, United Kingdom

NICOLAS GOLD, University College London, United Kingdom

CRISTINA BECCHIO, Department of Neurology, University Medical Center Hamburg-Eppendorf, Germany and Italian Institute of Technology, Italy

BENOÎT BARDY, EuroMov Digital Health in Motion, Univ. Montpellier IMT Mines Ales, France

NADIA BIANCHI-BERTHOUBE, University College London, United Kingdom

### 1 AN ABRIDGEMENT OF THE MAIN CATALOGUE OF 704 OPEN HUMAN MOVEMENT DATASETS

This is the second part of the supplementary materials for the ACM Computing Surveys publication:

*Human Movement Datasets: An Interdisciplinary Scoping Review*

Temitayo Olugbade, Marta Bieńkiewicz, Giulia Barbareschi, Vincenzo D'Amato, Luca Oneto, Antonio Camurri, Catherine Holloway, Mårten Björkman, Peter Keller, Martin Clayton, Amanda Williams, Nicolas Gold, Cristina Becchio, Benoît Bardy, Nadia Bianchi-Berthouze.

and any use of the supplementary materials should cite this publication.

In this document, 704 open human datasets are listed by their references, with a table (see Table 1) that shows the distribution of the datasets across year periods. See the main catalogue(in the supplementary materials) for the detailed

---

Authors' addresses: Temitayo Olugbade, [temitayo.olugbade.13@ucl.ac.uk](mailto:temitayo.olugbade.13@ucl.ac.uk), University College London, Gower Street, London, United Kingdom, WC1E 6BT; Marta Bieńkiewicz, EuroMov Digital Health in Motion, Univ. Montpellier IMT Mines Ales, Montpellier, France; Giulia Barbareschi, University College London, London, United Kingdom; Vincenzo D'Amato, Università di Genova, Genoa, Italy; Luca Oneto, Università di Genova, Genoa, Italy; Antonio Camurri, Università di Genova, Genoa, Italy; Catherine Holloway, University College London, London, United Kingdom; Mårten Björkman, KTH Royal Institute of Technology, Stockholm, Sweden; Peter Keller, Western Sydney University, Sydney, Australia; Martin Clayton, Durham University, Durham, United Kingdom; Amanda C de C Williams, University College London, London, United Kingdom; Nicolas Gold, University College London, London, United Kingdom; Cristina Becchio, Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany and Italian Institute of Technology, Genoa, Italy; Benoît Bardy, EuroMov Digital Health in Motion, Univ. Montpellier IMT Mines Ales, Montpellier, France; Nadia Bianchi-Berthouze, University College London, London, United Kingdom, [nadia.berthouze@ucl.ac.uk](mailto:nadia.berthouze@ucl.ac.uk).

record of each dataset, and please see the publication cited above for the background to the table.

Table 1. An abridgement of the main catalogue of 704 open human movement datasets

Year of publication	Datasets (References only)
Up till 2000	[346, 502]
2001 - 2005	[92, 112, 185, 213, 255, 257, 283, 295, 321, 387, 394, 444, 451, 495, 537, 544, 561, 564, 620, 643]
2006 - 2010	[5, 6, 12, 15, 23, 30, 33, 37, 39, 44, 45, 67, 73, 75, 81, 88, 91, 102, 119, 144, 154, 158, 160, 172, 175, 176, 182, 186, 205, 209, 216, 217, 251-254, 285, 319, 320, 330, 332, 340, 350, 357, 359, 364, 367, 370, 377, 384, 390, 395, 398, 402, 403, 411, 414, 425, 428, 429, 475, 481, 500, 505, 519, 523, 565, 570, 578, 580, 582, 596, 601, 615, 623, 647, 650, 651, 656, 667, 675, 678, 683, 684, 701]
2011 - 2015	[9, 13, 19, 20, 22, 25, 28, 29, 36, 41-43, 46, 48, 51-54, 57, 59, 63-65, 69, 70, 72, 74, 76, 80, 93-95, 100, 103-105, 108, 109, 111, 113, 114, 116, 117, 122-124, 128, 129, 138, 139, 141, 146, 148-151, 169, 173, 174, 177, 179, 187, 189, 194, 195, 204, 206, 210-212, 215, 219, 221, 226, 227, 229, 232, 237, 239, 240, 242-244, 246, 248, 249, 256, 258, 263-265, 267, 268, 272-274, 277-279, 288, 290, 294, 296, 302-305, 308-317, 323, 326, 327, 329, 331, 344, 345, 347, 348, 358, 360-363, 366, 368, 369, 373, 375, 376, 379, 380, 391, 396, 397, 401, 406, 412, 415-418, 420, 423, 427, 431-433, 436, 439-441, 443, 446, 452, 453, 470-472, 474, 477-479, 482-484, 487, 490, 491, 498, 499, 506, 507, 512, 513, 515-518, 522, 524, 526-528, 530, 531, 534, 536, 539, 545, 548, 550, 552, 554, 556, 560, 562, 563, 567, 573-575, 577, 581, 586, 588-591, 594, 599, 602, 606, 609, 613, 618, 619, 627, 631-633, 636, 637, 639, 640, 645, 649, 655, 657-659, 661, 664, 666, 673, 676, 677, 679, 680, 688, 690-692, 695-698]
2016 - 2020	[1, 7, 7, 8, 10, 11, 14, 16-18, 21, 26, 27, 31, 32, 34, 35, 38, 40, 47, 49, 50, 55, 56, 58, 60-62, 66, 68, 71, 77, 79, 83-87, 89, 90, 96-98, 101, 106, 107, 110, 115, 118, 120, 121, 125, 126, 130-137, 142, 143, 145, 147, 152, 153, 155-157, 159, 161-168, 170, 178, 180, 181, 183, 184, 188, 190-193, 196-203, 207, 208, 214, 218, 220, 222-225, 228, 230, 231, 233-235, 238, 241, 245, 247, 250, 259, 261, 266, 269-271, 275, 276, 280-282, 284, 286, 287, 289, 291-293, 297-301, 306, 307, 318, 322, 324, 325, 328, 333-339, 341-343, 349, 351-356, 365, 371, 372, 374, 378, 381-383, 385, 386, 388, 389, 392, 393, 399, 400, 404, 405, 408-410, 413, 419, 421, 424, 426, 430, 434, 435, 437, 438, 442, 445, 447-450, 454-469, 473, 476, 480, 485, 486, 488, 489, 492-494, 496, 497, 501, 503, 504, 508-511, 514, 520, 521, 525, 529, 532, 533, 535, 538, 540-543, 546, 547, 549, 551, 553, 555, 557-559, 566, 568, 569, 571, 572, 576, 579, 583-585, 587, 592, 593, 595, 597, 598, 600, 603-605, 607, 608, 610, 616, 617, 621, 622, 624, 625, 628-630, 634, 635, 638, 641, 642, 644, 646, 648, 652-654, 660, 662, 663, 665, 668-672, 674, 681, 682, 685-687, 689, 693, 694, 699, 700]
Early 2021	[4]
Unknown	[2, 3, 24, 78, 82, 99, 127, 140, 171, 236, 260, 262, 407, 422, 611, 612, 614, 626]

## REFERENCES

- [1] Yehya Abouelnaga, Hesham M. Eraqi, and Mohamed N. Moustafa. 2017. Real-time Distracted Driver Posture Classification. *arXiv* (2017).
- [2] ACCADOSU [n.d.]. ACCAD: Motion Capture Data. <https://accad.osu.edu/research/motion-lab/mocap-system-and-data>.
- [3] ActivityNet1.2 [n.d.]. ActivityNet. <http://activity-net.org/download.html>.

- [4] Nikolaos M. Adaloglou, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George Xydopoulos, Klimis Antzakas, Dimitris Papazachariou, and Petros none Daras. 2021. A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition. *IEEE Transactions on Multimedia* 22(10), c (2021), 1–14. <https://doi.org/10.1109/TMM.2021.3070438>
- [5] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and Daviv Reinitz. 2008. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 3 (2008), 555–560. <https://doi.org/10.1109/TPAMI.2007.70825>
- [6] Ulrich Von Agris and Karl-friedrich Kraiss. [n.d.]. Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition. In *Gesture in Human-Computer Interaction and Simulation*.
- [7] Insaf Ajili, Malik Malle, and Jean-Yves Didier. 2019. Human motions and emotions recognition inspired by LMA qualities. *The Visual Computer* 35, 10 (2019), 1411–1426. <https://doi.org/10.1007/s00371-018-01619-w>
- [8] Seymanur Akti, Gozde Ayse Tataroglu, and Hazim Kemal Ekenel. 2019. Vision-based Fight Detection from Surveillance Cameras. In *9th International Conference on Image Processing Theory, Tools and Applications, IPTA*. <https://doi.org/10.1109/IPTA.2019.8936070>
- [9] Xavier Alameda-Pineda, Jordi Sanchez-Riera, Johannes Wienke, Vojtěch Franc, Jan Čech, Kaustubh Kulkarni, Antoine Deleforge, and Radu Horaud. 2013. RAVEL: an annotated corpus for training robots with audiovisual abilities. *Journal on Multimodal User Interfaces* 7, 1 (2013), 79–91. <https://doi.org/10.1007/s12193-012-0111-y>
- [10] Xavier Alameda-Pineda, Jacopo Staiano, Ramanathan Subramanian, Ligia Batrinca, Elisa Ricci, Bruno Lepri, Oswald Lanz, and Nicu Sebe. 2016. SALSA: A novel dataset for multimodal group behavior analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 8 (2016), 1707–1720. <https://doi.org/10.1109/TPAMI.2015.2496269>
- [11] Jean Baptiste Alayrac, Piotr Bojanowski, Nishant Agrawal, Josef Sivic, Ivan Laptev, and Simon Lacoste-Julien. 2016. Unsupervised learning from narrated instruction videos. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2016.495>
- [12] Marcel Alcoverro, Josep Ramon Casas, and Montse Pardas. 2010. Skeleton and shape adjustment and tracking in multicamera environments. In *International Conference on Articulated Motion and Deformable Objects*. Springer, 88–97.
- [13] M Alfonse, A Ali, A S Elons, N L Badr, and M Aboul-Ela. 2015. Arabic sign language benchmark database for different heterogeneous sensors. *2015 5th International Conference on Information Communication Technology and Accessibility (ICTA)* (2015). <https://doi.org/10.1109/ICTA.2015.7426902>
- [14] Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2018. Kinder-Gator: The UF Kinect Database of Child and Adult Motion. *Eurographics* (2018), 13–16. <https://doi.org/10.2312/egs.20181033>
- [15] Kerem Altun and Billur Barshan. 2010. Human activity recognition using inertial/magnetic sensor units. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 6219 LNCS. 38–51. [https://doi.org/10.1007/978-3-642-14715-9\\_5](https://doi.org/10.1007/978-3-642-14715-9_5)
- [16] Claudia Álvarez-Aparicio, Ángel Manuel Guerrero-Higueras, Maria Carmen Calvo Olivera, Francisco J Rodríguez-Lera, Francisco Martín, and Vicente Matellán. 2018. Benchmark dataset for evaluation of range-based people tracker classifiers in mobile robots. *Frontiers in neurorobotics* 11 (2018), 72.
- [17] Mona Saleh Alzahrani, Salma Kammoun Jarraya, Manar Ali Salamah, and Hanène Ben-Abdallah. 2017. FallFree: Multiple fall scenario dataset of cane users for monitoring applications using kinect. In *13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*. IEEE, 327–333.
- [18] E Ambrosini, M Parati, E Peri, C De Marchis, C Nava, A Pedrocchi, G Ferriero, and S Ferrante. 2020. Changes in leg cycling muscle synergies after training augmented by functional electrical stimulation in subacute stroke survivors: a pilot study. *J Neuroeng Rehabil* 17, 1 (2020), 35. <https://doi.org/10.1186/s12984-020-00662-w>
- [19] Mohsen S Amiri. 2013. Non-Intrusive Human Activity Monitoring in a Smart Home Environment. In *5th International Conference on e-Health Networking, Applications and Services (Healthcom)*. 606–610.
- [20] Virginia O. Andersson and Ricardo M. Araujo. 2015. Person identification using anthropometric and gait data from kinect sensor. In *National Conference on Artificial Intelligence*, Vol. 1. 425–431.
- [21] Mykhaylo Andriluka, Umar Iqbal, Eldar Insafutdinov, Leonid Pishchulin, Anton Milan, Juergen Gall, and Bernt Schiele. 2017. PoseTrack: A Benchmark for Human Pose Estimation and Tracking. *arXiv* (2017).
- [22] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *Cvpr*. 3686–3693.
- [23] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. 2010. Monocular 3D pose estimation and tracking by detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 623–630. <https://doi.org/10.1109/CVPR.2010.5540156>
- [24] AniAgeProjectdataset [n.d.]. Traditional Dances Download. <https://www.euh2020aniage.org/testthaidancedownload>.
- [25] Muhammad Latif Anjum, Omar Ahmad, Stefano Rosa, Jingchun Yin, and Basilio Bona. 2014. Skeleton Tracking Based Complex Human Activity Recognition Using Kinect Camera. In *Social Robotics*, Michael Beetz, Benjamin Johnston, and Mary-Anne Williams (Eds.). Springer International Publishing, Cham, 23–33.
- [26] AnnotatedBharatnatyam 2017. Annotated Bharatnatyam Data Set. <http://hci.cse.iitkgp.ac.in/>.
- [27] João Antunes, Alexandre Bernardino, Asim Smailagic, and Daniel P Siewiorek. 2018. AHA-3D: A Labelled Dataset for Senior Fitness Exercise Recognition and Segmentation from 3D Skeletal Data. In *BMVC*. 332.
- [28] Tarik Arici, Sait Celebi, Ali S Aydin, and Talha T Temiz. 2014. Robust gesture recognition using feature pre-processing and weighted dynamic time warping. *Multimedia Tools and Applications* 72, 3 (2014), 3045–3062. <https://doi.org/10.1007/s11042-013-1591-9>

- [29] Andreas Aristidou, Efstathios Stavrakis, and Yiorgos Chrysanthou. 2014. LMA-Based Motion Retrieval for Folk Dance Cultural Heritage. In *Digital Heritage*, Marinos Ioannides, Nadia Magnenat-Thalmann, Eleanor Fink, Roko Žarnić, Alex-Yianing Yen, and Ewald Quak (Eds.). Springer International Publishing, Cham, 207–216.
- [30] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan, and Ashwin Thangali. 2008. The American Sign Language Lexicon Video Dataset. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*. IEEE. <https://doi.org/10.1109/CVPRW.2008.4563181>
- [31] M S H Aung, S Kaltwang, B Romera-Paredes, B Martinez, A Singh, M Cella, M Valstar, H Meng, A Kemp, M Shafizadeh, A C Elkins, N Kanakam, A de Rothschild, N Tyler, P J Watson, A C d. C Williams, M Pantic, and N Bianchi-Berthouze. 2016. The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal EmoPain Dataset. *IEEE Transactions on Affective Computing* 7 (2016). <https://doi.org/10.1109/TAFFC.2015.2462830>
- [32] Suzanne Aussems, Natasha Kwok, and Sotaro Kita. 2018. GestuRe and ACTION Exemplar (GRACE) video database: stimuli for research on manners of human locomotion and iconic gestures. *Behavior Research Methods* 50, 3 (2018), 1270–1284. <https://doi.org/10.3758/s13428-017-0942-2>
- [33] Edouard Auvinet, Caroline Rougier, Jean Meunier, Alain St-arnaud, and Jacqueline Rousseau. 2010. *Multiple cameras fall data set*. Technical Report.
- [34] Giuseppe Averta, Cosimo Della Santina, Gaetano Valenza, Antonio Bicchi, and Matteo Bianchi. 2020. Exploiting upper-limb functional principal components for human-like motion generation of anthropomorphic robots. *Journal of NeuroEngineering and Rehabilitation* 17, 1 (2020), 63. <https://doi.org/10.1186/s12984-020-00680-8>
- [35] Konstantinos Avgerinakis and Yiannis Kompatsiaris. 2016. Demcare action dataset for evaluating dementia patients in a home-based environment. *InImpact: The Journal of Innovation Impact* 6, 1 (2016), 83.
- [36] Sandra Avila, Nicolas Thome, Matthieu Cord, Eduardo Valle, and Arnaldo De A. Araújo. 2013. Pooling in image representation: The visual codeword point of view. *Computer Vision and Image Understanding* 117, 5 (2013), 453–465. <https://doi.org/10.1016/j.cviu.2012.09.007>
- [37] Héctor H Avilés, Wendy Aguilar, and Luis A Pineda. 2008. On the Selection of a Classification Technique for the Representation and Recognition of Dynamic Gestures. In *IBERAMIA*, Hector Geffner, Rui Prada, Isabel Machado Alexandre, and Nuno David (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 412–421.
- [38] Pablo Azagra, Yoan Mollard, Florian Golemo, Ana Cristina Murillo, Manuel Lopes, and Javier Civera. 2016. A Multimodal Human-Robot Interaction Dataset. In *NIPS 2016, workshop Future of Interactive Learning Machines*.
- [39] Marc Bächlin, Meir Plotnik, Daniel Roggen, Inbal Maidan, Jeffrey M. Hausdorff, Nir Giladi, and Gerhard Tröster. 2010. Wearable assistant for Parkinsons disease patients with the freezing of gait symptom. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 436–446. <https://doi.org/10.1109/TITB.2009.2036165>
- [40] Greet Baldewijns, Glen Debar, Gert Mertes, Bart Vanrumste, and Tom Croonenborghs. 2016. Bridging the gap between real-life data and simulated data by providing a highly realistic fall dataset for evaluating camera-based fall detection algorithms. *Healthcare technology letters* 3, 1 (2016), 6–11.
- [41] Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 2011. 3DPeS: 3D people dataset for surveillance and forensics. In *ACM Multimedia Conference and Co-Located Workshops*. 59–64. <https://doi.org/10.1145/2072572.2072590>
- [42] Oresti Banos, Rafael Garcia, Juan A. Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio Rojas, Alejandro Saez, and Claudia Villalonga. 2014. mHealthDroid: A novel framework for agile development of mobile health applications. In *Lecture Notes in Computer Science*. 91–98.
- [43] O. Banos, M. A. Toth, M. Damas, H. Pomares, I. Rojas, and O Amft. 2012. A benchmark dataset to evaluate sensor displacement in activity recognition. In *14th International Conference on Ubiquitous Computing*.
- [44] Tanja Bänziger, Marcello Mortillaro, and Klaus R Scherer. 2012. Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 5 (2012), 1161.
- [45] Tanja Bänziger and Klaus R Scherer. 2007. Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In *International conference on affective computing and intelligent interaction*. Springer, 476–487.
- [46] Igor Barros Barbosa, Marco Cristani, Alessio Del Bue, Loris Bazzani, and Vittorio Murino. 2012. Re-identification with RGB-D sensors. In *Lecture Notes in Computer Science*. Vol. 7583 LNCS. 433–442. [https://doi.org/10.1007/978-3-642-33863-2\\_43](https://doi.org/10.1007/978-3-642-33863-2_43)
- [47] Paola Barra, Carmen Bisogni, Michele Nappi, David Freire-Obregón, and Modesto Castrillón-Santana. 2020. Gotcha-I: A Multiview Human Videos Dataset. In *Security in Computing and Communications*, Sabu M Thampi, Gregorio Martinez Perez, Ryan Ko, and Danda B Rawat (Eds.). Springer Singapore, Singapore, 213–224.
- [48] Silvio Barra, Maria De Marsico, Virginio Cantoni, and Daniel Riccio. 2014. Using Mutual Information for Multi-Anchor Tracking of Human Beings. In *Biometric Authentication*, Virginio Cantoni, Dimo Dimov, and Massimo Tistarelli (Eds.). Springer International Publishing, Cham, 28–39.
- [49] Daniel Paul Barrett, Ran Xu, Haonan Yu, and Jeffrey Mark Siskind. 2016. Collecting and annotating the large continuous action dataset. *Machine Vision and Applications* 27, 7 (2016), 983–995. <https://doi.org/10.1007/s00138-016-0768-4>
- [50] Pablo Barros, Nikhil Churamani, Angelica Lim, and Stefan Wermter. 2019. The OMG-Empathy Dataset: Evaluating the Impact of Affective Behavior in Storytelling. In *8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.
- [51] Federico Bartoli, Giuseppe Lisanti, Lorenzo Seidenari, Svebor Karaman, and Alberto Del Bimbo. 2015. Museumvisitors: a dataset for pedestrian and group detection, gaze estimation and behavior understanding. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 19–27.
- [52] Florian Baumann, Irina Schulz, and Bodo Rosenhahn. 2014. Multi-sensor Acceleration-Based Action Recognition. In *Image Analysis and Recognition*, Aurélio Campilho and Mohamed Kamel (Eds.). Springer International Publishing, Cham, 48–57.

- [53] Kamil Behún, Adam Herout, and Alexander Páldy. 2014. Kinect-supported dataset creation for human pose estimation. In *30th Spring Conference on Computer Graphics*. 55–62.
- [54] Vasileios Belagiannis, Sikandar Amin, Mykhaylo Andriluka, Bernt Schiele, Nassir Navab, and Slobodan Ilic. 2014. 3D Pictorial Structures for Multiple Human Pose Estimation. In *International Conference on Computer Vision Pattern Recognition*.
- [55] Vasileios Belagiannis, Xinchao Wang, Horesh Beny Ben Shitrit, Kiyoshi Hashimoto, Ralf Stauder, Yoshimitsu Aoki, Michael Kranzfelder, Armin Schneider, Pascal Fua, Slobodan Ilic, Hubertus Feussner, and Nassir Navab. 2016. Parsing human skeletons in an operating room. *Machine Vision and Applications* 27, 7 (2016), 1035–1046. <https://doi.org/10.1007/s00138-016-0792-4>
- [56] Atef Ben-Youssef, Chloé Clavel, Slim Essid, Miriam Bilac, Marine Chamoux, and Angelica Lim. 2017. UE-HRI: a new dataset for the study of user engagement in spontaneous human-robot interactions. In *19th ACM international conference on multimodal interaction*. 464–472.
- [57] Ben Belfold and Ian Reid. 2011. Stable multi-target tracking in real-time surveillance video. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 3457–3464. <https://doi.org/10.1109/CVPR.2011.5995667>
- [58] T R Bennett, H C Massey, J Wu, S A Hasnain, and R Jafari. 2016. MotionSynthesis Toolset (MoST): An Open Source Tool and Data Set for Human Motion Data Synthesis and Validation. *IEEE Sensors Journal* 16 (2016). <https://doi.org/10.1109/JSEN.2016.2562599>
- [59] Enrique Bermejo Nievas, Oscar Deniz Suarez, Gloria Bueno García, and Rahul Sukthankar. 2011. Violence Detection in Video Using Computer Vision Techniques. In *Computer Analysis of Images and Patterns*, Pedro Real, Daniel Diaz-Pernil, Helena Molina-Abril, Ainhoa Berciano, and Walter Kropatsch (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 332–339.
- [60] Cigdem Beyan, Nicolò Carissimi, Francesca Capozzi, Sebastiano Vascon, Matteo Bustreo, Antonio Pierro, Cristina Becchio, and Vittorio Murino. 2016. Detecting emergent leader in a meeting environment using nonverbal visual features only. In *18th ACM International Conference on Multimodal Interaction*. 317–324. <https://doi.org/10.1145/2993148.2993175>
- [61] C Beyan, M Shahid, and V Murino. 2020. RealVAD: A Real-world Dataset and A Method for Voice Activity Detection by Body Motion Analysis. *IEEE Transactions on Multimedia PP* (2020). <https://doi.org/10.1109/TMM.2020.3007350>
- [62] Sukalyan Bhakat and Ganesh Ramakrishnan. 2019. Anomaly detection in surveillance videos. In *ACM India Joint International Conference on Data Science and Management of Data*. 252–255. <https://doi.org/10.16511/j.cnki.qhdxxb.2020.22.008>
- [63] Alina Bialkowski, Simon Denman, Sridha Sridharan, Clinton Fookes, and Patrick Lucey. 2012. A database for person re-identification in multi-camera surveillance networks. In *International Conference on Digital Image Computing Techniques and Applications, DICTA 2012*. IEEE. <https://doi.org/10.1109/DICTA.2012.6411689>
- [64] Victoria Bloom, Vasileios Argyriou, and Dimitrios Makris. 2014. G3di: A gaming interaction dataset with a real time detection and evaluation framework. In *European Conference on Computer Vision*. Springer, 698–712.
- [65] Victoria Bloom, Dimitrios Makris, and Vasileios Argyriou. 2012. G3D: a gaming action dataset and real time action recognition evaluation framework. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2012), 7–12. <https://doi.org/10.1109/CVPRW.2012.6239175>
- [66] C Bloomer, S Wang, and K Kontson. 2020. Kinematic analysis of motor learning in upper limb body-powered bypass prosthesis training. *PloS one* 15, 1 (2020), e0226563. <https://doi.org/10.1371/journal.pone.0226563>
- [67] Scott Blunsden and R. B. Fisher. 2010. The BEHAVE video dataset: ground truthed video for multi-person behavior classification. *Annals of the BMVA* 4 (2010), 1–12.
- [68] João Borges, Sandro Queirós, Bruno Oliveira, Helena Torres, Nelson Rodrigues, Victor Coelho, Johannes Pallauf, José Henrique Brito, José Mendes, and Jaime C Fonseca. 2020. A system for the generation of in-car human body pose datasets. *Machine Vision and Applications* 32, 1 (2020), 4. <https://doi.org/10.1007/s00138-020-01131-z>
- [69] Ricard Borràs, Àgata Lapedriza, and Laura Igual. 2012. Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition. In *Image Analysis and Recognition*, Aurélio Campilho and Mohamed Kamel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 98–105.
- [70] Alessandro Borreo, Leonardo Onofri, and Paolo Soda. 2015. A multi-environment dataset for activity of daily living recognition in video streams. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 747–750.
- [71] Elif Bozkurt, Hossein Khaki, Sinan Keçeci, B Berker Türker, Yücel Yemez, and Engin Erzin. 2017. The JESTKOD database: an affective multimodal database of dyadic interactions. *Language Resources and Evaluation* 51, 3 (2017), 857–872. <https://doi.org/10.1007/s10579-016-9377-0>
- [72] William Brendel, Alan Fern, and Sinisa Todorovic. 2011. Probabilistic event logic for interval-based event recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Ii* (2011), 3329–3336. <https://doi.org/10.1109/CVPR.2011.5995491>
- [73] Alexia Briassouli, Vagia Tsiminaki, and Ioannis Kompatsiaris. 2009. Human motion analysis via statistical motion processing and sequential change detection. *EURASIP Journal on Image and Video Processing* 2009 (2009), 1–16.
- [74] Geert Brône and Bert Oben. 2015. InSight Interaction: a multimodal and multifocal dialogue corpus. *Language Resources and Evaluation* 49, 1 (2015), 195–214. <https://doi.org/10.1007/s10579-014-9283-2>
- [75] Gabriel J Brostow, Jamie Shotton, Julien Fauqueur, and Roberto Cipolla. 2008. Segmentation and recognition using structure from motion point clouds. In *European conference on computer vision*. Springer, 44–57.
- [76] Thomas Brox and Jitendra Malik. 2011. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 3 (2011), 500–513. <https://doi.org/10.1109/TPAMI.2010.143>
- [77] Lilli Bruckschén, Sabrina Amft, Julian Tanke, Jürgen Gall, and Maren Bennewitz. 2019. Detection of Generic Human-Object Interactions in Video Streams. In *Social Robotics*, Miguel A Salichs, Shuzhi Sam Ge, Emilia Ivanova Barakova, John-John Cabibihan, Alan R Wagner, Álvaro Castro-González, and Hongsheng He (Eds.). Springer International Publishing, Cham, 108–118.

- [78] BSLdataset [n.d.]. British Sign Language Corpus Project. <https://bslcorpusproject.org/data/>.
- [79] Hannah Bull, Annelies Braffort, and Michèle Gouiffès. 2020. MEDI-API-SKEL - A 2D-skeleton video database of french sign language with aligned french subtitles. In *12th International Conference on Language Resources and Evaluation, Conference Proceedings*, 6063–6068.
- [80] Magnus Burenius, Josephine Sullivan, and Stefan Carlsson. 2013. 3D pictorial structures for multiple view articulated pose estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3618–3625. <https://doi.org/10.1109/CVPR.2013.464>
- [81] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation* 42, 4 (2008), 335–359.
- [82] CABankMOVINCorpusdataset [n.d.]. CABank MOVIN Corpus. <https://ca.talkbank.org/access/MOVIN.html>.
- [83] Angelo Cafaro, Johannes Wagner, Tobias Baur, Soumia Dermouche, Mercedes Torres Torres, Catherine Pelachaud, Elisabeth André, and Michel Valstar. 2017. The NoXi database: Multimodal recordings of mediated novice-expert interactions. In *19th ACM International Conference on Multimodal Interaction*. 350–359. <https://doi.org/10.1145/3136755.3136780>
- [84] Enrico Calabrese, Gemma Taverni, Christopher Awai Easthope, Sophie Skriabine, Federico Corradi, Luca Longinotti, Kynan Eng, and Tobi Delbruck. 2019. DHP19: Dynamic vision sensor 3D human pose dataset. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 0.
- [85] Andres Felipe Calvo, German Andres Holguin, and Henry Medeiros. 2019. Human Activity Recognition Using Multi-modal Data Fusion. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Ruben Vera-Rodriguez, Julian Fierrez, and Aythami Morales (Eds.). Springer International Publishing, Cham, 946–953.
- [86] Ryan Cameron, Zheming Zuo, Graham Sexton, and Longzhi Yang. 2018. A Fall Detection/Recognition System and an Empirical Study of Gradient-Based Feature Extraction Approaches. In *Advances in Computational Intelligence Systems*, Fei Chao, Steven Schockaert, and Qingfu Zhang (Eds.). Springer International Publishing, Cham, 276–289.
- [87] Ruben Cantarero, Maria J Santofimia, David Villa, Roberto Requena, Maria Campos, Francisco Florez-Revuelta, Jean-Christophe Nebel, Jesus Martinez-del Rincon, and Juan C Lopez. 2016. Kinect and Episodic Reasoning for Human Action Recognition. In *International Conference on Distributed Computing and Artificial Intelligence*, Sigeru Omatu, Ali Semalat, Grzegorz Bocewicz, Paweł Sitek, Izabela E Nielsen, Julián A García García, and Javier Bajo (Eds.). Springer International Publishing, Cham, 147–154.
- [88] Liangliang Cao, Zicheng Liu, and Thomas S. Huang. 2010. Cross-dataset action detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1998–2005.
- [89] Zhe Cao, Hang Gao, Karttikeya Mangalam, Qi-Zhi Cai, Minh Vo, and Jitendra Malik. 2020. Long-Term Human Motion Prediction with Scene Context. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 387–404.
- [90] Marianna Capecci, Maria Gabriella Ceravolo, Francesco Ferracuti, Sabrina Iarlori, Andrea Monteriù, Luca Romeo, and Federica Verdini. 2019. The KIMORE dataset: Kinematic assessment of Movement and clinical scores for remote monitoring of physical REhabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 7 (2019), 1436–1448.
- [91] A Cappelletti, B Lepri, N Mana, F Pianesi, and M Zancanaro. 2008. A multimodal data collection of daily activities in a real instrumented apartment. In *LREC 2008 Workshop on Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*.
- [92] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. 2005. The AMI meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction*. Springer, 28–39.
- [93] Vincenzo Carletti, Pasquale Foggia, Gennaro Percannella, Alessia Saggese, and Mario Vento. 2013. Recognition of human actions from RGB-D videos using a reject option. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 436–445. [https://doi.org/10.1007/978-3-642-41190-8\\_47](https://doi.org/10.1007/978-3-642-41190-8_47)
- [94] Josep Maria Carmona and Joan Climent. 2012. A Performance Evaluation of HMM and DTW for Gesture Recognition. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Luis Alvarez, Marta Mejail, Luis Gomez, and Julio Jacobo (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 236–243.
- [95] Sandra Carvalho, Jorge Leite, Santiago Galdo-Alvarez, and Oscar F Goncalves. 2012. The emotional movie database (EMDB): A self-report and psychophysiological study. *Applied Psychophysiology and Biofeedback* 37, 4 (2012), 279–294.
- [96] Eduardo Casilari, Jose A Santoyo-Ramón, and Jose M Cano-García. 2017. Umafall: A multisensor dataset for the research on automatic fall detection. *Procedia Computer Science* 110 (2017), 32–39.
- [97] Gregory Castanon, Nathan Shnidman, Tim Anderson, and Jeffrey Byrne. 2019. Out the Window: A Crowd-Sourced Dataset for Activity Classification in Surveillance Video. *arXiv* (2019).
- [98] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2017. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing* (2017).
- [99] CGANormativeGaitdataset [n.d.]. CGA Normative Gait Database. <http://www.clinicalgaitanalysis.com/data/>.
- [100] Pedro Chahuara, Anthony Fleury, François Portet, and Michel Vacher. 2012. Using Markov Logic Network for On-Line Activity Recognition from Non-visual Home Automation Sensors. In *Ambient Intelligence*, Fabio Paternò, Boris de Ruyter, Panos Markopoulos, Carmen Santoro, Evert van Loenen, and Kris Luyten (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 177–192.
- [101] Punarjay Chakravarty and Tinne Tuytelaars. 2016. Cross-modal supervision for learning active speaker detection in video. In *Lecture Notes in Computer Science*. [https://doi.org/10.1007/978-3-319-46454-1\\_18](https://doi.org/10.1007/978-3-319-46454-1_18)

- [102] Antoni B. Chan, Zhang Sheng John Liang, and Nuno Vasconcelos. 2008. Privacy preserving crowd monitoring: Counting people without people models or tracking. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. <https://doi.org/10.1109/CVPR.2008.4587569>
- [103] Imen Charfi, Johel Miteran, Julien Dubois, Mohamed Atri, and Rached Tourki. 2012. Definition and performance evaluation of a robust SVM based fall detection solution. In *8th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2012r*. IEEE, 218–224. <https://doi.org/10.1109/SITIS.2012.155>
- [104] Imen Charfi, Johel Miteran, Julien Dubois, Mohamed Atri, and Rached Tourki. 2013. Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. *Journal of Electronic Imaging* 22, 4 (2013), 041106. <https://doi.org/10.1117/1.jei.22.4.041106>
- [105] T Chattopadhyay, S Sengupta, A Sinha, and N Rampuria. 2011. Creation and Analysis of a Corpus of Text Rich Indian TV Videos. In *International Conference on Document Analysis and Recognition*. <https://doi.org/10.1109/ICDAR.2011.174>
- [106] Anargyros Chatzitofis, Leonidas Saroglou, Prodromos Boutis, Petros Drakoulis, Nikolaos Zioulis, Shishir Subramanyam, Bart Kevelham, Caecilia Charbonnier, Pablo Cesar, and Dimitrios Zarpalas. 2020. HUMAN4D: A Human-Centric Multimodal Dataset for Motions Immersive Media. *IEEE Access* (2020).
- [107] Tatjana Chavdarova, Pierre Baqué, Stéphane Bouquet, Andrii Maksai, Cijo Jose, Louis Lettry, Pascal Fua, Luc Van Gool, and François Fleuret. 2017. The Wildtrack Multi-Camera Person Dataset. *arXiv* (2017).
- [108] C Chen, R Jafari, and N Kehtarnavaz. 2015. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. (2015). <https://doi.org/10.1109/ICIP.2015.7350781>
- [109] Gengjie Chen, Pierre-Luc St-Charles, Wassim Bouachir, and Guillaume-Alexandre Bilodeau Robert Bergevin. 2015. Reproducible Evaluation of Pan-tilt-zoom Tracking. In *International Conference on Image Processing (ICIP)*. IEEE, 1–5.
- [110] Junwen Chen, Haiting Hao, Hanbin Hong, and Yu Kong. 2020. RIT-18: A Novel Dataset for Compositional Group Activity Understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 362–363.
- [111] Ke Chen, Chen Change Loy, Shaogang Gong, and Tao Xiang. 2012. Feature mining for localised crowd counting. In *British Machine Vision Conference*. <https://doi.org/10.5244/C.26.21>
- [112] Lei Chen, R Travis Rose, Ying Qiao, Irene Kimbara, Fey Parrill, Haleema Welji, Tony Xu Han, Jilin Tu, Zhongqiang Huang, Mary Harper, et al. 2005. VACE multimodal meeting corpus. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 40–51.
- [113] Lulu Chen, Hong Wei, and James Ferryman. [n.d.]. *Pattern Recognition Letters* ([n. d.]).
- [114] Dong Seon Cheng, Hugues Salamin, Pietro Salvagnini, Marco Cristani, Alessandro Vinciarelli, and Vittorio Murino. 2014. Predicting online lecture ratings based on gesturing and vocal behavior. *Journal on Multimodal User Interfaces* 8, 2 (2014), 151–160. <https://doi.org/10.1007/s12193-013-0142-z>
- [115] Huaining Cheng and Soon M Chung. 2018. Action recognition from point cloud patches using discrete orthogonal moments. *Multimedia Tools and Applications* 77, 7 (2018), 8213–8236. <https://doi.org/10.1007/s11042-017-4711-0>
- [116] Hong Cheng, Zicheng Liu, Yang Zhao, Guo Ye, and Xinghai Sun. 2014. Real world activity summary for senior home monitoring. *Multimedia Tools and Applications* 70, 1 (2014), 177–197. <https://doi.org/10.1007/s11042-012-1162-5>
- [117] Zhongwei Cheng, Lei Qin, Yituo Ye, Qingming Huang, and Qi Tian. 2012. Human Daily Action Analysis with Multi-view and Color-Depth Data. In *ECCV Workshops and Demonstrations*, Andrea Fusiello, Vittorio Murino, and Rita Cucchiara (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 52–61.
- [118] Roman Chereshevnev and Attila Kertész-Farkas. 2018. HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks. In *Analysis of Images, Social Networks and Texts*, Wil M P van der Aalst, Dmitry I Ignatov, Michael Khachay, Sergei O Kuznetsov, Victor Lempitsky, Irina A Lomazova, Natalia Loukachevitch, Amedeo Napoli, Alexander Panchenko, Panos M Pardalos, Andrey V Savchenko, and Stanley Wasserman (Eds.). Springer International Publishing, Cham, 131–141.
- [119] Wongun Choi, Khuram Shahid, and Silvio Savarese. 2009. What are they doing?: Collective activity classification using spatio-temporal relationship among people. In *IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*. IEEE, 1282–1289. <https://doi.org/10.1109/ICCVW.2009.5457461>
- [120] Huang-cheng Chou, Wei-cheng Lin, Chyi-chang Li, and Chi-chun Lee. 2017. NNIME : The NTHU-NTUA Chinese Interactive Multimodal Emotion Corpus. In *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 292–298.
- [121] Julia F. Christensen, Anna Lambrechts, and Manos Tsakiris. 2019. The Warburg Dance Movement Library—The WADAMO Library: A Validation Study. *Perception* 48, 1 (2019), 26–57. <https://doi.org/10.1177/0301006618816631>
- [122] Julia F. Christensen, Marcos Nadal, Camilo José Cela-Conde, and Antoni Gomila. 2014. A norming study and library of 203 dance movements. *Perception* 43, 2-3 (2014), 178–206. <https://doi.org/10.1068/p7581>
- [123] Addwiteey Chrungoo, S S Manimaran, and Balaraman Ravindran. 2014. Activity Recognition for Natural Human Robot Interaction. In *Social Robotics*, Michael Beetz, Benjamin Johnston, and Mary-Anne Williams (Eds.). Springer International Publishing, Cham, 84–94.
- [124] Jia Luen Chua, Yoong Choon Chang, and Wee Keong Lim. 2015. A simple vision-based fall detection technique for indoor video surveillance. *Signal, Image and Video Processing* 9, 3 (2015), 623–633. <https://doi.org/10.1007/s11760-013-0493-7>
- [125] Jihoon Chung, Cheng-hsin Wu, Hsuan-ru Yang, Yu-Wing Tai, and Chi-Keung Tang. 2020. HAA500: Human-Centric Atomic Action Dataset with Curated Videos. *arXiv* (2020).
- [126] Martin Clayton, Laura Leante, and Simone Tarsitani. 2018. IEMP North Indian Raga. <https://doi.org/10.17605/OSF.IO/KS325>.
- [127] CMUGraphicsLabMotionCapturedataset [n.d.]. CMU Graphics Lab Motion Capture Database. <http://mocap.cs.cmu.edu/>.

- [128] Davide Conigliaro, Paolo Rota, Francesco Setti, Chiara Bassetti, Nicola Conci, Nicu Sebe, and Marco Cristani. 2015. The s-hock dataset: Analyzing crowds at the stadium. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2039–2047.
- [129] Christopher Conly, Paul Doliotis, Pat Jangyodsuk, Rommel Alonzo, and Vassilis Athitsos. 2013. Toward a 3D body part detection video dataset and hand tracking benchmark. In *6th International Conference on Pervasive Technologies Related to Assistive Environments*. 1–6.
- [130] A Cook, B Mandal, D Berry, and M Johnson. 2019. Towards Automatic Screening of Typical and Atypical Behaviors in Children With Autism. *IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (2019). <https://doi.org/10.1109/DSAA.2019.00065>
- [131] Claudio Coppola, Serhan Cosar, Diego R Faria, and Nicola Bellotto. 2020. Social Activity Recognition on Continuous RGB-D Video Sequences. *International Journal of Social Robotics* 12, 1 (2020), 201–215. <https://doi.org/10.1007/s12369-019-00541-y>
- [132] Claudio Coppola, Diego R. Faria, Urbano Nunes, and Nicola Bellotto. 2016. Social activity recognition based on probabilistic merging of skeleton features with proximity priors from RGB-D data. In *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 5055–5061. <https://doi.org/10.1109/IROS.2016.7759742>
- [133] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schi. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1080/17843286.1974.11735726>
- [134] Juan Abdon Miranda Correa, Mojtaba Khomami Abadi, Nicolae Sebe, and Ioannis Patras. 2018. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing* (2018).
- [135] Alexandre Coste, Benoit G Bardy, Stefan Janaqi, Piotr Slowiński, Krasimira Tsaneva-Atanasova, Juliette Lozano Goupil, and Ludovic Marin. 2020. Decoding identity from motion: how motor similarities colour our perception of self and others. *Psychological Research* (2020). <https://doi.org/10.1007/s00426-020-01290-8>
- [136] Serhan Coşar and Nicola Bellotto. 2020. Human Re-Identification with a Robot Thermal Camera Using Entropy-Based Sampling. *Journal of Intelligent Robotic Systems* 98, 1 (2020), 85–102. <https://doi.org/10.1007/s10846-019-01026-w>
- [137] R Crawford, L Gizzi, A Dieterich, Á Ni Mhuiris, and D Falla. 2018. Age-related changes in trunk muscle activity and spinal and lower limb kinematics during gait. *PLoS one* 13, 11 (2018), e0206514. <https://doi.org/10.1371/journal.pone.0206514>
- [138] Marco Cristani, Loris Bazzani, Giulia Paggetti, Andrea Fossati, Diego Tosato, Alessio Del Bue, Gloria Menegaz, and Vittorio Murino. 2011. Social interaction discovery by statistical analysis of F-formations. In *British Machine Vision Conference 2011*. 1–12. <https://doi.org/10.5244/C25.23>
- [139] Manuel P Cuellar, Maria Ros, Maria J Martin-Bautista, Y Le Borgne, and Gianluca Bontempi. 2015. An Approach for the Evaluation of Human Activities in Physical Therapy Scenarios. In *Mobile Networks and Management*, Ramón Agüero, Thomas Zinner, Rossitza Goleva, Andreas Timm-Giel, and Phuoc Tran-Gia (Eds.). Springer International Publishing, Cham, 401–414.
- [140] CVBASE06dataset [n.d.]. CVBASE '06 Dataset. <http://vision.fe.uni-lj.si/cvbase06/downloads.html>.
- [141] Božidara Cvetković, Boštjan Kaluža, Mitja Luštrek, and Matjaž Gams. 2012. Multi-Classifer Adaptive Training: Specialising an Activity Recognition Classifier Using Semi-supervised Learning. In *Ambient Intelligence*, Fabio Paternò, Boris de Ruyter, Panos Markopoulos, Carmen Santoro, Evert van Loenen, and Kris Luyten (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 193–207.
- [142] M Dallel, V Havard, D Baudry, and X Savatier. 2020. InHARD - Industrial Human Action Recognition Dataset in the Context of Industrial Collaborative Robotics. In *IEEE International Conference on Human-Machine Systems (ICHMS)*. <https://doi.org/10.1109/ICHMS49158.2020.9209531>
- [143] Roeland De Geest, Efstratios Gavves, Amir Ghodrati, Zhenyang Li, Cees Snoek, and Tinne Tuytelaars. 2016. Online action detection. In *European Conference on Computer Vision*. Springer, 269–284.
- [144] Fernando De La Torre, Jessica Hodgins, Adam W. Bargteil, Xavier Martin, Justin C. Macey, Alex Collado, and Pep Beltran. 2008. *Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database*. Technical Report April. 19 pages.
- [145] Jan de Wit, Bram Willemsen, Mirjam de Haas, Emiel Kraemer, Paul Vogt, Marije Merckens, Reinjet Oostdijk, Chani Savelberg, Sabine Verdult, and Pieter Wolfert. 2019. Playing charades with a robot: collecting a large dataset of human gestures through HRI. In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 634–635.
- [146] Brian DeCann, Arun Ross, and Jeremy Dawson. 2013. Investigating gait recognition in the short-wave infrared (SWIR) spectrum: dataset and challenges. In *Biometric and Surveillance Technology for Human and Activity Identification X*, Vol. 8712. International Society for Optics and Photonics, 87120J.
- [147] Arnaud Decatoire, Sophie-Anne Beauprez, Jean Pylouster, Patrick Lacouture, Yannick Blandin, and Christel Bidet-Ildei. 2019. PLAViMoP: How to standardize and simplify the use of point-light displays. *Behavior Research Methods* 51, 6 (2019), 2573–2596. <https://doi.org/10.3758/s13428-018-1112-x>
- [148] Benoît Delachaux, Julien Rebetez, Andres Perez-Uribe, and Héctor Fabio Satizábal Mejía. 2013. Indoor Activity Recognition by Combining One-vs.-All Neural Network Classifiers Exploiting Wearable and Depth Sensors. In *Advances in Computational Intelligence*, Ignacio Rojas, Gonzalo Joya, and Joan Cabestany (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 216–223.
- [149] Claire-Hélène Demarty, Cédric Penet, Guillaume Gravier, and Mohammad Soleymani. 2012. A Benchmarking Campaign for the Multimodal Detection of Violent Scenes in Movies. In *ECCV Workshops and Demonstrations*, Andrea Fusiello, Vittorio Murino, and Rita Cucchiara (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 416–425.
- [150] B E Demiröz, İ A O Eroğlu, A A Salah, and L Akarun. 2012. Feature-based tracking on a multi-omnidirectional camera dataset. In *5th International Symposium on Communications, Control and Signal Processing*. <https://doi.org/10.1109/ISCCSP.2012.6217867>



- [151] Jingjing Deng, Xianghua Xie, and Shangming Zhou. 2014. Conversational Interaction Recognition Based on Bodily and Facial Movement. In *Image Analysis and Recognition*, Aurélio Campilho and Mohamed Kamel (Eds.). Springer International Publishing, Cham, 237–245.
- [152] Andrea D'Eusanio, Stefano Pini, Guido Borghi, Roberto Vezzani, and Rita Cucchiara. 2019. Manual Annotations on Depth Maps for Human Pose Estimation. In *ICIA*, Elisa Ricci, Samuel Rota Bulò, Cees Snoek, Oswald Lanz, Stefano Messelodi, and Nicu Sebe (Eds.). Springer International Publishing, Cham, 233–244.
- [153] Elham Dolatabadi, Ying Xuan Zhi, Bing Ye, Marge Coahran, Giorgia Lupinacci, Alex Mihailidis, Rosalie Wang, and Babak Taati. 2017. The toronto rehab stroke pose dataset to detect compensation during stroke rehabilitation therapy. In *11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 375–381.
- [154] Piotr Dollár, Christian Wojek, Bernt Schiele, and Pietro Perona. 2009. Pedestrian detection: A benchmark. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. 304–311. <https://doi.org/10.1109/CVPRW.2009.5206631>
- [155] Alex Dominguez-Sanchez, Sergio Orts-Escolano, and Miguel Cazorla. 2017. Recognizing Pedestrian Direction Using Convolutional Neural Networks. In *Advances in Computational Intelligence*, Ignacio Rojas, Gonzalo Joya, and Andreu Catala (Eds.). Springer International Publishing, Cham, 235–245.
- [156] Jianmin Dong and Zhongmin Cai. 2016. User Authentication Using Motion Sensor Data from Both Wearables and Smartphones. In *Chinese Conference on Biometric Recognition*, Zhisheng You, Jie Zhou, Yunhong Wang, Zhenan Sun, Shiguang Shan, Weishi Zheng, Jianjiang Feng, and Qijun Zhao (Eds.). Springer, Cham, 756–764.
- [157] Junting Dong, Qing Shuai, Yuanqing Zhang, Xian Liu, Xiaowei Zhou, and Hujun Bao. 2020. Motion Capture from Internet Videos. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 210–227.
- [158] Ellen Douglas-Cowie, Roddy Cowie, Ian Sneddon, Cate Cox, Orla Lowry, Margaret Mcrorie, Jean-Claude Martin, Laurence Devillers, Sarkis Abrilian, Anton Batliner, et al. 2007. The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In *International conference on affective computing and intelligent interaction*. Springer, 488–500.
- [159] N Dounskaia, Y Shimansky, B K Ganter, and M E Vidt. 2020. A simple joint control pattern dominates performance of unconstrained arm movements of daily living tasks. *PLoS one* 15, 7 (2020), e0235813. <https://doi.org/10.1371/journal.pone.0235813>
- [160] Philippe Dreuw, David Rybach, Thomas Deselaers, Morteza Zahedi, and Hermann Ney. 2007. Speech recognition techniques for a sign language recognition system. In *Annual Conference of the International Speech Communication Association, INTERSPEECH*. 705–708.
- [161] Camille Dupont, Luis Tobias, and Bertrand Luvion. 2017. Crowd-11: A dataset for fine grained crowd behaviour analysis. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 9–16.
- [162] Vibekananda Dutta and Teresa Zielinska. 2019. Predicting Human Actions Taking into Account Object Affordances. *Journal of Intelligent Robotic Systems* 93, 3 (2019), 745–761. <https://doi.org/10.1007/s10846-018-0815-7>
- [163] V Dutta and T Zielinska. 2020. Prognosing Human Activity Using Actions Forecast and Structured Database. *IEEE Access* 8 (2020). <https://doi.org/10.1109/ACCESS.2020.2963933>
- [164] Neelam Dwivedi, Dushyant Kumar Singh, and Dharmender Singh Kushwaha. 2020. Orientation Invariant Skeleton Feature (OISF): a new feature for Human Activity Recognition. *Multimedia Tools and Applications* 79, 29 (2020), 21037–21072. <https://doi.org/10.1007/s11042-020-08902-w>
- [165] Mahsa Ehsanpour, Alireza Abedin, Fatemeh Saleh, Javen Shi, Ian Reid, and Hamid Rezatofighi. 2020. Joint Learning of Social Groups, Individuals Action and Sub-group Activities in Videos. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 177–195.
- [166] Amani Elaoud, Walid Barhoumi, Ezzeddine Zagrouba, and Brahim Agrebi. 2020. Skeleton-based comparison of throwing motion for handball players. *Journal of Ambient Intelligence and Humanized Computing* 11, 1 (2020), 419–431. <https://doi.org/10.1007/s12652-019-01301-6>
- [167] Omar Elharrouss, Noor Almaadeed, and Somaya Al-Maadeed. 2020. Mhad: multi-human action dataset. In *Fourth International Congress on Information and Communication Technology*. Springer, 333–341.
- [168] A Elhayek, E de Aguiar, A Jain, J Thompson, L Pishchulin, M Andriluka, C Bregler, B Schiele, and C Theobalt. 2017. MARCOmI—ConvNet-Based MARKer-Less Motion Capture in Outdoor and Indoor Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017). <https://doi.org/10.1109/TPAMI.2016.2557779>
- [169] Chris Ellis, Syed Zain Masood, Marshall F Tappen, Joseph J LaViola, and Rahul Sukthankar. 2013. Exploring the Trade-off Between Accuracy and Observational Latency in Action Recognition. *International Journal of Computer Vision* 101, 3 (2013), 420–436. <https://doi.org/10.1007/s11263-012-0550-7>
- [170] Atis Elsts, Tilo Burghardt, Dallen Byrne, Massimo Camplani, Dima Damen, Xenofon Fafoutis, Sion Hannuna, William Harwin, Michael Holmes, and Balazs Janko. 2018. A guide to the sphere 100 homes study dataset. *arXiv* (2018).
- [171] EmotionalBodyMotiondataset [n.d.]. Emotional Body Motion Database. <http://ebmdb.tuebingen.mpg.de/>.
- [172] Markus Enzweiler and Dariu M. Gavrilă. 2009. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 12 (2009), 2179–2195. <https://doi.org/10.1109/TPAMI.2008.260>
- [173] Sergio Escalera, Xavier Baró, Jordi González, Miguel A Bautista, Meysam Madadi, Miguel Reyes, Victor Ponce-López, Hugo J Escalante, Jamie Shotton, and Isabelle Guyon. 2014. ChaLearn Looking at People Challenge 2014: Dataset and Results. In *ECCV*, Lourdes Agapito, Michael M Bronstein, and Carsten Rother (Eds.). Springer International Publishing, Cham, 459–473.
- [174] Sergio Escalera, Cristian Sminchisescu, Richard Bowden, Stan Sclaroff, Jordi González, Xavier Baró, Miguel Reyes, Isabelle Guyon, Vassilis Athitsos, Hugo Escalante, Leonid Sigal, and Antonis Argyros. 2013. ChaLearn multi-modal gesture recognition 2013: grand challenge and workshop

- summary. (2013), 365–368. <https://doi.org/10.1145/2522848.2532597>
- [175] Andreas Ess, Bastian Leibe, Konrad Schindler, and Luc van Gool. 2009. Robust multiperson tracking from a mobile platform. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 10 (2009), 1831–1846. <https://doi.org/10.1109/TPAMI.2009.109>
- [176] Andreas Ess, Bastian Leibe, and Luc Van Gool. 2007. Depth and appearance for mobile scene analysis. In *Proceedings of the IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2007.4409092>
- [177] Slim Essid, Xinyu Lin, Marc Gowing, Georgios Kordelas, Anil Aksay, Philip Kelly, Thomas Fillon, Qianni Zhang, Alfred Dielmann, Vlado Kitanovski, Robin Tournemene, Aymeric Masurelle, Ebroul Izquierdo, Noel E O'Connor, Petros Daras, and Gaël Richard. 2013. A multi-modal dance corpus for research into interaction between humans in virtual environments. *Journal on Multimodal User Interfaces* 7, 1 (2013), 157–170. <https://doi.org/10.1007/s12193-012-0109-5>
- [178] Vito Evola and Joanna Skubisz. 2019. Coordinated Collaboration and Nonverbal Social Interactions: A Formal and Functional Analysis of Gaze, Gestures, and Other Body Movements in a Contemporary Dance Improvisation Performance. *Journal of Nonverbal Behavior* 43, 4 (2019), 451–479. <https://doi.org/10.1007/s10919-019-00313-2>
- [179] Nazli Farajidavar, Teófilo De Campos, Josef Kittler, and Fei Yan. 2011. Transductive transfer learning for action recognition in tennis games. In *IEEE International Conference on Computer Vision*. 1548–1553. <https://doi.org/10.1109/ICCVW.2011.6130434>
- [180] Hayden Faulkner and Anthony Dick. 2017. TennisSet: A Dataset for Dense Fine-Grained Event Recognition, Localisation and Description. In *International Conference on Digital Image Computing: Techniques and Applications*. 1–8. <https://doi.org/10.1109/DICTA.2017.8227494>
- [181] Na Feng, Zikai Song, Junqing Yu, Yi-Ping Phoebe Chen, Yizhu Zhao, Yunfeng He, and Tao Guan. 2020. SSET: a dataset for shot segmentation, event detection, player tracking in soccer videos. *Multimedia Tools and Applications* 79, 39 (2020), 28971–28992. <https://doi.org/10.1007/s11042-020-09414-3>
- [182] James Ferryman and Anna-Louise Ellis. 2014. Performance evaluation of crowd image analysis using the PETS2009 dataset. *Pattern Recognition Letters* 44 (2014), 3–15.
- [183] T Fettrow, H Reimann, D Grenet, E Thompson, J Crenshaw, J Higginson, and J Jeka. 2019. Interdependence of balance mechanisms during bipedal locomotion. *PLoS one* 14, 12 (2019), e0225902. <https://doi.org/10.1371/journal.pone.0225902>
- [184] Panagiotis Paraskevas Filintisis, Niki Efthymiou, Petros Koutras, Gerasimos Potamianos, and Petros Maragos. 2019. Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child–Robot Interaction. *IEEE Robotics and Automation Letters* 4, 4 (2019), 4011–4018. <https://doi.org/10.1109/lra.2019.2930434>
- [185] Robert B Fisher. 2004. The PETS04 surveillance ground-truth data sets. In *Proc. 6th IEEE international workshop on performance evaluation of tracking and surveillance*. 1–5.
- [186] François Fleuret, Jérôme Berclaz, Richard Lengagne, and Pascal Fua. 2008. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 2 (2008), 267–282. <https://doi.org/10.1109/TPAMI.2007.1174>
- [187] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, and Hermann Ney. 2012. RWTH-PHOENIX-weather: A large vocabulary sign language recognition and translation corpus. In *8th International Conference on Language Resources and Evaluation, LREC 2012*. 3785–3789.
- [188] Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. 2017. Automatically Classifying User Engagement for Dynamic Multi-party Human–Robot Interaction. *International Journal of Social Robotics* 9, 5 (2017), 659–674. <https://doi.org/10.1007/s12369-017-0414-y>
- [189] Nesrine Fourati and Catherine Pelachaud. 2014. Emily: Emotional body expression in daily actions database. In *9th International Conference on Language Resources and Evaluation, LREC*. 3486–3493.
- [190] Annalisa Franco, Antonio Magnani, and Dario Maio. 2017. Joint Orientations from Skeleton Data for Human Activity Recognition. In *Image Analysis and Processing*, Sebastiano Battiato, Giovanni Gallo, Raimondo Schettini, and Filippo Stanco (Eds.). Springer International Publishing, Cham, 152–162.
- [191] Reginaldo K Fukuchi, Claudiane A Fukuchi, and Marcos Duarte. 2017. A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics. *PeerJ* 5 (2017), e3298.
- [192] Daisuke Furuki and Ken Takiyama. 2020. A data-driven approach to decompose motion data into task-relevant and task-irrelevant components in categorical outcome. *Scientific Reports* 10, 1 (2020).
- [193] Alexander Gabriel, Serhan Coşar, Nicola Bellotto, and Paul Baxter. 2019. A dataset for action recognition in the wild. In *Annual Conference Towards Autonomous Robotic Systems*. Springer, 362–374.
- [194] Salvatore Gaglio, Giuseppe Lo Re, and Marco Morana. 2014. Human Activity Recognition Process Using 3-D Posture Data. 45, 5 (2014), 586–597. <https://doi.org/10.1109/THMS.2014.2377111>
- [195] Varun Ganapathi, Christian Plagemann, Daphne Koller, and Sebastian Thrun. 2012. Real-Time Human Pose Tracking from Range Data. In *ECCV*, Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 738–751.
- [196] Chenqiang Gao, Yinhe Du, Jiang Liu, Jing Lv, Luyu Yang, Deyu Meng, and Alexander G Hauptmann. 2016. Infar dataset: Infrared action recognition at different times. *Neurocomputing* 212 (2016), 36–47.
- [197] Chenqiang Gao, Luyu Yang, Yinhe Du, Zeming Feng, and Jiang Liu. 2016. From constrained to unconstrained datasets: an evaluation of local action descriptors and fusion strategies for interaction recognition. *World Wide Web* 19, 2 (2016), 265–276. <https://doi.org/10.1007/s11280-015-0348-y>
- [198] Jibin Gao, Wei-Shi Zheng, Jia-Hui Pan, Chengying Gao, Yaowei Wang, Wei Zeng, and Jianhuang Lai. 2020. An Asymmetric Modeling for Action Assessment. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham,

- 222–238.
- [199] Zan Gao, Tao-tao Han, Hua Zhang, Yan-bing Xue, and Guang-ping Xu. 2018. MMA: a multi-view and multi-modality benchmark dataset for human action recognition. *Multimedia Tools and Applications* 77, 22 (2018), 29383–29404. <https://doi.org/10.1007/s11042-018-5833-8>
- [200] G Gaziv, L Noy, Y Liron, and U Alon. 2017. A reduced-dimensionality approach to uncovering dyadic modes of body motion in conversations. *PLoS one* 12, 1 (2017), e0170786. <https://doi.org/10.1371/journal.pone.0170786>
- [201] M D Geil, Z Safaeepour, B Giavedoni, and C P Coulter. 2020. Walking kinematics in young children with limb loss using early versus traditional prosthetic knee prescription protocols. *PLoS one* 15, 4 (2020), e0231401. <https://doi.org/10.1371/journal.pone.0231401>
- [202] Saeed Ghorbani, Nikolaus F Troje, Anne Thaler, C V Mar, and Konrad Kording. 2020. MoVi: A Large Multipurpose Motion and Video Dataset. *arXiv* (2020).
- [203] S Ginosar, A Bar, G Kohavi, C Chan, A Owens, and J Malik. 2019. Learning Individual Styles of Conversational Gesture. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2019.00361>
- [204] T Giraud, M Soury, J Hua, A Delaborde, M Tahon, D A G Jauregui, V Eyharabide, E Filaire, C Le Scanff, L Devillers, B Isableu, and J C Martin. 2013. Multimodal Expressions of Stress during a Public Speaking Task: Collection, Annotation and Global Analyses. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (2013). <https://doi.org/10.1109/ACII.2013.75>
- [205] Nikolaos Gkalelis, Hansung Kim, Adrian Hilton, Nikos Nikolaidis, and Ioannis Pitas. 2009. The i3DPost multi-view and 3D human action/interaction database. In *6th European Conference for Visual Media Production*. 159–168. <https://doi.org/10.1109/CVMP.2009.19>
- [206] Michael Glodek, Georg Layher, Felix Heilemann, Florian Gawrilowicz, Günther Palm, Friedhelm Schwenker, and Heiko Neumann. 2015. uulMAD – A Human Action Recognition Dataset for Ground-Truth Evaluation and Investigation of View Invariances. In *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*, Friedhelm Schwenker, Stefan Scherer, and Louis-Philippe Morency (Eds.). Springer International Publishing, Cham, 77–91.
- [207] Øyvind Gløersen and Peter Federolf. 2016. Predicting missing marker trajectories in human motion data using marker intercorrelations. *PLoS one* 11, 3 (2016), e0152616.
- [208] Arturo Gomez Chavez, Andrea Ranieri, Davide Chiarella, Enrica Zereik, Anja Babić, and Andreas Birk. 2019. CADDY underwater stereo-vision dataset for human–robot interaction (HRI) in the context of diver activities. *Journal of Marine Science and Engineering* 7, 1 (2019), 16.
- [209] Lena Gorelick, Moshe Blank, Eli Shechtman, Michal Irani, and Ronen Basri. 2007. Actions as space-time shapes. *IEEE transactions on pattern analysis and machine intelligence* 29, 12 (2007), 2247–2253.
- [210] Ilaria Gori, Jivko Sinapov, Priyanka Khante, Peter Stone, and J K Aggarwal. 2015. Robot-Centric Activity Recognition ‘in the Wild’. In *Social Robotics*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Springer International Publishing, Cham, 224–234.
- [211] S Gourgari, G Goudelis, K Karpouzis, and S Kollias. 2013. THETIS: Three Dimensional Tennis Shots a Human Action Dataset. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. <https://doi.org/10.1109/CVPRW.2013.102>
- [212] Harry J. Griffin, Min S.H. Aung, Bernardino Romera-Paredes, Ciaran McLoughlin, Gary McKeown, William Curran, and Nadia Bianchi-Berthouze. 2013. Laughter type recognition from whole body motion. In *Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*. 349–355. <https://doi.org/10.1109/ACII.2013.64>
- [213] Ralph Gross and Jianbo Shi. 2001. The CMU motion of body (MOBO) database. (2001).
- [214] Chunhui Gu, Chen Sun, David A. Ross, Carl Vondrick, Caroline Pantofaru, Yeqing Li, Sudheendra Vijayanarasimhan, George Toderici, Susanna Ricco, Rahul Sukthankar, Cordelia Schmid, and Jitendra Malik. 2017. AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions. In *IEEE Conference on Computer Vision and Pattern Recognition*.
- [215] G Guerra-Filho and A Biswas. 2011. The human motion database: A cognitive and parametric sampling of human motion. In *IEEE International Conference on Automatic Face Gesture Recognition (FG)*. <https://doi.org/10.1109/FG.2011.5771381>
- [216] Hatice Gunes and Massimo Piccardi. 2006. Bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *International Conference on Pattern Recognition*. 1148–1153. <https://doi.org/10.1109/ICPR.2006.39>
- [217] Abhinav Gupta, Aniruddha Kembhavi, and Larry S. Davis. 2009. Observing human-object interactions: Using spatial and functional compatibility for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 10 (2009), 1775–1789. <https://doi.org/10.1109/TPAMI.2009.83>
- [218] Arpan Gupta and Sakthi Balan M. 2018. Temporal Cricket Stroke Localization from Untrimmed Highlight Videos. In *11th Indian Conference on Computer Vision, Graphics and Image Processing*. <https://doi.org/10.1145/3293353.3293415>
- [219] Isabelle Guyon, Vassilis Athitsos, Pat Jangyodsuk, and Hugo Jair Escalante. 2014. *Machine Vision and Applications* 25, 8 (2014), 1929–1951. <https://doi.org/10.1007/s00138-014-0596-3>
- [220] T Hachaj, M R Ogiela, and M Piekarczyk. 2017. The open online repository of karate motion capture data: A tool for scientists and sport educators. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (2017). <https://doi.org/10.1109/SSCI.2017.8285270>
- [221] Simon Hadfield and Richard Bowden. 2013. Hollywood 3D: Recognizing actions in 3D natural scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3398–3405. <https://doi.org/10.1109/CVPR.2013.436>
- [222] Thomas Hanke, Marc Schuder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in Size and Depth. In *9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*. 75–82.

- [223] Albert Haque, Boya Peng, Zelun Luo, Alexandre Alahi, Serena Yeung, and Li Fei-Fei. 2016. Towards Viewpoint Invariant 3D Human Pose Estimation. In *ECCV*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 160–177.
- [224] Steen Harsted, Anders Holsgaard-Larsen, Lise Hestbæk, Eleanor Boyle, and Henrik Hein Lauridsen. 2019. Concurrent validity of lower extremity kinematics and jump characteristics captured in pre-school children by a markerless 3D motion capture system. *Chiropractic Manual Therapies* 27, 1 (2019), 39. <https://doi.org/10.1186/s12998-019-0261-z>
- [225] Md Kamrul Hasan, Wasifur Rahman, Amir Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, and Louis-Philippe Morency. 2019. Ur-funny: A multimodal language dataset for understanding humor. *arXiv* (2019).
- [226] Mahdi Hashemzadeh, Gang Pan, and Min Yao. 2014. Counting moving people in crowds using motion statistics of feature-points. *Multimedia Tools and Applications* 72, 1 (2014), 453–487. <https://doi.org/10.1007/s11042-013-1367-2>
- [227] Atsushi Hashimoto, Tetsuro Sasada, Yoko Yamakata, Shinsuke Mori, and Michihiko Minoh. 2014. Kusk dataset: Toward a direct understanding of recipe text and human cooking activity. In *2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 583–588.
- [228] Mohamed Hassan, Vasileios Choutas, DIMITRIOS Tzionas, and Michael Black. 2019. Resolving 3D human pose ambiguities with 3D scene constraints. In *IEEE International Conference on Computer Vision*. 2282–2292. <https://doi.org/10.1109/ICCV.2019.00237>
- [229] Tal Hassner, Yossi Itcher, and Orit Kliper-Gross. 2012. Violent flows: Real-time detection of violent crowd behavior. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 1–6. <https://doi.org/10.1109/CVPRW.2012.6239348>
- [230] Dilana Hazer-Rau, Sascha Meudt, Andreas Daucher, Jennifer Spohrs, Holger Hoffmann, Friedhelm Schwenker, and Harald C. Traue. 2020. The uulmmac database—a multimodal affective corpus for affective computing in human-computer interaction. *Sensors (Switzerland)* 20, 8 (2020), 1–33. <https://doi.org/10.3390/s20082308>
- [231] Jian He, Cheng Zhang, Xinlin He, and Ruihai Dong. 2020. Visual Recognition of traffic police gestures with convolutional pose machine and handcrafted features. *Neurocomputing* 390 (2020), 248–259. <https://doi.org/10.1016/j.neucom.2019.07.103>
- [232] Fabian Caba Heilbron, Victor Escorcia, Bernard Ghanem, and Juan Carlos Nieves. 2015. ActivityNet: A large-scale video benchmark for human activity understanding. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 961–970. <https://doi.org/10.1109/CVPR.2015.7298698>
- [233] Danica Hendry, Kevin Chai, Amity Campbell, Luke Hopper, Peter O’Sullivan, and Leon Straker. 2020. Development of a Human Activity Recognition System for Ballet Tasks. *Sports Medicine - Open* 6, 1 (2020), 10. <https://doi.org/10.1186/s40798-020-0237-5>
- [234] Erik Herrmann, Han Du, André Antakli, Dmitri Rubinstein, René Schubotz, Janis Sprenger, Somayeh Hosseini, Noshaba Cheema, Ingo Zinnikus, and Martin Manns. 2019. Motion Data and Model Management for Applied Statistical Motion Synthesis. In *STAG*. 79–88.
- [235] Nikolas Hesse, Sergi Pujades, Javier Romero, Michael J Black, Christoph Bodensteiner, Michael Arens, Ulrich G Hofmann, Uta Tacke, Mijna Hadders-Algra, Raphael Weinberger, Wolfgang Müller-Felber, and A Sebastian Schroeder. 2018. Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis. In *Medical Image Computing and Computer Assisted Intervention*, Alejandro F Frangi, Julia A Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger (Eds.). Springer International Publishing, Cham, 792–800.
- [236] HIDUMDdataset [n.d.]. Human Identification at a Distance: UMD Database. <http://www.umiacs.umd.edu/labs/pirl/hid/data.html>.
- [237] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. 2011. Person re-identification by descriptive and discriminative classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 6688 LNCS. 91–102. [https://doi.org/10.1007/978-3-642-21227-7\\_9](https://doi.org/10.1007/978-3-642-21227-7_9)
- [238] Basma Hisham and Alaa Hamouda. 2019. Supervised learning classifiers for Arabic gestures recognition using Kinect V2. *SN Applied Sciences* 1, 7 (2019), 768. <https://doi.org/10.1007/s42452-019-0771-2>
- [239] Chiung Ching Ho, Hu Ng, Wooi Haw Tan, Kok Why Ng, Hau Lee Tong, Timothy Tzen Vun Yap, Pei Fen Chong, C. Eswaran, and Junaidi Abdullah. 2013. MMU GASPFA: A COTS multimodal biometric database. *Pattern Recognition Letters* 34, 15 (2013), 2043–2050. <https://doi.org/10.1016/j.patrec.2013.01.027>
- [240] Minh Hoai and Andrew Zisserman. 2015. Thread-Safe: Towards Recognizing Human Actions Across Shot Boundaries. In *ACCV*, Daniel Cremers, Ian Reid, Hideo Saito, and Ming-Hsuan Yang (Eds.). Springer International Publishing, Cham, 222–237.
- [241] Ramona Hoffmann, Bertram Taetz, Markus Miezal, Gabriele Bleser, and Sigrid Leyendecker. 2020. On optical data-guided optimal control simulations of human motion. *Multibody System Dynamics* 48, 1 (2020), 105–126.
- [242] M Hofmann and D M Gavrilu. 2012. Multi-view 3D Human Pose Estimation in Complex Environment. *International Journal of Computer Vision* 96, 1 (2012), 103–124. <https://doi.org/10.1007/s11263-011-0451-1>
- [243] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. 2014. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation* 25, 1 (2014), 195–206. <https://doi.org/10.1016/j.jvcir.2013.02.006>
- [244] Martin Hofmann, Shamik Sural, and Gerhard Rigoll. 2011. Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *19th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, WSCG. 99–104.
- [245] Sarah Hood, Marshall K Ishmael, Andrew Gunnell, K B Foreman, and Tommaso Lenzi. 2020. A kinematic and kinetic dataset of 18 above-knee amputees walking at various speeds. *Scientific data* 7, 1 (2020), 1–8.
- [246] Emdad Hossain and Girija Chetty. 2013. Multimodal Feature Learning for Gait Biometric Based Human Identity Recognition. In *Neural Information Processing*, Minhoo Lee, Akira Hirose, Zeng-Guang Hou, and Rhee Man Kil (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 721–728.

- [247] Tao Hu, Xinyan Zhu, Shaohua Wang, and Lian Duan. 2019. Human interaction recognition using spatial-temporal salient feature. *Multimedia Tools and Applications* 78, 20 (2019), 28715–28735. <https://doi.org/10.1007/s11042-018-6074-6>
- [248] Chun Hao Huang, Edmond Boyer, Nassir Navab, and Slobodan Ilic. 2014. Human shape and pose tracking using keyframes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3446–3453. <https://doi.org/10.1109/CVPR.2014.440>
- [249] Dong Huang, Shitong Yao, Yi Wang, and Fernando De La Torre. 2014. Sequential max-margin event detectors. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8691 LNCS. 410–424. [https://doi.org/10.1007/978-3-319-10578-9\\_27](https://doi.org/10.1007/978-3-319-10578-9_27)
- [250] Qingqiu Huang, Yu Xiong, Anyi Rao, Jiase Wang, and Dahua Lin. 2020. Movienet: A holistic dataset for movie understanding. *arXiv* (2020).
- [251] Hayley Hung and Gokul Chittaranjan. 2010. The idiap wolf corpus: exploring group behaviour in a competitive role-playing game. In *Proceedings of the 18th ACM international conference on Multimedia*. 879–882.
- [252] Tm Huynh, Mario Fritz, and Bernt Schiele. 2008. Discovery of activity patterns using topic models. In *0th International Conference on Ubiquitous Computing*. 10–19. <https://doi.org/10.1145/1409635.1409638>
- [253] Tâm Hunh, Ulf Blanke, and Bernt Schiele. 2007. Scalable recognition of daily activities with wearable sensors. In *International Symposium on Location-and Context-Awareness*. Springer, 50–67.
- [254] Tâm Hunh and Bernt Schiele. 2006. Unsupervised discovery of structure in activity data using multiple eigenspaces. In *International Symposium on Location-and Context-Awareness*. Springer, 151–167.
- [255] Bon-Woo Hwang, Sungmin Kim, and Seong-Wan Lee. 2005. 2D and 3D full-body gesture database for analyzing daily human gestures. In *International Conference on Intelligent Computing*. Springer, 611–620.
- [256] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. 2015. Multispectral pedestrian detection: Benchmark dataset and baseline. In *IEEE conference on computer vision and pattern recognition*. 1037–1045.
- [257] ICSdataset 2003. ICS Action Database. <http://www.miubiq.cs.titech.ac.jp/action/>.
- [258] Haroon Idrees, Amir R. Zamir, Yu Gang Jiang, Alex Gorban, Ivan Laptev, Rahul Sukthankar, and Mubarak Shah. 2017. The THUMOS challenge on action recognition for videos “in the wild”. *Computer Vision and Image Understanding* 155 (2017), 1–23. <https://doi.org/10.1016/j.cviu.2016.10.018>
- [259] IITKharagpurdataset 2017. Annotated Bharatnatyam Data Set. <http://hci.cse.iitkgp.ac.in/>.
- [260] IMDailyDepthActivitydataset [n.d.]. IM-DailyDepthActivity Dataset. <http://imlab.postech.ac.kr/databases.htm>.
- [261] Javed Imran and Balasubramanian Raman. 2020. Deep motion templates and extreme learning machine for sign language recognition. *The Visual Computer* 36, 6 (2020), 1233–1246. <https://doi.org/10.1007/s00371-019-01725-3>
- [262] IMSportingBehaviorsdataset [n.d.]. IM-Sporting Behaviors. <https://portals.au.edu.pk/imc/Pages/Datasets.aspx>.
- [263] Sozo Inoue, Naonori Ueda, Yasunobu Nohara, and Naoki Nakashima. 2015. Mobile activity recognition for a whole day: recognizing real nursing activities with big dataset. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1269–1280.
- [264] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. 2014. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (2014), 1325–1339. <https://doi.org/10.1109/TPAMI.2013.248>
- [265] Alexandros Iosifidis, Anastasios Tefas, and Ioannis Pitas. 2012. Activity-based person identification using fuzzy representation and discriminant learning. *IEEE Transactions on Information Forensics and Security* 7, 2 (2012), 530–542. <https://doi.org/10.1109/TIFS.2011.2175921>
- [266] Umar Iqbal, Anton Milan, and Juergen Gall. 2017. PoseTrack : Joint Multi-Person Pose Estimation and Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2011–2020.
- [267] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. 2012. The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security* 7, 5 (2012), 1511–1521.
- [268] Yumi Iwashita, Koichi Ogawara, and Ryo Kurazume. 2014. Identification of people walking along curved trajectories. *Pattern Recognition Letters* 48 (2014), 60–69. <https://doi.org/10.1016/j.patrec.2014.04.004>
- [269] Shrajal Jain, Aditya Rustagi, Sumeet Saurav, Ravi Saini, and Sanjay Singh. 2020. Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment. *Neural Computing and Applications* (2020). <https://doi.org/10.1007/s00521-020-05405-5>
- [270] Jinhyeok Jang, Dohyung Kim, Cheonshu Park, Minsu Jang, Jaeyeon Lee, and Jaehong Kim. 2020. ETRI-Activity3D: A Large-Scale RGB-D Dataset for Robots to Recognize Daily Activities of the Elderly. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 10990–10997.
- [271] David Jardim, Luis Nunes, and Miguel Sales Dias. 2016. Automatic Human Activity Segmentation and Labeling in RGBD Videos. In *Intelligent Decision Technologies*, Ireneusz Czarnowski, Alfonso Mateos Caballero, Robert J Howlett, and Lakhmi C Jain (Eds.). Springer International Publishing, 383–394.
- [272] D B Jayagopi, S Sheiki, D Klotz, J Wienke, J Odobez, S Wrede, V Khalidov, L Nyugen, B Wrede, and D Gatica-Perez. 2013. The vernissage corpus: A conversational Human-Robot-Interaction dataset. In *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. <https://doi.org/10.1109/HRI.2013.6483545>
- [273] H Jhuang, J Gall, S Zuffi, C Schmid, and M J Black. 2013. Towards Understanding Action Recognition. In *IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2013.396>
- [274] Yanli Ji, Hong Cheng, Yali Zheng, and Haoxin Li. 2015. Learning contrastive feature distribution model for interaction recognition. *Journal of Visual Communication and Image Representation* 33 (2015), 340–349. <https://doi.org/10.1016/j.jvcir.2015.10.001>

- [275] Y Ji, Y Yang, F Shen, H T Shen, and W Zheng. 2020. Arbitrary-view Human Action Recognition: A Varying-view RGB-D Action Dataset. *IEEE Transactions on Circuits and Systems for Video Technology* PP (2020). <https://doi.org/10.1109/TCSVT.2020.2975845>
- [276] Baoxiong Jia, Yixin Chen, Siyuan Huang, Yixin Zhu, and Song-Chun Zhu. 2020. LEMMA: A Multi-view Dataset for Learning Multi-agent Multi-task Activities. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 767–786.
- [277] Zhuolin Jiang, Zhe Lin, and Larry Davis. 2012. Recognizing human actions by learning and matching shape-motion prototype trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 3 (2012), 533–547. <https://doi.org/10.1109/TPAMI.2011.147>
- [278] Cheng-Bin Jin, Shengzhe Li, Trung Dung Do, and Hakil Kim. 2015. Real-Time Human Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras. In *PCM*, Yo-Sung Ho, Jitao Sang, Yong Man Ro, Junmo Kim, and Fei Wu (Eds.). Springer International Publishing, Cham, 330–339.
- [279] Kristiina Jokinen and Graham Wilcock. 2015. Modelling User Experience in Human-Robot Interactions. In *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*, Ronald Böck, Francesca Bonin, Nick Campbell, and Ronald Poppe (Eds.). Springer International Publishing, Cham, 45–56.
- [280] G D Jones, D C James, M Thacker, and D A Green. 2018. Parameters that remain consistent independent of pausing before gait-initiation during normal rise-to-walk behaviour delineated by sit-to-walk and sit-to-stand-and-walk. *PLoS one* 13, 10 (2018), e0205346. <https://doi.org/10.1371/journal.pone.0205346>
- [281] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. 2019. Panoptic Studio: A Massively Multiview System for Social Interaction Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 1 (2019), 190–204. <https://doi.org/10.1109/TPAMI.2017.2782743>
- [282] Luis Jure, Martín Rocamora, Simone Tarsitani, and Martin Clayton. 2018. IEMP Uruguayan Candombe. <https://doi.org/10.17605/OSF.IO/WFX7K>.
- [283] Agnes Just, Olivier Bernier, and Sébastien Marcel. 2004. *HMM and IOHMM for the recognition of mono-and bi-manual 3D hand gestures*. Technical Report. IDIAP.
- [284] Abdolrahim Kadkhodamohammadi and Nicolas Padoy. 2020. A generalizable approach for multi-view 3D human pose regression. *Machine Vision and Applications* 32, 1 (2020), 6. <https://doi.org/10.1007/s00138-020-01120-2>
- [285] Boštjan Kaluža, Violeta Mirchevska, Erik Dovgan, Mitja Luštrek, and Matjaž Gams. 2010. An agent-based approach to care in independent living. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 177–186. [https://doi.org/10.1007/978-3-642-16917-5\\_18](https://doi.org/10.1007/978-3-642-16917-5_18)
- [286] H G Kang and J B Dingwell. 2016. Differential Changes with Age in Multiscale Entropy of Electromyography Signals from Leg Muscles during Treadmill Walking. *PLoS one* 11, 8 (2016), e0162034. <https://doi.org/10.1371/journal.pone.0162034>
- [287] Katerina Karagiannaki, Athanasia Panousopoulou, and Panagiotis Tsakalides. 2016. A benchmark study on feature selection for human activity recognition. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 105–108. <https://doi.org/10.1145/2968219.2971421>
- [288] Andrej Karpathy, George Toderici, Sachin Shetty, Tommy Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1725–1732. <https://doi.org/10.1109/CVPR.2014.223>
- [289] Parinaz Kasebzadeh, Gustaf Hendeby, Carsten Fritsche, Fredrik Gunnarsson, and Fredrik Gustafsson. 2017. IMU dataset for motion and device mode classification. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 1–8.
- [290] Dimitrios Kastaniotis, Ilias Theodorakopoulos, George Economou, and Spiros Fotopoulos. 2013. Gait-based gender recognition using pose information for real time applications. In *18th International Conference on Digital Signal Processing, DSP*. IEEE, 1–6. <https://doi.org/10.1109/ICDSP.2013.6622766>
- [291] Christina Katsimerou, Joris Albeda, Alina Huldgtren, Ingrid Heynderickx, and Judith A. Redi. 2016. Crowdsourcing empathetic intelligence: The case of the annotation of EMMA database for emotion and mood recognition. *ACM Transactions on Intelligent Systems and Technology* 7, 4 (2016). <https://doi.org/10.1145/2897369>
- [292] A Kaur, A Mustafa, L Mehta, and A Dhall. 2018. Prediction and Localization of Student Engagement in the Wild. *2018 Digital Image Computing: Techniques and Applications (DICTA)* (2018). <https://doi.org/10.1109/DICTA.2018.8615851>
- [293] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The Kinetics Human Action Video Dataset. *arXiv* (2017). <https://doi.org/10.1145/3123266.3138874>
- [294] Vahid Kazemi, Magnus Burenius, Hossein Azizpour, and Josephine Sullivan. 2013. Multi-view body part recognition with random forests. In *British Machine Vision Conference 2013*. 1–11. <https://doi.org/10.5244/C.27.48>
- [295] Yan Ke, Rahul Sukthankar, and Martial Hebert. 2005. Efficient visual event detection using volumetric features. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 1. IEEE, 166–173.
- [296] Bruce D. Keefe, Matthias Villing, Chris Racey, Samantha L. Strong, Joanna Wincenciak, and Nick E. Barraclough. 2014. A database of whole-body action videos for the study of action, emotion, and untrustworthiness. *Behavior research methods* 46, 4 (2014), 1042–1051. <https://doi.org/10.3758/s13428-013-0439-6>
- [297] Vili Kellokumpu, Markus Särkiniemi, and Guoying Zhao. 2017. Affective Gait Recognition and Baseline Evaluation from Real World Samples. In *ACCV Workshops*, Chu-Song Chen, Jiwen Lu, and Kai-Kuang Ma (Eds.). Springer International Publishing, Cham, 567–575.

- [298] Dohyung Kim, Minsu Jang, and Jaehong Kim. 2016. Example-Based Retrieval System for Human Motion Data. In *6th International Conference on IT Convergence and Security (ICITCS)*. IEEE, 1–2.
- [299] Hansung Kim, Alun Evans, Josep Blat, and Adrian Hilton. 2018. Multimodal Visual Data Registration for Web-Based Visualization in Media Production. *IEEE Transactions on Circuits and Systems for Video Technology* 28, 4 (2018), 863–877. <https://doi.org/10.1109/TCSVT.2016.2642825>
- [300] Yeonho Kim and Daijin Kim. 2018. Real-time dance evaluation by markerless human pose estimation. *Multimedia Tools and Applications* 77, 23 (2018), 31199–31220. <https://doi.org/10.1007/s11042-018-6068-4>
- [301] A Kitzi, S Schröter, E Naroska, G Stockmanns, R Viga, and A Grabmaier. 2017. MoveHN-A database to support the development of motion based biosignal processing systems. In *5th European Signal Processing Conference (EUSIPCO)*. <https://doi.org/10.23919/EUSIPCO.2017.8081628>
- [302] Andrea Kleinsmith, Nadia Bianchi-Berthouze, and Anthony Steed. 2011. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 41, 4 (2011), 1027–1038. <https://doi.org/10.1109/TSMCB.2010.2103557>
- [303] O Kliper-Gross, T Hassner, and L Wolf. 2012. The Action Similarity Labeling Challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (2012). <https://doi.org/10.1109/TPAMI.2011.209>
- [304] Brian A. Knarr, Trisha M. Kesar, Darcy S. Reisman, Stuart A. Binder-Macleod, and Jill S. Higginson. 2013. Changes in the activation and function of the ankle plantar flexor muscles due to gait retraining in chronic stroke survivors. *Journal of neuroengineering and rehabilitation* 10 (2013), 12. <https://doi.org/10.1186/1743-0003-10-12>
- [305] K Kolykhalova, A Camurri, G Völpe, M Sanguineti, E Puppo, and R Niewiadomski. 2015. A multimodal dataset for the analysis of movement qualities in karate martial art. *7th International Conference on Intelligent Technologies for Interactive Entertainment (INTEIN)* (2015).
- [306] Quan Kong, Ziming Wu, Ziwei Deng, Martin Klunkigt, Bin Tong, and Tomokazu Murakami. 2019. MMAct: A large-scale dataset for cross modal human action understanding. In *IEEE International Conference on Computer Vision*. 8657–8666. <https://doi.org/10.1109/ICCV.2019.00875>
- [307] Yu Kong and Yun Fu. 2017. Max-Margin Heterogeneous Information Machine for RGB-D Action Recognition. *International Journal of Computer Vision* 123, 3 (2017), 350–371. <https://doi.org/10.1007/s11263-016-0982-6>
- [308] Yu Kong, Yunde Jia, and Yun Fu. 2012. Learning human interaction by interactive phrases. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7572 LNCS, PART 1 (2012), 300–313. [https://doi.org/10.1007/978-3-642-33718-5\\_22](https://doi.org/10.1007/978-3-642-33718-5_22)
- [309] Julian Francisco Pieter Kooij, Nicolas Schneider, Fabian Flohr, and Dariu M Gavrila. 2014. Context-Based Pedestrian Path Prediction. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 618–633.
- [310] Hema Swetha Koppula, Rudhir Gupta, and Ashutosh Saxena. 2013. Learning human activities and object affordances from RGB-D videos. *International Journal of Robotics Research* 32, 8 (2013), 951–970. <https://doi.org/10.1177/0278364913478446>
- [311] Barbara Krausz and Christian Bauckhage. 2012. Loveparade 2010: Automatic video analysis of a crowd disaster. *Computer Vision and Image Understanding* 116, 3 (2012), 307–319. <https://doi.org/10.1016/j.cviu.2011.08.006>
- [312] Hilde Kuehne, Ali Arslan, and Thomas Serre. 2014. The Language of Actions : Recovering the Syntax and Semantics of Goal-Directed Human Activities. In *IEEE conference on computer vision and pattern recognition*. 780–787.
- [313] Hilde Kuehne, Hueihan Jhuang, E Garrote, Tomaso Poggio, and Thomas Serre. 2011. HMDB51: A Large Video Database for Human Motion Recognition. In *International Conference on Computer Vision*, Wolfgang E Nagel, Dietmar H Kröner, and Michael M Resch (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 571–582.
- [314] Marek Kulbacki, Jakub Segen, Kamil Wereszczyński, and Adam Gudyś. 2014. VMAS: Massive Dataset of Multi-camera Video for Learning, Classification and Recognition of Human Actions. In *Intelligent Information and Database Systems*, Ngoc Thanh Nguyen, Boonwat Attachoo, Bogdan Trawiński, and Kulwadee Somboonviwat (Eds.). Springer International Publishing, Cham, 565–574.
- [315] A. Kurakin, Z. Zhang, and Z. Liu. 2012. A real time system for dynamic hand gesture recognition with a depth sensor. In *European Signal Processing Conference*. IEEE, 1975–1979.
- [316] Bogdan Kwolek and Michal Kepski. 2014. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine* 117, 3 (2014), 489–501. <https://doi.org/10.1016/j.cmpb.2014.09.005>
- [317] Bogdan Kwolek, Tomasz Krzeszowski, Agnieszka Michalczyk, and Henryk Josinski. 2014. 3D Gait Recognition Using Spatio-Temporal Motion Descriptors, Ngoc Thanh Nguyen, Boonwat Attachoo, Bogdan Trawiński, and Kulwadee Somboonviwat (Eds.). Springer International Publishing, Cham, 595–604.
- [318] Steven Landry and Myounghoon Jeon. 2020. Interactive sonification strategies for the motion and emotion of dance performances. *Journal on Multimodal User Interfaces* 14, 2 (2020), 167–186. <https://doi.org/10.1007/s12193-020-00321-3>
- [319] Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. 2008. Learning realistic human actions from movies. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.
- [320] Ivan Laptev and Patrick Pérez. 2007. Retrieving actions in movies. In *Proceedings of the IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2007.4409105>
- [321] Guillaume Lathoud, Jean-Marc Odobez, and Daniel Gatica-Perez. 2004. AV16. 3: An audio-visual corpus for speaker localization and tracking. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 182–195.
- [322] Ryan Layne, Sion Hannuna, Massimo Camplani, Jake Hall, Timothy M Hospedales, Tao Xiang, Majid Mirmehdi, and Dima Damen. 2017. A dataset for persistent multi-target multi-camera tracking in RGB-D. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 47–55.

- [323] Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. 2015. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking. *arXiv* (2015), 1–15. <http://arxiv.org/abs/1504.01942>
- [324] Karina Lebel, Patrick Boissy, Hung Nguyen, and Christian Duval. 2017. Inertial measurement systems for segments and joints kinematics assessment: towards an understanding of the variations in sensors accuracy. *BioMedical Engineering OnLine* 16, 1 (2017), 56. <https://doi.org/10.1186/s12938-017-0347-6>
- [325] Jinkue Lee and Hoeryong Jung. 2020. TUHAD: Taekwondo Unit Technique Human Action Dataset with Key Frame-Based CNN Action Recognition. *Sensors* 20, 17 (2020), 4871.
- [326] Iulia Lefter, Gertjan J Burghouts, and Leon J M Rothkrantz. 2014. An audio-visual dataset of human-human interactions in stressful situations. *Journal on Multimodal User Interfaces* 8, 1 (2014), 29–41. <https://doi.org/10.1007/s12193-014-0150-7>
- [327] Daniel Lightley, Moi Hoon Yap, Jessica Coulson, Yoann Barnouin, and Jamie S McPhee. 2015. Benchmarking human motion analysis using kinect one: An open source dataset. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1–7.
- [328] Séverin Lemaignan, Charlotte E R Edmunds, Emmanuel Senft, and Tony Belpaeme. 2018. The PInSoRo dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PLoS one* 13, 10 (2018), e0205999.
- [329] Samuel Lemercier, Mathieu Moreau, Mehdi Moussaïd, Guy Theraulaz, Stéphane Donikian, and Julien Pettré. 2011. Reconstructing Motion Capture Data for Human Crowd Study. In *Motion in Games*, Jan M Allbeck and Petros Faloutsos (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 365–376.
- [330] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. 2007. Crowds by example. *Computer Graphics Forum* 26, 3 (2007), 655–664. <https://doi.org/10.1111/j.1467-8659.2007.01089.x>
- [331] Heike Leutheuser, Dominik Schulhaus, and Bjoern M. Eskofier. 2013. Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. *PLoS ONE* 8, 10 (2013), 1–12. <https://doi.org/10.1371/journal.pone.0075196>
- [332] J J Lewis, S G Nikolov, A Loza, E Fernandez Canga, N Cvejic, J Li, A Cardinali, C N Canagarajah, D R Bull, T Riley, D Hickman, M I Smith, and Moira Smith. 2006. *The Eden Project Multi-Sensor Data Set*. Technical Report.
- [333] Roberto Leyva, Victor Sanchez, and Chang-Tsun Li. 2017. The LV dataset: A realistic surveillance video dataset for abnormal event detection. In *5th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 1–6.
- [334] Li Li, Ziyang Xie, and Xu Xu. 2020. MOPED25: A multimodal dataset of full-body pose and motion in occupational tasks. *Journal of Biomechanics* 113 (2020), 110086.
- [335] Michael H. Li, Tiago A. Mestre, Susan H. Fox, and Babak Taati. 2018. Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with deep learning pose estimation. *Journal of neuroengineering and rehabilitation* 15, 1 (2018), 1–13.
- [336] S Li, L Cui, C Zhu, B Li, N Zhao, and T Zhu. 2016. Emotion recognition using Kinect motion capture data of human gaits. *PeerJ* 4 (2016), e2364.
- [337] Shengchao Li, Lin Zhang, and Xiumin Diao. 2020. Deep-Learning-Based Human Intention Prediction Using RGB Images and Optical Flow. *Journal of Intelligent Robotic Systems* 97, 1 (2020), 95–107. <https://doi.org/10.1007/s10846-019-01049-3>
- [338] Shu-Jie Li, Hai-Sheng Zhu, Li-Ping Zheng, and Lin Li. 2020. A Perceptual-Based Noise-Agnostic 3D Skeleton Motion Data Refinement Network. *IEEE Access* 8 (2020), 52927–52940.
- [339] Wenhui Li, Yongkang Wong, An-An Liu, Yang Li, Yu-Ting Su, and Mohan Kankanalli. 2017. Multi-camera action dataset for cross-camera action recognition benchmarking. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 187–196.
- [340] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. 2010. Action recognition based on a bag of 3D points. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*. IEEE, 9–14. <https://doi.org/10.1109/CVPRW.2010.5543273>
- [341] Xing Li, Zhenjie Hou, Jiuzhen Liang, and Chen Chen. 2020. Human action recognition based on 3D body mask and depth spatial-temporal maps. *Multimedia Tools and Applications* 79 (2020), 35761–35778. <https://doi.org/10.1007/s11042-020-09593-z>
- [342] Y Li, J Cheng, W Feng, and D Tao. 2016. Feature fusion of triaxial acceleration signals and depth maps for human action recognition. *IEEE International Conference on Information and Automation (ICIA)* (2016). <https://doi.org/10.1109/ICInfA.2016.7832012>
- [343] Ya Li, Jianhua Tao, Linlin Chao, Wei Bao, and Yazhu Liu. 2017. CHEAVD: a Chinese natural emotional audio-visual database. *Journal of Ambient Intelligence and Humanized Computing* 8, 6 (2017), 913–924. <https://doi.org/10.1007/s12652-016-0406-z>
- [344] Ivan Lillo, Alvaro Soto, and Juan Carlos Niebles. 2014. Discriminative Hierarchical Modeling of Spatio-Temporally Composable Human Activities. In *IEEE conference on computer vision and pattern recognition*. 812–819.
- [345] Yan-Ching Lin, Min-Chun Hu, Wen-Huang Cheng, Yung-Huan Hsieh, and Hong-Ming Chen. 2012. Human action recognition and retrieval using sole depth information. In *MM'12*. 1053–1056. <https://doi.org/10.1145/2393347.2396381>
- [346] James Little and Jeffrey Boyd. 1998. Recognizing people by their gait: the shape of motion. *Videre: Journal of computer vision research* 1, 2 (1998), 1–32.
- [347] An An Liu, Wei Zhi Nie, Yu Ting Su, Li Ma, Tong Hao, and Zhao Xuan Yang. 2015. Coupled hidden conditional random fields for RGB-D human action recognition. *Signal Processing* 112 (2015), 74–82. <https://doi.org/10.1016/j.sigpro.2014.08.038>
- [348] An An Liu, Yu Ting Su, Ping Ping Jia, Zan Gao, Tong Hao, and Zhao Xuan Yang. 2015. Multiple/single-view human action recognition via part-induced multitask structural learning. *IEEE Transactions on Cybernetics* 45, 6 (2015), 1194–1208. <https://doi.org/10.1109/TCYB.2014.2347057>
- [349] Chunhui Liu, Yueyu Hu, Yanghao Li, Sijie Song, and Jiaying Liu. 2017. PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding. *arXiv* (2017). arXiv:1703.07475



- [350] Jingen Liu, Jiebo Luo, and Mubarak Shah. 2009. Recognizing realistic actions from videos in the Wild. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, Vol. 2009 IEEE. IEEE, 1996–2003. <https://doi.org/10.1109/CVPRW.2009.5206744>
- [351] Jun Liu, Amir Shahroudy, Mauricio Lisboa Perez, Gang Wang, Ling-Yu Duan, and Alex Kot Chichung. 2019. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8828, c (2019), 1–1. <https://doi.org/10.1109/tpami.2019.2916873>
- [352] Jiaying Liu, Sijie Song, Chunhui Liu, Yanghao Li, and Yueyu Hu. 2020. A Benchmark Dataset and Comparison Study for Multi-modal Human Action Analytics. *ACM Transactions on Multimedia Computing, Communications, and Applications* 16, 2 (2020), 41:1–41:24.
- [353] Phoebe Liu, Dylan F Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2018. Learning proactive behavior for interactive social robots. *Autonomous Robots* 42, 5 (2018), 1067–1085. <https://doi.org/10.1007/s10514-017-9671-8>
- [354] Shenglan Liu, Shuai Guo, Wei Wang, Hong Qiao, Yang Wang, and Wenbo Luo. 2020. Multi-view laplacian eigenmaps based on bag-of-neighbors for RGB-D human emotion recognition. *Information Sciences* 509 (2020), 243–256. <https://doi.org/10.1016/j.ins.2019.08.035>
- [355] Shenglan Liu, Xiang Liu, Gao Huang, Hong Qiao, Lianyu Hu, Dong Jiang, Aibin Zhang, Yang Liu, and Ge Guo. 2020. FSD-10: A fine-grained classification dataset for figure skating. *Neurocomputing* 413 (2020), 360–367.
- [356] Yuanliu Liu, Bo Peng, Peipei Shi, He Yan, Yong Zhou, Bing Han, Yi Zheng, Chao Lin, Jianbin Jiang, and Yin Fan. 2018. iqiwi-vid: A large dataset for multi-modal person identification. *arXiv* (2018).
- [357] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. 2007. A long-term evaluation of sensing modalities for activity recognition. In *International conference on Ubiquitous computing*. Springer, 483–500.
- [358] David López-Fernández, Francisco José Madrid-Cuevas, Ángel Carmona-Poyato, Manuel Jesús Marín-Jiménez, and Rafael Muñoz-Salinas. 2014. The AVA multi-view dataset for gait recognition. In *International workshop on activity monitoring by multiple distributed sensing*. Springer, 26–39.
- [359] Alexander Loui, Jiebo Luo, Shih Fu Chang, Dan Ellis, Wei Jiang, Lyndon Kennedy, Keansub Lee, and Akira Yanagawa. 2007. Kodak’s consumer video benchmark data set: Concept definition and annotation. In *Proceedings of the ACM International Multimedia Conference and Exhibition*. 245–254. <https://doi.org/10.1145/1290082.1290117>
- [360] Cewu Lu, Jianping Shi, and Jiaya Jia. 2013. Abnormal event detection at 150 FPS in MATLAB. In *IEEE International Conference on Computer Vision*. 2720–2727. <https://doi.org/10.1109/ICCV.2013.338>
- [361] Jiwen Lu, Gang Wang, and Pierre Moulin. 2014. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE Transactions on Information Forensics and Security* 9, 1 (2014), 51–61. <https://doi.org/10.1109/TIFS.2013.2291969>
- [362] Pengfei Lu and Matt Huenerfauth. 2014. Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation. *Computer Speech and Language* 28, 3 (2014), 812–831. <https://doi.org/10.1016/j.csl.2013.10.004>
- [363] Wei Lwun Lu, Jo Anne Ting, Kevin P. Murphy, and James J. Little. 2011. Identifying players in broadcast sports videos using conditional random fields. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 3249–3256. <https://doi.org/10.1109/CVPR.2011.5995562>
- [364] Manuel J Lucena, José Manuel Fuertes, and Nicolás Pérez De La Blanca. 2007. Human motion characterization using spatio-temporal features. In *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 72–79.
- [365] Jinna Lv, Wu Liu, Lili Zhou, Bin Wu, and Huadong Ma. 2018. Multi-stream Fusion Model for Social Relation Recognition from Videos. In *MultiMedia Modeling*, Klaus Schoeffmann, Thanarat H Chalidabhongse, Chong Wah Ngo, Supavadee Aramvith, Noel E O’Connor, Yo-Sung Ho, Moncef Gabbouj, and Ahmed Elgammal (Eds.). Springer International Publishing, Cham, 355–368.
- [366] Xin Ma, Haibo Wang, Bingxia Xue, and Yibin Li. 2015. Spatial-Temporal Feature Fusion for Human Fall Detection. In *Computer Vision*, Honbin Zha, Xilin Chen, Liang Wang, and Qiguang Miao (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 438–447.
- [367] Yingliang Ma, Helena M Paterson, and Frank E Pollick. 2006. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods* 38, 1 (2006), 134–141.
- [368] Zhigang Ma, Yi Yang, Feiping Nie, Nicu Sebe, Shuicheng Yan, and Alexander G Hauptmann. 2014. Harnessing Lab Knowledge for Real-World Action Recognition. *International Journal of Computer Vision* 109, 1 (2014), 60–73. <https://doi.org/10.1007/s11263-014-0717-5>
- [369] Lucia Maddalena and Alfredo Petrosino. 2015. Towards benchmarking scene background initialization. In *International conference on image analysis and processing*. Springer, 469–476.
- [370] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. 2010. Anomaly detection in crowded scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1975–1981. <https://doi.org/10.1109/CVPR.2010.5539872>
- [371] Maria Mahmood, Ahmad Jalal, and Kibum Kim. 2019. WHITE STAG model: wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors. *Multimedia Tools and Applications* 79, 11 (2019), 6919–6950. <https://doi.org/10.1007/s11042-019-08527-8>
- [372] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael Black. 2019. AMASS: Archive of motion capture as surface shapes. In *IEEE International Conference on Computer Vision*. 5441–5450. <https://doi.org/10.1109/ICCV.2019.00554>
- [373] Yasushi Makihara, Hidetoshi Mannami, and Yasushi Yagi. 2011. Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database. In *ACCV*, Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 440–451.
- [374] L Maman, E Ceccaldi, N Lehmann-Willenbrock, L Likforman-Sulem, M Chetouani, G Volpe, and G Varni. 2020. GAME-ON: A Multimodal Dataset for Cohesion and Group Analysis. *IEEE Access* 8 (2020). <https://doi.org/10.1109/ACCESS.2020.3005719>

- [375] Rok Mandeljc, Stanislav Kovacic, Matej Kristan, and Janez Pers̃. 2012. *MVL Lab5: Multi-modal Indoor Person Localization Dataset*. Technical Report.
- [376] Christian Mandery, Ömer Terlemeç, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. 2015. The KIT whole-body human motion database. In *17th International Conference on Advanced Robotics, ICAR 2015*. IEEE, 329–336. <https://doi.org/10.1109/ICAR.2015.7251476>
- [377] Valeria Manera, Ben Schouten, Cristina Becchio, Bruno G Bara, and Karl Verfaillie. 2010. Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. *Behavior research methods* 42, 1 (2010), 168–178.
- [378] Valeria Manera, Tabea von der Lühe, Leonhard Schilbach, Karl Verfaillie, and Cristina Becchio. 2016. Communicative interactions in point-light displays: Choosing among multiple response alternatives. *Behavior Research Methods* 48, 4 (2016), 1580–1590. <https://doi.org/10.3758/s13428-015-0669-x>
- [379] Al Mansur, Yasushi Makihara, and Yasushi Yagi. 2013. Inverse dynamics for action recognition. *IEEE Transactions on Cybernetics* 43, 4 (2013), 1226–1236. <https://doi.org/10.1109/TSMCB.2012.2226879>
- [380] Alice Mantoan, Claudio Pizzolato, Massimo Sartori, Zimi Sawacha, Claudio Cobelli, and Monica Reggiani. 2015. MOtoNMS: A MATLAB toolbox to process motion data for neuromusculoskeletal modeling and simulation. *Source code for biology and medicine* 10, 1 (2015), 12.
- [381] Marco Marcon, Marco Brando Mario Paracchini, and Stefano Tubaro. 2019. A framework for interpreting, modeling and recognizing human body gestures through 3D eigenpostures. *International Journal of Machine Learning and Cybernetics* 10, 5 (2019), 1205–1226. <https://doi.org/10.1007/s13042-018-0801-1>
- [382] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. 2016. Hand gesture recognition with jointly calibrated Leap Motion and depth sensor. *Multimedia Tools and Applications* 75, 22 (2016), 14991–15015. <https://doi.org/10.1007/s11042-015-2451-6>
- [383] Roberto Marroquin, Julien Dubois, and Christophe Nicolle. 2019. WiseNET: An indoor multi-camera multi-space dataset with contextual information and annotations for people detection and tracking. *Data in brief* 27 (2019), 104654.
- [384] Marcin Marszałek, Ivan Laptev, and Cordelia Schmid. 2009. Actions in context. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. IEEE, 2929–2936. <https://doi.org/10.1109/CVPRW.2009.5206557>
- [385] Manuel Martin, Alina Roitberg, Monica Haurilet, Matthias Horne, Simon Reiß, Michael Voit, and Rainer Stiefelhagen. 2019. Driveact: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles. In *IEEE international conference on computer vision*. 2801–2810.
- [386] Pierre-Etienne Martin, Jenny Benois-Pineau, Renaud Péteri, and Julien Morlier. 2020. Fine grained sport action recognition with Twin spatio-temporal convolutional neural networks. *Multimedia Tools and Applications* 79, 27 (2020), 20429–20447. <https://doi.org/10.1007/s11042-020-08917-3>
- [387] A. M. Martinez, R. B. Wilbur, R. Shay, and A. C. Kak. 2002. Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In *4th IEEE International Conference on Multimodal Interfaces*. 167–172. <https://doi.org/10.1109/ICMI.2002.1166987>
- [388] Angel Martínez-González, Michael Villamizar, Olivier Canévet, and Jean-Marc Odobez. 2019. Investigating Depth Domain Adaptation for Efficient Human Pose Estimation. In *ECCV Workshops*, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 346–363.
- [389] Lourdes Martínez-Villaseñor, Hiram Ponce, and Ricardo Abel Espinosa-Loera. 2018. Multimodal Database for Human Activity Recognition and Fall Detection. *Multidisciplinary Digital Publishing Institute Proceedings* 2, 19 (2018), 1237. <https://doi.org/10.3390/proceedings2191237>
- [390] Silke Matthes, Thomas Hanke, Jakob Storz, Eleni Efthimiou, Athanasia-Lida Dimou, Panagiotis Karioris, Annelies Braffort, Annick Choisier, Julia Pelhate, and Eva Safar. 2010. Elicitation Tasks and Materials designed for Dicta-Sign’s Multi-lingual Corpus. In *4th International workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*. 158–163.
- [391] J S Matthis and B R Fajen. 2013. Humans exploit the biomechanics of bipedal gait during visually guided walking over complex terrain. *Proc Biol Sci* 280, 1762 (2013), 20130700. <https://doi.org/10.1098/rspb.2013.0700>
- [392] J S Matthis, J L Yates, and M M Hayhoe. 2018. Gaze and the Control of Foot Placement When Walking in Natural Terrain. *Curr Biol* 28, 8 (2018), 1224–1233.e5. <https://doi.org/10.1016/j.cub.2018.03.008>
- [393] Pauline Maurice, Adrien Malaisé, Clélie Amiot, Nicolas Paris, Guy-Junior Richard, Olivier Rochel, and Serena Ivaldi. 2019. Human movement and ergonomics: An industry-oriented dataset for collaborative robotics. *The International Journal of Robotics Research* 38, 14 (2019), 1529–1537.
- [394] Iain McCowan, Daniel Gatica-Perez, Samy Bengio, Darren Moore, and Hervé Bourlard. 2004. Towards computer understanding of human interactions. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 56–75.
- [395] Gary McKeown, Michel F. Valstar, Roderick Cowie, and Maja Pantic. 2010. The semaine corpus of emotionally coloured character interactions. In *IEEE International Conference on Multimedia and Expo, ICME 2010*. IEEE, 1079–1084. <https://doi.org/10.1109/ICME.2010.5583006>
- [396] MCTdataset 2014. Multi-Camera Object Tracking (MCT) Challenge. <http://www.mct2014.com/Datasets.html>.
- [397] Alban Meffre, Christophe Collet, Nicolas Lachiche, and Pierre Gançarski. 2012. Real-Time Fall Detection Method Based on Hidden Markov Modelling. In *Image and Signal Processing, Abderrahim Elmoataz, Driss Mammass, Olivier Lezoray, Fathallah Nouboud, and Driss Aboutajdine (Eds.)*. Springer Berlin Heidelberg, Berlin, Heidelberg, 521–530.
- [398] Ramin Mehran, Alexis Oyama, and Mubarak Shah. 2009. Abnormal crowd behavior detection using social force model. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. IEEE, 935–942. <https://doi.org/10.1109/CVPRW.2009.5206641>
- [399] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. 2017. Monocular 3D human pose estimation in the wild using improved CNN supervision. In *International Conference on 3D Vision, 3DV*. 506–516. <https://doi.org/10.1109/3DV.2017.00064>

- [400] Dushyant Mehta, Oleksandr Sotnychenko, Franziska Mueller, Weipeng Xu, Srinath Sridhar, Gerard Pons-Moll, and Christian Theobalt. 2018. Single-shot multi-person 3D pose estimation from monocular RGB. In *International Conference on 3D Vision, 3DV 2018*. IEEE, 120–130. <https://doi.org/10.1109/3DV.2018.00024>
- [401] Michał Meina, Andrzej Janusz, Krzysztof Rykaczewski, Dominik Ślęzak, Bartosz Celmer, and Adam Krasuski. 2015. Tagging firefighter activities at the emergency scene: Summary of AIAA'15 data mining competition at Knowledge Pit. In *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 367–373.
- [402] Ross Messing, Chris Pal, and Henry Kautz. 2009. Activity recognition using the velocity histories of tracked keypoints. In *2009 IEEE 12th international conference on computer vision*. IEEE, 104–111.
- [403] Angeliki Metallinou, Chi-Chun Lee, Carlos Busso, Sharon Carnicke, and Shrikanth Narayanan. 2010. The USC Creative IT database: A Multimodal Database of Theatrical Improvisation. *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* May (2010), 1–4.
- [404] Pascal Mettes, Jan C. van Gemert, and Cees G.M. Snoek. 2016. Spot on: Action localization from point-supervised proposals. In *Lecture Notes in Computer Science*. 437–453. [https://doi.org/10.1007/978-3-319-46454-1\\_27](https://doi.org/10.1007/978-3-319-46454-1_27)
- [405] Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth, and Konrad Schindler. 2016. MOT16: A Benchmark for Multi-Object Tracking. (2016), 1–12.
- [406] Leandro Miranda, Thales Vieira, Dimas Martinez, Thomas Lewiner, Antonio W. Vieira, and Mario F.M. Campos. 2012. Real-time gesture recognition from depth data through key poses learning and decision forests. In *Brazilian Symposium of Computer Graphic and Image Processing*. IEEE, 268–275. <https://doi.org/10.1109/SIBGRAPI.2012.44>
- [407] MMDBdataset [n.d.].
- [408] G Moisan, M Descarreaux, and V Cantin. 2020. The influence of footwear on walking biomechanics in individuals with chronic ankle instability. *PLoS one* 15, 9 (2020), e0239621. <https://doi.org/10.1371/journal.pone.0239621>
- [409] Sarah Morrison-Smith, Aishat Aloba, Hangwei Lu, Brett Benda, Shaghayegh Esmaeili, Gianne Flores, Jesse Smith, Nikita Soni, Isaac Wang, and Rejin Joy. 2020. MMGatorAuth: A Novel Multimodal Dataset for Authentication Interactions in Gesture and Voice. In *International Conference on Multimodal Interaction*. 370–377.
- [410] Fatemeh Mortazavi and Ali Nadian-Ghomsheh. 2019. Continuous online exercise monitoring and assessment system with visual guidance feedback for stroke rehabilitation. *Multimedia Tools and Applications* 78, 22 (2019), 32055–32085. <https://doi.org/10.1007/s11042-019-08020-2>
- [411] Djamel Mostefa, Nicolas Moreau, Khalid Choukri, Gerasimos Potamianos, Stephen M. Chu, Amrith Tyagi, Josep R. Casas, Jordi Turmo, Luca Cristoforetti, Francesco Tobia, Aristodemos Pnevmatikakis, Vassilis Mylonakis, Fotios Talantzis, Susanne Burger, Rainer Stiefelhagen, Keni Bernardin, and Cedrick Rochet. 2007. The CHIL audiovisual corpus for lecture and meeting analysis inside smart rooms. *Language Resources and Evaluation* 41, 3–4 (2007), 389–407. <https://doi.org/10.1007/s10579-007-9054-4>
- [412] MRSC12dataset 2012. Kinect Gestures. <https://msropendata.com/datasets/27be4b01-e2cf-46ea-add6-ac77a1f61616>.
- [413] Rifat Muhammad Mueid, Chandrama Ahmed, and Md. Atiqur Rahman Ahad. 2016. Pedestrian activity classification using patterns of motion and histogram of oriented gradient. *Journal on Multimodal User Interfaces* 10, 4 (2016), 299–305. <https://doi.org/10.1007/s12193-015-0178-3>
- [414] Meinard Muller, Tido Roder, Michael Clausen, and year = 2007 title = Mocap Database HDM05. [n.d.]. . Technical Report.
- [415] Philipp M. Muller, Sikandar Amin, Prateek Verma, Mykhaylo Andriluka, and Andreas Bulling. 2015. Emotion recognition from embedded bodily expressions and speech during dyadic interactions. *International Conference on Affective Computing and Intelligent Interaction, ACII 2015* (2015), 663–669. <https://doi.org/10.1109/ACII.2015.7344640>
- [416] Matteo Munaro, Gioia Ballin, Stefano Michieletto, and Emanuele Menegatti. 2013. 3D flow estimation for human action recognition from colored point clouds. *Biologically Inspired Cognitive Architectures* 5 (2013), 42–51. <https://doi.org/10.1016/j.bica.2013.05.008>
- [417] Matteo Munaro, Alberto Basso, Andrea Fossati, Luc Van Gool, and Emanuele Menegatti. 2014. 3D reconstruction of freely moving persons for re-identification with a depth sensor. In *IEEE International Conference on Robotics and Automation*. IEEE, 4512–4519. <https://doi.org/10.1109/ICRA.2014.6907518>
- [418] Matteo Munaro and Emanuele Menegatti. 2014. Fast RGB-D people tracking for service robots. *Autonomous Robots* 37, 3 (2014), 227–242. <https://doi.org/10.1007/s10514-014-9385-0>
- [419] Yuji Nagashima, Keiko Watanabe, Daisuke Hara, Yasuo Horiuchi, Shinji Sako, and Akira Ichikawa. 2020. Constructing a Highly Accurate Japanese Sign Language Motion Database Including Dialogue. In *HCI International 2020 Posters*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 76–81.
- [420] Athira Nambiar, Matteo Taiana, Dario Figueira, Jacinto C Nascimento, and Alexandre Bernardino. 2014. A multi-camera video dataset for research on high-definition surveillance. *International Journal of Machine Intelligence and Sensory Signal Processing* 1, 3 (2014), 267–286.
- [421] Alfredo Nazabal and Antonio Artés. 2017. Active Sensing in Human Activity Recognition. In *Advances in Computational Intelligence*, Ignacio Rojas, Gonzalo Joya, and Andreu Catala (Eds.). Springer International Publishing, Cham, 157–166.
- [422] NearMissDRdataset [n.d.]. Drive Recorder Database. <https://www.jsae.or.jp/hiyari/>.
- [423] Farhood Negin, Fırat Özdemir, Ceyhan Burak Akgül, Kamer Ali Yüksel, and Aytül Erçil. 2013. A Decision Forest Based Feature Selection Framework for Action Recognition from RGB-Depth Cameras. In *Image Analysis and Recognition*, Mohamed Kamel and Aurélio Campilho (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 648–657.
- [424] Benjamin A Newman, Reuben M Aronson, Siddhartha S Srinivasa, Kris Kitani, and Henny Admoni. 2018. Harmonic: A multimodal dataset of assistive human-robot collaboration. *arXiv* (2018).

- [425] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin. 2007. ETISEO, performance evaluation for video surveillance systems. In *IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2007*. 476–481. <https://doi.org/10.1109/AVSS.2007.4425357>
- [426] Trong-Nguyen Nguyen and Jean Meunier. 2018. Walking gait dataset: point clouds, skeletons and silhouettes. *DIRO, University of Montreal, Tech. Rep 1379* (2018).
- [427] B Ni, Gang Wang, and P Moulin. 2011. RGBD-HuDaAct: A color-depth video database for human daily activity recognition. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 1147–1153. <https://doi.org/10.1109/ICCVW.2011.6130379>
- [428] Bingbing Ni, Shuicheng Yan, and Ashraf Kassim. 2009. Recognizing human group activities with localized causalities. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. IEEE, 1470–1477. <https://doi.org/10.1109/CVPRW.2009.5206853>
- [429] Juan Carlos Niebles, Chih-Wei Chen, and Li Fei-Fei. 2010. Modeling temporal structure of decomposable motion segments for activity classification. In *European conference on computer vision*. Springer, 392–405.
- [430] Friedrich Niemann, Christopher Reining, Fernando Moya Rueda, Nilah Ravi Nair, Janine Anika Steffens, Gernot A Fink, and Michael ten Hompel. 2020. LARa: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes. *Sensors* 20, 15 (2020), 4083.
- [431] Radoslaw Niewiadomski, Maurizio Mancini, Tobias Baur, Giovanna Varni, Harry Griffin, and Min S H Aung. 2013. MMLI: Multimodal Multiperson Corpus of Laughter in Interaction. In *Human Behavior Understanding*, Albert Ali Salah, Hayley Hung, Oya Aran, and Hatice Gunes (Eds.). Springer International Publishing, Cham, 184–195.
- [432] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, Hung Hsuan Huang, and Shogo Okada. 2014. Predicting influential statements in group discussions using speech and head motion information. In *International Conference on Multimodal Interaction*. 136–143. <https://doi.org/10.1145/2663204.2663248>
- [433] Shahriar Nirjon, Chris Greenwood, Carlos Torres, Stefanie Zhou, John A. Stankovic, Hee Jung Yoon, Ho Kyeong Ra, Can Basaran, Taejoon Park, and Sang H. Son. 2014. Kintense: A robust, accurate, real-time and evolving system for detecting aggressive actions from streaming 3D skeleton data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom) Kintense*. IEEE, 2–10. <https://doi.org/10.1145/2517351.2517396>
- [434] Jens Nirme, Magnus Haake, Agneta Gulz, and Marianne Gullberg. 2020. Motion capture-based animated characters for the study of speech–gesture integration. *Behavior Research Methods* 52, 3 (2020), 1339–1354. <https://doi.org/10.3758/s13428-019-01319-w>
- [435] Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. 2016. EmoReact: a multimodal approach and dataset for recognizing emotional responses in children. In *18th ACM International conference on multimodal interaction*. 137–144.
- [436] Stavros Ntalampiras, Dejan Arsić, Martin Hofmann, Maria Andersson, and Todor Ganchev. 2014. PROMETHEUS: heterogeneous sensor database in support of research on human behavioral patterns in unrestricted environments. *Signal, Image and Video Processing* 8, 7 (2014), 1211–1231. <https://doi.org/10.1007/s11760-012-0346-9>
- [437] R W Nuckols, K Z Takahashi, D J Farris, S Mizrachi, R Riemer, and G S Sawicki. 2020. Mechanics of walking and running up and downhill: A joint-level perspective to guide design of lower-limb exoskeletons. *PLoS one* 15, 8 (2020), e0231996. <https://doi.org/10.1371/journal.pone.0231996>
- [438] Joao Ferreira Nunes, Pedro Miguel Moreira, and João Manuel R S Tavares. 2019. GRIDDS-a gait recognition image and depth dataset. In *ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*. Springer, 343–352.
- [439] Catharine Oertel, Fred Cummins, Jens Edlund, Petra Wagner, and Nick Campbell. 2013. D64: a corpus of richly recorded conversational interaction. *Journal on Multimodal User Interfaces* 7, 1 (2013), 19–28. <https://doi.org/10.1007/s12193-012-0108-6>
- [440] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, Rene Vidal, and Ruzena Bajcsy. 2013. Berkeley MHAD: A comprehensive Multimodal Human Action Database. *Proceedings of IEEE Workshop on Applications of Computer Vision* (2013), 53–60. <https://doi.org/10.1109/WACV.2013.6474999>
- [441] Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia Chih Chen, Jong Taek Lee, Saurajit Mukherjee, J. K. Aggarwal, Hyungtae Lee, Larry Davis, Eran Swears, Xioyang Wang, Qiang Ji, Kishore Reddy, Mubarak Shah, Carl Vondrick, Hamed Pirsiavash, Deva Ramanan, Jenny Yuen, Antonio Torralba, Bi Song, Anesco Fong, Amit Roy-Chowdhury, and Mita Desai. 2011. A large-scale benchmark dataset for event recognition in surveillance video. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2* (2011), 3153–3160. <https://doi.org/10.1109/CVPR.2011.5995586>
- [442] Takuya Ohashi, Yosuke Ikegami, and Yoshihiko Nakamura. 2020. Synergetic reconstruction from 2D pose and 3D motion for wide-space multi-person video motion capture in the wild. *Image and Vision Computing* 104 (2020). <https://doi.org/10.1016/j.imavis.2020.104028>
- [443] Olukunle Ojetola, Elena Gaura, and James Brusey. 2015. Data set for fall events and daily activities from inertial sensors. In *6th ACM Multimedia Systems Conference, MMSys 2015*. 243–248. <https://doi.org/10.1145/2713168.2713198>
- [444] Kenji Okuma, Ali Taleghani, Nando De Freitas, James J Little, and David G Lowe. 2004. A boosted particle filter: Multitarget detection and tracking. In *European conference on computer vision*. Springer, 28–39.
- [445] T Olugbade, J Newbold, R Johnson, E Volta, P Alborno, R Niewiadomski, M Dillon, G Volpe, and N Bianchi-Berthouze. 2020. Automatic Detection of Reflective Thinking in Mathematical Problem Solving based on Unconstrained Bodily Exploration. *IEEE Transactions on Affective Computing* PP (2020). <https://doi.org/10.1109/TAFFC.2020.2978069>
- [446] Omar Oreifej and Zicheng Liu. 2013. HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 716–723. <https://doi.org/10.1109/CVPR.2013.98>
- [447] Gerardo Ortega and Asli Özyürek. 2020. Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods* 52, 1 (2020), 51–67. <https://doi.org/10.3758/s13428-019-01204-6>
- [448] Juan Diego Ortega, Neslihan Kose, Paola Cañas, Min-An Chao, Alexander Unnervik, Marcos Nieto, Oihana Otaegui, and Luis Salgado. 2020. DMD: A Large-Scale Multi-modal Driver Monitoring Dataset for Attention and Alertness Analysis. In *ECCV Workshops*, Adrien Bartoli and Andrea

- Fusiello (Eds.). Springer International Publishing, Cham, 387–405.
- [449] Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A Mollineda. 2018. Vision-based gait impairment analysis for aided diagnosis. *Medical Biological Engineering Computing* 56, 9 (2018), 1553–1564. <https://doi.org/10.1007/s11517-018-1795-2>
- [450] S Orter, D K Ravi, N B Singh, F Vogl, W R Taylor, and N König Ignasiak. 2019. A method to concatenate multiple short time series for evaluating dynamic behaviour during walking. *PLoS one* 14, 6 (2019), e0218594. <https://doi.org/10.1371/journal.pone.0218594>
- [451] Alice J O’Toole, Joshua Harms, Sarah L Snow, Dawn R Hurst, Matthew R Pappas, Janet H Ayyad, and Hervé Abdi. 2005. A video database of moving faces and people. *IEEE Transactions on pattern analysis and machine intelligence* 27, 5 (2005), 812–816.
- [452] Paul Over, Jon Fiscus, Greg Sanders, and Barbara Shaw. 2013. *TRECVID 2012 – An Overview of the Goals, Tasks, Data, Evaluation Mechanisms, and Metrics*. Technical Report.
- [453] Ahmet Turan Özdemir and Billur Barshan. 2014. Detecting falls with wearable sensors using machine learning techniques. *Sensors (Switzerland)* 14, 6 (2014), 10691–10708. <https://doi.org/10.3390/s140610691>
- [454] Oğulcan Özdemir, Ahmet Alp Kindiroğlu, Necati Cihan Camgöz, and Lale Akarun. 2020. Bosphorussign22k sign language recognition dataset. *arXiv* (2020).
- [455] G Ozkaya, H R Jung, I S Jeong, M R Choi, M Y Shin, X Lin, W S Heo, M S Kim, E Kim, and K K Lee. 2018. Three-dimensional motion capture data during repetitive overarm throwing practice. *Sci Data* 5 (2018), 180272. <https://doi.org/10.1038/sdata.2018.272>
- [456] Federico Pala, Riccardo Satta, Giorgio Fumera, and Fabio Roli. 2016. Multimodal Person Reidentification Using RGB-D Cameras. *IEEE Transactions on Circuits and Systems for Video Technology* 26, 4 (2016), 788–799. <https://doi.org/10.1109/TCSVT.2015.2424056>
- [457] Cristina Palmero, Albert Clapés, Chris Bahnsen, Andreas Møgelmo, Thomas B Moeslund, and Sergio Escalera. 2016. Multi-modal RGB–Depth–Thermal Human Body Segmentation. *International Journal of Computer Vision* 118, 2 (2016), 217–239. <https://doi.org/10.1007/s11263-016-0901-x>
- [458] E Papageorgiou, C Simon-Martinez, G Molenaers, E Ortibus, A Van Campenhout, and K Desloovere. 2019. Are spasticity, weakness, selectivity, and passive range of motion related to gait deviations in children with spastic cerebral palsy? A statistical parametric mapping study. *PLoS one* 14, 10 (2019), e0223363. <https://doi.org/10.1371/journal.pone.0223363>
- [459] Stylianos Paraschiakos, Ricardo Cachucho, Matthijs Moed, Diana van Heemst, Simon Mooijaart, Eline P Slagboom, Arno Knobbe, and Marian Beekman. 2020. Activity recognition using wearable sensors for tracking the elderly. *User Modeling and User-Adapted Interaction* 30, 3 (2020), 567–605. <https://doi.org/10.1007/s11257-020-09268-2>
- [460] Cheul Young Park, Narae Cha, Soowon Kang, Auk Kim, Ahsan Habib Khandoker, Leontios Hadjileontiadis, Alice Oh, Yong Jeong, and Uichin Lee. 2020. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data* 7, 1 (2020).
- [461] P Parmar and B T Morris. 2019. What and How Well You Performed? A Multitask Learning Approach to Action Quality Assessment. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). <https://doi.org/10.1109/CVPR.2019.00039>
- [462] Luis Patino, Tom Cane, Alain Vallee, and James Ferryman. 2016. PETS 2016: Dataset and Challenge. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.
- [463] Adrian Penate-Sanchez, David Freire-Obregón, Adrián Lorenzo-Melián, Javier Lorenzo-Navarro, and Modesto Castrillón-Santana. 2020. TGC20ReId: A dataset for sport event re-identification in the wild. *Pattern Recognition Letters* 138 (2020), 355–361.
- [464] Olga Perepelkina, Evdokia Kazimirova, and Maria Konstantinova. 2018. RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing. In *Speech and Computer*, Alexey Karpov, Oliver Jokisch, and Rodmonga Potapova (Eds.). Springer International Publishing, Cham, 501–510.
- [465] Asanka G Perera, Yee Wei Law, and Javaan Chahl. 2018. UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition. In *ECCV*, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 117–128.
- [466] Asanka G Perera, Yee Wei Law, and Javaan Chahl. 2019. Drone-Action: An Outdoor Recorded Drone Video Dataset for Action Recognition. *Drones* 3, 4 (2019), 82.
- [467] Asanka G Perera, Yee Wei Law, Titilayo T Ogunwa, and Javaan Chahl. 2020. A Multiviewpoint Outdoor Dataset for Human Action Recognition. *IEEE Transactions on Human-Machine Systems* (2020).
- [468] Jessica Permatasari, Tee Connie, and Thian Song Ong. 2020. Inertial sensor fusion for gait recognition with symmetric positive definite Gaussian kernels analysis. *Multimedia Tools and Applications* 79, 43 (2020), 32665–32692. <https://doi.org/10.1007/s11042-020-09438-9>
- [469] Jessica Permatasari, Tee Connie, and Ong Thian Song. 2020. The MMUISD Gait Database and Performance Evaluation Compared to Public Inertial Sensor Gait Databases. In *Computational Science and Technology*, Rayner Alfred, Yuto Lim, Haviluddin Haviluddin, and Chin Kim On (Eds.). Springer Singapore, Singapore, 189–198.
- [470] PETS2006dataset 2006. PETS 2006 Benchmark Data.
- [471] PETS2007dataset 2007. PETS 2007 Benchmark Data.
- [472] Svein Arne Pettersen, Dag Johansen, Håvard Johansen, Vegard Berg-Johansen, Vamsidhar Reddy Gaddam, Asgeir Mortensen, Ragnar Langseth, Carsten Griwodz, Håkon Kvale Stensland, and Pål Halvorsen. 2014. Soccer video and player position dataset. In *5th ACM Multimedia Systems Conference, MMSys 2014*. 18–23. <https://doi.org/10.1145/2557642.2563677>
- [473] Huy Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. 2019. A Deep Learning Approach for Real-Time 3D Human Action Recognition from Skeletal Data. In *Image Analysis and Recognition*, Fakhri Karray, Aurélio Campilho, and Alfred Yu (Eds.). Springer International Publishing, Cham, 18–32.

- [474] Alessandro Pieropan, Giampiero Salvi, Karl Pauwels, and Hedvig Kjellström. 2014. A dataset of human manipulation actions. In *ICRA Workshop Auton. Grasping Manipulation*.
- [475] Stefano Pinardi and Roberto Bisiani. 2010. Movement Recognition with Intelligent Multisensor Analysis, a Lexical Approach. In *Intelligent Environments (Workshops)*. 170–177. <https://doi.org/10.3233/978-1-60750-638-6-170>
- [476] Silvia L Pinteá, Jian Zheng, Xilin Li, Paulina J M Bank, Jacobus J van Hilten, and Jan C van Gemert. 2018. Hand-Tremor Frequency Estimation in Videos. In *ECCV Workshops*, Laura Leal-Taixé and Stefan Roth (Eds.). Springer International Publishing, Cham, 213–228.
- [477] Hamed Pirsiavash, Carl Vondrick, and Antonio Torralba. 2014. Assessing the Quality of Actions. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 556–571.
- [478] Rainer Planinc and Martin Kampel. 2013. Robust Fall Detection by Combining 3D Data and Fuzzy Logic. In *ACCV*, Jong-Il Park and Junmo Kim (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 121–132.
- [479] Rainer Planinc and Martin Kampel. 2015. Human Centered Scene Understanding Based on 3D Long-Term Tracking Data. In *Pattern Recognition and Image Analysis*, Roberto Paredes, Jaime S Cardoso, and Xosé M Pardo (Eds.). Springer International Publishing, Cham, 93–100.
- [480] Rainer Polak, Simone Tarsitani, and Martin Clayton. 2018. IEMP Malian Jembe. <https://doi.org/10.17605/OSF.IO/M652X>.
- [481] Gerard Pons-Moll, Andreas Baak, Thomas Helten, Meinard Müller, Hans Peter Seidel, and Bodo Rosenhahn. 2010. Multisensor-fusion for 3D full-body human motion capture. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 663–670. <https://doi.org/10.1109/CVPR.2010.5540153>
- [482] Gerard Pons-Moll, Laura Leal-Taixé, Juergen Gall, and Bodo Rosenhahn. 2012. Data-Driven Manifolds for Outdoor Motion Capture, Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, Laura Leal-Taixé, and Bodo Rosenhahn (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 305–328.
- [483] Alberto Pozo, Jesús García, and Miguel A Patricio. 2012. INEF12Basketball Dataset and the Group Behavior Recognition Issue. In *Distributed Computing and Artificial Intelligence*, Sigeru Omatu, Juan F De Paz Santana, Sara Rodríguez González, Jose M Molina, Ana M Bernardos, and Juan M Corchado Rodríguez (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 151–160.
- [484] Gaurav N Pradhan, Navzer Engineer, Mihai Nadin, and B Prabhakaran. [n.d.]. *Universal Access in the Information Society* ([n.d.]). <https://doi.org/10.1007/s10209-010-0205-9>
- [485] Majid Ali Khan Quaid and Ahmad Jalal. 2020. Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. *Multimedia Tools and Applications* 79, 9 (2020), 6061–6083. <https://doi.org/10.1007/s11042-019-08463-7>
- [486] C A Rábago, J Aldridge Whitehead, and J M Wilken. 2016. Evaluation of a Powered Ankle-Foot Prosthesis during Slope Ascent Gait. *PLoS one* 11, 12 (2016), e0166815. <https://doi.org/10.1371/journal.pone.0166815>
- [487] C A Rábago, J B Dingwell, and J M Wilken. 2015. Reliability and Minimum Detectable Change of Temporal-Spatial, Kinematic, and Dynamic Stability Measures during Perturbed Gait. *PLoS one* 10, 11 (2015), e0142083. <https://doi.org/10.1371/journal.pone.0142083>
- [488] Hamidreza Rabiee, Javad Haddadnia, and Hossein Mousavi. 2016. Crowd behavior representation: an attribute-based approach. *SpringerPlus* 5, 1 (2016), 1179. <https://doi.org/10.1186/s40064-016-2786-0>
- [489] Hossein Rahmani, Arif Mahmood, Du Huynh, and Ajmal Mian. 2016. Histogram of Oriented Principal Components for Cross-View Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 12 (2016), 2430–2443. <https://doi.org/10.1109/TPAMI.2016.2533389>
- [490] Hossein Rahmani, Arif Mahmood, Du Q Huynh, and Ajmal Mian. 2014. HOPC: Histogram of Oriented Principal Components of 3D Pointclouds for Action Recognition. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 742–757.
- [491] S. Ramagiri, R. Kavi, and V. Kulathumani. 2011. Real-time multi-view human action recognition using a wireless camera network. In *5th ACM/IEEE International Conference on Distributed Smart Cameras, ICDS-C 2011*. IEEE, 1–6. <https://doi.org/10.1109/ICDS-C.2011.6042901>
- [492] Rahul Raman, Pankaj K Sa, Banshidhar Majhi, and Sambit Bakshi. 2017. Acquisition and corpus description of NITR conscious walk dataset. *ACM SIGBioinformatics Rec.* 7, 1 (2017), Article 1. <https://doi.org/10.1145/3056351.3056352>
- [493] Vignesh Ramanathan, Jonathan Huang, Sami Abu-El-Hajja, Alexander Gorban, Kevin Murphy, and Li Fei-Fei. 2016. Detecting events and key actors in multi-person videos. (2016). <https://doi.org/10.1109/CVPR.2016.332>
- [494] Hiranmayi Ranganathan, Shayok Chakraborty, and Sethuraman Panchanathan. 2016. Multimodal emotion recognition using deep learning architectures. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. IEEE. <https://doi.org/10.1109/WACV.2016.7477679>
- [495] Cen Rao, Alper Yilmaz, and Mubarak Shah. 2002. View-Invariant Representation and Recognition of Actions. *International Journal of Computer Vision* 50, 2 (2002), 203–226. <https://doi.org/10.1023/A:1020350100748>
- [496] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K Tsotsos. 2019. PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *IEEE International Conference on Computer Vision*. 6262–6271.
- [497] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior. In *IEEE International Conference on Computer Vision Workshops*. 206–213.
- [498] Kishore K Reddy and Mubarak Shah. 2013. Recognizing 50 human action categories of web videos. *Machine Vision and Applications* 24, 5 (2013), 971–981. <https://doi.org/10.1007/s00138-012-0450-4>
- [499] Attila Reiss and Didier Stricker. 2012. Introducing a new benchmarked dataset for activity monitoring. In *16th international symposium on wearable computers*. 108–109.
- [500] Boris Reuderink, Mannes Poel, Khiet Truong, Ronald Poppe, and Maja Pantic. 2008. Decision-level fusion for audio-visual laughter detection. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 137–148.

- [501] Michael Riegler, Duc-Tien Dang-Nguyen, Bård Winther, Carsten Griwodz, Konstantin Pogorelov, and Pål Halvorsen. 2016. Heimdallr: a dataset for sport analysis. In *7th International Conference on Multimedia Systems*. 1–6.
- [502] Gerhard Rigoll, Andreas Kosmala, and Stefan Eickeler. 1997. High performance real-time gesture recognition using hidden markov models. In *International Gesture Workshop*. Springer, 69–80.
- [503] Omar Rihawi, Djamel Merad, and Jean-Luc Damoiseaux. 2017. 3D-AD: 3D-autism dataset for repetitive behaviours with kinect sensor. In *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–6.
- [504] Isidoros Rodomagoulakis, Nikolaos Kardaris, Vassilis Pitsikalis, Antonis Arvanitakis, and Petros Maragos. 2016. A multimedia gesture dataset for human robot communication: Acquisition, tools and recognition results. In *IEEE International Conference on Image Processing (ICIP)*. IEEE, 3066–3070.
- [505] D Roggen, A Calatroni, M Rossi, T Holleczeck, K Förster, G Tröster, P Lukowicz, D Bannach, G Pirkel, A Ferscha, J Doppler, C Holzmann, M Kurz, G Holl, R Chavarriaga, H Sagha, H Bayati, M Creatura, and J d. R Millàn. 2010. Collecting complex activity datasets in highly rich networked sensor environments. *Seventh International Conference on Networked Sensing Systems (INSS)* (2010). <https://doi.org/10.1109/INSS.2010.5573462>
- [506] Anna Rohrbach, Marcus Rohrbach, Niket Tandon, and Bernt Schiele. 2015. A dataset for Movie Description. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 07-12-June. 3202–3212. <https://doi.org/10.1109/CVPR.2015.7298940>
- [507] Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. 2012. A database for fine grained activity detection of cooking activities. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2012), 1194–1201. <https://doi.org/10.1109/CVPR.2012.6247801>
- [508] Marcus Rohrbach, Anna Rohrbach, Michaela Regneri, Sikandar Amin, Mykhaylo Andriluka, Manfred Pinkal, and Bernt Schiele. 2016. Recognizing Fine-Grained and Composite Activities Using Hand-Centric Features and Script Data. *International Journal of Computer Vision* 119, 3 (2016), 346–373. <https://doi.org/10.1007/s11263-015-0851-8>
- [509] Laura Romeo, Roberto Marani, Antonio Pettiti, Annalisa Milella, Tiziana D’Orazio, and Grazia Cicirelli. 2020. Image-Based Mobility Assessment in Elderly People from Low-Cost Systems of Cameras: A Skeletal Dataset for Experimental Evaluations. In *Ad-Hoc, Mobile, and Wireless Networks*, Luigi Alfredo Grieco, Gennaro Boggia, Giuseppe Piro, Yaser Jararweh, and Claudia Campolo (Eds.). Springer International Publishing, Cham, 125–130.
- [510] Cristina Romero-González, Álvaro Villena, Daniel González-Medina, Jesus Martínez-Gómez, Luis Rodríguez-Ruiz, and Ismael García-Varea. 2017. InLiDa: A 3D Lidar Dataset for People Detection and Tracking in Indoor Environments. In *VISIGRAPP*. 484–491.
- [511] Franco Ronchetti, Facundo Quiroga, César Armando Estrebo, Laura Cristina Lanzarini, and Alejandro Rosete. 2016. LSA64: an Argentinian sign language dataset. In *XXII Congreso Argentino de Ciencias de la Computación (CACIC)*.
- [512] Paolo Rota, Nicola Conci, and Nicu Sebe. 2012. Real Time Detection of Social Interactions in Surveillance Videos. In *ECCV Workshops and Demonstrations*, Andrea Fusiello, Vittorio Murino, and Rita Cucchiara (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 111–120.
- [513] P Rota, N Conci, N Sebe, and J M Rehg. 2015. Real-life violent social interaction detection. In *IEEE International Conference on Image Processing (ICIP)*. <https://doi.org/10.1109/ICIP.2015.7351446>
- [514] E Roth, M Möncks, T Bohné, and L Pumplun. 2020. Context-Aware Cyber-Physical Assistance Systems in Industrial Systems: A Human Activity Recognition Approach. In *IEEE International Conference on Human-Machine Systems (ICHMS)*. <https://doi.org/10.1109/ICHMS49158.2020.9209488>
- [515] Anindya Roy, Camille Guinaudeau, Hervé Bredin, and Claude Barras. 2014. TVD: a reproducible and multiply aligned TV series dataset. In *LREC*.
- [516] Don J Rude, Stephen Adams, and Peter A Beling. 2015. A benchmark dataset for depth sensor based activity recognition in a manufacturing process. *IFAC-PapersOnLine* 48, 3 (2015), 668–674.
- [517] Dmitry Rudoy and Lihí Zelnik-Manor. 2012. Viewpoint Selection for Human Actions. *International Journal of Computer Vision* 97, 3 (2012), 243–254. <https://doi.org/10.1007/s11263-011-0484-5>
- [518] Simon Ruffieux, Denis Lalanne, and Elena Mugellini. 2013. ChAirGest: A challenge for multimodal mid-air gesture recognition for close HCI. In *ACM International Conference on Multimodal Interaction*. 483–488. <https://doi.org/10.1145/2522848.2532590>
- [519] Michael S Ryoo and Jake K Aggarwal. 2009. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *IEEE 12th international conference on computer vision*. IEEE, 1593–1600.
- [520] M S Ryoo and Larry Matthies. 2016. First-Person Activity Recognition: Feature, Temporal Structure, and Prediction. *International Journal of Computer Vision* 119, 3 (2016), 307–328. <https://doi.org/10.1007/s11263-015-0847-4>
- [521] Andrea Sabo, Sina Mehdizadeh, Kimberley-Dale Ng, Andrea Iaboni, and Babak Taati. 2020. Assessment of Parkinsonian gait in older adults with dementia via human pose tracking in video data. *Journal of NeuroEngineering and Rehabilitation* 17, 1 (2020), 97. <https://doi.org/10.1186/s12984-020-00728-9>
- [522] Seyed Morteza Safdarnejad, Xiaoming Liu, Lalita Udpa, Brooks Andrus, John Wood, and Dean Craven. 2015. Sports videos in the wild (SVW): A video dataset for sports analysis. In *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–7.
- [523] SaGAdataset 2010. Bavarian Archive for Speech Signals: Bielefeld Speech and Gesture Alignment Corpus - SaGA. <https://www.phonetik.uni-muenchen.de/Bas/BasSaGAeng.html>.
- [524] Mukesh Saini, Seshadri Padmanabha Venkatagiri, Wei Tsang Ooi, and Mun Choon Chan. 2013. The jiku mobile video dataset. In *ACM multimedia systems conference*. 108–113.
- [525] Rajkumar Saini, Pradeep Kumar, Barjinder Kaur, Partha Pratim Roy, Debi Prosad Dogra, and K C Santosh. 2019. Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare. *International Journal of Machine Learning and Cybernetics* 10, 9 (2019), 2529–2540. <https://doi.org/10.1007/s13042-018-0887-5>

- [526] David A Salter, Amir Tamrakar, Behjat Siddiquie, Mohamed R Amer, Ajay Divakaran, Brian Lande, and Darius Mehri. 2015. The tower game dataset: A multimodal dataset for analyzing social interaction predicates. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 656–662.
- [527] Daniel Sánchez, Miguel Ángel Bautista, and Sergio Escalera. 2015. HuPBA8k+: Dataset and ECOC-Graph-Cut based segmentation of human limbs. *Neurocomputing* 150, Part A (2015), 173–188. <https://doi.org/10.1016/j.neucom.2014.07.069>
- [528] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2011. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia* 14, 3 PART 2 (2011), 816–832. <https://doi.org/10.1109/TMM.2011.2181941>
- [529] Martin Sandau, Rikke V Heimbürger, Karl E Jensen, Thomas B Moeslund, Henrik Aanæs, Tine Alkjær, and Erik B Simonsen. 2016. Reliable gait recognition using 3D reconstructions and random forests—An anthropometric approach. *Journal of forensic sciences* 61, 3 (2016), 637–648.
- [530] Peter Sandilands, Myung Geol Choi, and Taku Komura. 2012. Capturing Close Interactions with Objects Using a Magnetic Motion Capture System and a RGBD Sensor. In *Motion in Games*, Marcelo Kallmann and Kostas Bekris (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 220–231.
- [531] Maria J Santofimia, Jesus Martinez-del Rincon, and Jean-Christophe Nebel. 2012. Common-Sense Knowledge for a Computer Vision System for Human Action Recognition. In *Ambient Assisted Living and Home Care*, José Bravo, Ramón Hervás, and Marcela Rodríguez (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 159–166.
- [532] José Antonio Santoyo-Ramón, Eduardo Casilari, and José Manue Cano-García. 2018. Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection with supervised learning. *Sensors (Switzerland)* 18, 4 (2018). <https://doi.org/10.3390/s18041155>
- [533] Tomasz Sapiński, Dorota Kamińska, Adam Pelikant, Cagri Ozcinar, Egils Avots, and Gholamreza Anbarjafari. 2019. Multimodal Database of Emotional Speech, Video and Gestures. In *Pattern Recognition and Information Forensics*, Zhaoxiang Zhang, David Suter, Yingli Tian, Alexandra Branzan Albu, Nicolas Sidère, and Hugo Jair Escalante (Eds.). Springer International Publishing, Cham, 153–163.
- [534] Benjamin Sapp, David Weiss, and Ben Taskar. 2011. Parsing human motion with stretchable models. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1281–1288. <https://doi.org/10.1109/CVPR.2011.5995607>
- [535] Faegheh Sardari, Adeline Paiement, and Majid Mirmehdi. 2019. View-Invariant Pose Analysis for Human Movement Assessment from RGB Data. In *ICIAP*, Elisa Ricci, Samuel Rota Bulò, Cees Snoek, Oswald Lanz, Stefano Meseleodi, and Nicu Sebe (Eds.). Springer International Publishing, Cham, 237–248.
- [536] M M Sardsehmukh, M T Kolte, P N Chatur, and D S Chaudhari. 2014. 3-D dataset for Human Activity Recognition in video surveillance. In *IEEE Global Conference on Wireless Computing Networking (GCWCN)*. IEEE, 75–78.
- [537] Sudeep Sarkar, P Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W Bowyer. 2005. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence* 27, 2 (2005), 162–177.
- [538] Manolis Savva, Angel X. Chang, Pat Hanrahan, Matthew Fisher, and Matthias Nießner. 2016. PiGraphs: Learning interaction snapshots from observations. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–12. <https://doi.org/10.1145/2992138.2992147>
- [539] Nikolaos Savva and Nadia Bianchi-Berthouze. 2012. Automatic Recognition of Affective Body Movement in a Video Game Scenario. In *Intelligent Technologies for Interactive Entertainment*, Antonio Camurri and Cristina Costa (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 149–159.
- [540] Monika Schach and Alexander Gepperth. 2020. On Multi-modal Fusion for Freehand Gesture Recognition. In *ICANN*, Igor Farkaš, Paolo Masulli, and Stefan Wermter (Eds.). Springer International Publishing, Cham, 862–873.
- [541] Laura Schiphorst, Metehan Doyran, Sabine Molenaar, Albert Ali Salah, and Sjaak Brinkkemper. 2020. Video2Report: A Video Database for Automatic Reporting of Medical Consultancy Sessions. In *15th IEEE International Conference on Automatic Face and Gesture Recognition, FG*, 552–556. <https://doi.org/10.1109/FG47880.2020.00020>
- [542] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. 2018. Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. In *International Conference on Multimodal Interaction*. 400–408. <https://doi.org/10.1145/3242969.3242985>
- [543] Céline Schreiber and Florent Moissenet. 2019. A multimodal dataset of human gait at different walking speeds established on injury-free adult participants. *Scientific data* 6, 1 (2019), 1–7.
- [544] Christian Schüldt, Ivan Laotev, and Barbara Caputo. 2004. Recognizing human actions : A local SVM approach Recognizing Human Actions. In *17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 3. 3–7. <https://doi.org/10.1109/ICPR.2004.1334462>
- [545] Arne Schumann and Eduardo Monari. 2014. A soft-biometrics dataset for person tracking and re-identification. In *11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 193–198.
- [546] Jesse Scott, Bharadwaj Ravichandran, Christopher Funk, Robert T Collins, and Yanxi Liu. 2020. From Image to Stability: Learning Dynamics from Human Pose. In *ECCV*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 536–554.
- [547] N Seethapathi and M Srinivasan. 2019. Step-to-step variations in human running reveal how humans run without falling. *Elife* 8 (2019). <https://doi.org/10.7554/eLife.38371>
- [548] Lorenzo Seidenari, Vincenzo Varano, Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. 2013. Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 479–485. <https://doi.org/10.1109/CVPRW.2013.77>
- [549] Mariano Serrao, Giorgia Chini, Matteo Bergantino, Diego Sarnari, Carlo Casali, Carmela Conte, Alberto Ranavolo, Christian Marcotulli, Martina Rinaldi, and Gianluca Coppola. 2018. Dataset on gait patterns in degenerative neurological diseases. *Data in brief* 16 (2018), 806–816.



- [550] Ricky J Sethi. 2015. Towards defining groups and crowds in video using the atomic group actions dataset. In *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2925–2929.
- [551] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. 2016. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. In *IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1109/CVPR.2016.115>
- [552] Noor Shaker, Stylianos Asteriadis, Georgios N Yannakakis, and Kostas Karpouzis. 2011. A Game-Based Corpus for Analysing the Interplay between Game Context and Player Experience. In *Affective Computing and Intelligent Interaction*, Sidney D’Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 547–556.
- [553] Dian Shao, Yue Zhao, Bo Dai, and Dahua Lin. 2020. Finegym: A hierarchical video dataset for fine-grained action understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2616–2625.
- [554] Jing Shao, Chen Change Loy, and Xiaogang Wang. 2014. Scene-independent group profiling in crowd. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2014.285>
- [555] Guangyao Shen, Xin Wang, Xuguang Duan, Hongzhi Li, and Wenwu Zhu. 2020. MEmoR: A Dataset for Multimodal Emotion Reasoning in Videos. In *28th ACM International Conference on Multimedia*. Association for Computing Machinery, Seattle, WA, USA, 493–502. <https://doi.org/10.1145/3394171.3413909>
- [556] Haoquan Shen, Shou-I Yu, Yi Yang, Deyu Meng, and Alexander Hauptmann. 2014. Unsupervised Video Adaptation for Parsing Human Motion. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 347–360.
- [557] Bowen Shi, Aurora Martinez, Del Rio, Jonathan Keane, Jonathan Michaux, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2018. American Sign Language Fingerspelling Recognition in the Wild. In *IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 145–152.
- [558] Bowen Shi, Aurora Martinez Del Rio, Jonathan Keane, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2019. Fingerspelling recognition in the wild with iterative visual attention. In *IEEE International Conference on Computer Vision*. 5399–5408. <https://doi.org/10.1109/ICCV.2019.00550>
- [559] Minho Shim, Young Hwi Kim, Kyungmin Kim, and Seon Joo Kim. 2018. Teaching Machines to Understand Baseball Games: Large-Scale Baseball Video Database for Multiple Video Understanding Tasks. In *ECCV*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 420–437.
- [560] Atsushi Shimada, Kazuaki Kondo, Daisuke Deguchi, Géraldine Morin, and Helman Stern. 2013. Kitchen Scene Context Based Gesture Recognition: A Contest in ICPR2012. In *Advances in Depth Image Analysis and Applications*, Xiaoyi Jiang, Olga Regina Pereira Bellon, Dmitry Goldgof, and Takeshi Oishi (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 168–185.
- [561] Thomas F Shipley and Jonathan S Brumberg. 2005. Markerless motion-capture for point-light displays. Retrieved April 4 (2005), 2005.
- [562] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul J.M. Havinga. 2014. Fusion of smartphone motion sensors for physical activity recognition. *Sensors* 14, 6 (2014), 10146–10176. <https://doi.org/10.3390/s140610146>
- [563] Muhammad Shoaib, Ralf Dragon, and Joern Ostermann. 2011. Context-aware visual analysis of elderly activity in a cluttered home environment. *EURASIP Journal on Advances in Signal Processing* 2011, 1 (2011), 129. <https://doi.org/10.1186/1687-6180-2011-129>
- [564] Jamie D Shutler, Michael G Grant, Mark S Nixon, and John N Carter. 2004. On a large sequence-based human gait database. In *Applications and science in soft computing*. Springer, 339–346.
- [565] Leonid Sigal, Alexandru O Balan, and Michael J Black. 2010. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision* 87, 1-2 (2010), 4.
- [566] Gunnar A Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, and Abhinav Gupta. 2016. Hollywood in Homes: Crowdsourcing Data Collection for Activity Understanding. In *ECCV*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 510–526.
- [567] Damien Simonnet, Michal Lewandowski, Sergio A Velastin, James Orwell, and Esin Turkbeyler. 2012. Re-identification of Pedestrians in Crowds Using Dynamic Time Warping. In *ECCV Workshops and Demonstrations*, Andrea Fusiello, Vittorio Murino, and Rita Cucchiara (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 423–432.
- [568] Bharat Singh, Tim K. Marks, Michael Jones, Oncel Tuzel, and Ming Shao. 2016. A Multi-stream Bi-directional Recurrent Neural Network for Fine-Grained Action Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1961–1970. <https://doi.org/10.1109/CVPR.2016.216>
- [569] Jasvinder Pal Singh, Sanjeev Jain, Sakshi Arora, and Uday Pratap Singh. 2020. Reconstruction of occluded ROI in multi-person gait based on numerical methods. *Multimedia Systems* 26, 3 (2020), 249–266. <https://doi.org/10.1007/s00530-019-00641-9>
- [570] Sanchit Singh, Sergio A. Velastin, and Hossein Ragheb. 2010. MuHAVI: A multicamera human action video dataset for the evaluation of action recognition methods. In *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*. IEEE, 48–55. <https://doi.org/10.1109/AVSS.2010.63>
- [571] Joyeeta Singha, Amarjit Roy, and Rabul Hussain Laskar. 2018. Dynamic hand gesture recognition using vision-based approach for human-computer interaction. *Neural Computing and Applications* 29, 4 (2018), 1129–1141. <https://doi.org/10.1007/s00521-016-2525-z>
- [572] Mohamed Mostafa Soliman, Mohamed Hussein Kamal, Mina Abd El-Massih Nashed, Youssef Mohamed Mostafa, Bassel Safwat Chawky, and Dina Khattab. 2019. Violence Recognition from Videos using Deep Learning Techniques. In *IEEE 9th International Conference on Intelligent Computing and Information Systems, ICICIS*. IEEE, 80–85. <https://doi.org/10.1109/ICICIS46948.2019.9014714>
- [573] Yale Song, David Demirdjian, and Randall Davis. 2011. Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*. 500–506. <https://doi.org/10.1109/FG.2011.5771448>

- [574] Yan Song, Jinhui Tang, Fan Liu, and Shuicheng Yan. 2014. Body surface context: A new robust feature for action recognition from depth videos. *IEEE Transactions on Circuits and Systems for Video Technology* 24, 6 (2014), 952–965. <https://doi.org/10.1109/TCSVT.2014.2302558>
- [575] K Soomro, A R Zamir, and M Shah. 2012. UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild. *arXiv* (2012).
- [576] S I Soraya, S Chuang, Y Tseng, T İk, and Y Ching. 2019. A Comprehensive Multisensor Dataset Employing RGBD Camera, Inertial Sensor and Web Camera. In *20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. <https://doi.org/10.23919/APNOMS.2019.8892906>
- [577] Luciano Spinello and Kai O. Arras. 2011. People detection in RGB-D data. In *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 3838–3843. <https://doi.org/10.1109/IROS.2011.6048835>
- [578] Jonathan Starck and Adrian Hilton. 2007. Surface Capture for Performance-Based Animation. *IEEE computer graphics and applications* 27, 3 (2007), 21–31.
- [579] Thanos G Stavropoulos, Georgios Meditskos, Stelios Andreadis, Konstantinos Avgerinakis, Katerina Adam, and Ioannis Kompatsiaris. 2020. Semantic event fusion of computer vision and ambient sensor data for activity recognition to support dementia care. *Journal of Ambient Intelligence and Humanized Computing* 11, 8 (2020), 3057–3072. <https://doi.org/10.1007/s12652-016-0437-5>
- [580] Thomas Stiefmeier, Daniel Roggen, Georg Ogris, Paul Lukowicz, and Gerhard Tröster. 2008. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing* 7, 2 (2008), 42–50.
- [581] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 2015. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *13th ACM Conference on Embedded Networked Sensor Systems*. 127–140. <https://doi.org/10.1145/2809695.2809718>
- [582] Petra Maria Strauß, Holger Hoffmann, Wolfgang Minker, Heiko Neumann, Gunther Palm, Stefan Scherer, Harald C. Traue, and Ulrich Weidenbacher. 2008. The PIT corpus of German multi-party dialogues. In *6th International Conference on Language Resources and Evaluation, LREC 2008*. 2442–2445.
- [583] Yuting Su, Yang Li, and Anan Liu. 2019. Open-view human action recognition based on linear discriminant analysis. *Multimedia Tools and Applications* 78, 1 (2019), 767–782. <https://doi.org/10.1007/s11042-018-5657-6>
- [584] Waqas Sultani, Chen Chen, and Mubarak Shah. 2018. Real-world Anomaly Detection in Surveillance Videos. In *IEEE conference on computer vision and pattern recognition*. <https://doi.org/10.1145/3297001.3297034>
- [585] Jiande Sun, Yufei Wang, Jing Li, Wenbo Wan, De Cheng, and Huaxiang Zhang. 2018. View-invariant gait recognition based on kinect skeleton feature. *Multimedia Tools and Applications* 77, 19 (2018), 24909–24935. <https://doi.org/10.1007/s11042-018-5722-1>
- [586] Li Sun, Mingli Song, Jiajun Bu, and Chun Chen. 2012. Pose Estimation with Motionlet LLC Coding. In *PCM*, Weisi Lin, Dong Xu, Anthony Ho, Jianxin Wu, Ying He, Jianfei Cai, Mohan Kankanhalli, and Ming-Ting Sun (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 435–443.
- [587] Pei Sun, Henrik Kretschmar, Xerxes Dotiwalla, Aurélien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. 2020. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2443–2451. <https://doi.org/10.1109/CVPR42600.2020.00252>
- [588] Xiaofan Sun, Jeroen Lichtenauer, Michel Valstar, Anton Nijholt, and Maja Pantic. 2011. A Multimodal Database for Mimicry Analysis. In *Affective Computing and Intelligent Interaction*, Sidney D’Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 367–376.
- [589] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. 2011. Human Activity Detection from RGBD Images. *Artificial Intelligence abs/1107.0* (2011), 47–55.
- [590] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. 2012. Unstructured human activity detection from RGBD images. In *IEEE International Conference on Robotics and Automation*. IEEE, 842–849. <https://doi.org/10.1109/ICRA.2012.6224591>
- [591] Adam Świtoński, Bartosz Piórkowski, Henryk Josiński, Konrad Wojciechowski, and Aldona Drabik. 2015. Recognition of Human Gestures Represented by Depth Camera Motion Sequences. In *New Trends in Intelligent Information and Database Systems*, Dariusz Barbucha, Ngoc Thanh Nguyen, and John Batubara (Eds.). Springer International Publishing, Cham, 33–42.
- [592] Timo Sztyler and Heiner Stuckenschmidt. 2016. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *IEEE International Conference on Pervasive Computing and Communications, PerCom 2016*. IEEE. <https://doi.org/10.1109/PERCOM.2016.7456521>
- [593] Omid Taheri, Nima Ghorbani, Michael J Black, and Dimitrios Tzionas. 2020. GRAB: A dataset of whole-body human grasping of objects. In *European Conference on Computer Vision*. Springer, 581–600.
- [594] Valtteri Takala and Matti Pietikäinen. 2012. CMV100: A dataset for people tracking and re-identification in sparse camera networks. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 1387–1390.
- [595] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. 2018. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications* 10, 1 (2018), 4.
- [596] Daoliang Tan, Kaiqi Huang, Shiqi Yu, and Tieniu Tan. 2006. Efficient night gait recognition based on template matching. *Proceedings - International Conference on Pattern Recognition* 3 (2006), 1000–1003. <https://doi.org/10.1109/ICPR.2006.478>
- [597] Chang Tang, Wanqing Li, Pichao Wang, and Lizhe Wang. 2018. Online human action recognition based on incremental learning of weighted covariance descriptors. *Information Sciences* 467 (2018), 219–237. <https://doi.org/10.1016/j.ins.2018.08.003>
- [598] Julian Tanke, Oh-Hun Kwon, Patrick Stotko, Radu Alexandru Rosu, Michael Weinmann, Hassan Errami, Sven Behnke, Maren Bennewitz, Reinhard Klein, and Andreas Weber. 2019. Bonn Activity Maps: Dataset Description. *arXiv* (2019).

- [599] Ashish Tawari, Cuong Tran, Anup Doshi, Thorsten Zander, and Mohan Trivedi. 2012. Distributed multisensory signals acquisition and analysis in dyadic interactions. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Austin, Texas, USA, 2261–2266. <https://doi.org/10.1145/2212776.2223786>
- [600] Jawad Tayyub, Majd Hawasly, David C Hogg, and Anthony G Cohn. 2017. CLAD: a complex and long activities dataset with rich crowdsourced annotations. *arXiv* (2017).
- [601] Moritz Tenorth, Jan Bandouch, and Michael Beetz. 2009. The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition Moritz. In *2009 IEEE 12th International Conference on Computer Vision Workshops*. 1089–1096.
- [602] Ilias Theodorakopoulos, Dimitris Kastaniotis, George Economou, and Spiros Fotopoulos. 2014. Pose-based human action recognition via sparse representation in dissimilarity space. *Journal of Visual Communication and Image Representation* 25, 1 (2014), 12–23. <https://doi.org/10.1016/j.jvci.2013.03.008>
- [603] Lucas Thomeer, Shanyuanye Guan, Hans Gray, Anthony Schache, Richard de Steiger, and Marcus Pandy. 2020. Six-Degree-of-Freedom Tibiofemoral and Patellofemoral Joint Motion During Activities of Daily Living. *Annals of Biomedical Engineering* 49, 4 (2020), 1183–1198. <https://doi.org/10.1007/s10439-020-02646-2>
- [604] G Thrane, M Alt Murphy, and K S Sunnerhagen. 2018. Recovery of kinematic arm function in well-performing people with subacute stroke: a longitudinal cohort study. *J Neuroeng Rehabil* 15, 1 (2018), 67. <https://doi.org/10.1186/s12984-018-0409-4>
- [605] G Thrane, K S Sunnerhagen, and M A Murphy. 2020. Upper limb kinematics during the first year after stroke: the stroke arm longitudinal study at the University of Gothenburg (SALGOT). *J Neuroeng Rehabil* 17, 1 (2020), 76. <https://doi.org/10.1186/s12984-020-00705-2>
- [606] Atousa Torabi, Christopher Pal, Hugo Laroche, and Aaron Courville. 2015. Using Descriptive Video Services to Create a Large Data Source for Video Annotation Research. *arXiv* (2015). <http://arxiv.org/abs/1503.01070>
- [607] Dinh Tuan Tran, Hirotake Yamazoe, and Joo-Ho Lee. 2020. Multi-scale affined-HOF and dimension selection for view-unconstrained action recognition. *Applied Intelligence* 50, 5 (2020), 1468–1486. <https://doi.org/10.1007/s10489-019-01572-8>
- [608] Thanh-Hai Tran, Thi-Lan Le, Dinh-Tan Pham, Van-Nam Hoang, Van-Minh Khong, Quoc-Toan Tran, Thai-Son Nguyen, and Cuong Pham. 2018. A multi-modal multi-view dataset for human fall analysis and preliminary investigation on modality. In *24th International Conference on Pattern Recognition (ICPR)*. IEEE, 1947–1952.
- [609] TRECVIDdataset 2015.
- [610] M Trinidad-Fernández, D Beckwée, A Cuesta-Vargas, M González-Sánchez, F A Moreno, J González-Jiménez, E Joos, and P Vaes. 2020. Validation, Reliability, and Responsiveness Outcomes Of Kinematic Assessment With An RGB-D Camera To Analyze Movement In Subacute And Chronic Low Back Pain. *Sensors (Basel)* 20, 3 (2020). <https://doi.org/10.3390/s20030689>
- [611] UCFARGDdataset [n.d.]. UCF-ARG Data Set. <https://www.crcv.ucf.edu/data/UCF-ARG.php>.
- [612] UCSDAnomalyDetectiondataset [n.d.]. UCSD Anomaly Detection Dataset. <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>.
- [613] Wallace Ugulino, Débora Cardador, Katia Vega, Eduardo Velloso, Ruy Milidiú, and Hugo Fuks. 2012. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. In *SBLA*, Leliane N Barros, Marcelo Finger, Aurora T Pozo, Gustavo A Giménez-Lugo, and Marcos Castilho (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 52–61.
- [614] UMNCCrowdActivitydataset [n.d.].
- [615] Ákos Utasi and Andrea Kovács. 2010. Recognizing human actions by using spatio-temporal motion descriptors. In *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 366–375.
- [616] Aleksandar Vakanski, Hyung Pil Jun, David Paul, and Russell Baker. 2018. A data set of human body movements for physical rehabilitation exercises. *Data* 3, 1 (2018). <https://doi.org/10.3390/data3010002>
- [617] Tim A Valk, Leonora J Mouton, Egbert Otten, and Raoul M Bongers. 2019. Fixed muscle synergies and their potential to improve the intuitive control of myoelectric assistive technology for upper extremities. *Journal of NeuroEngineering and Rehabilitation* 16, 1 (2019), 6. <https://doi.org/10.1186/s12984-018-0469-5>
- [618] N. P. van der Aa, X. Luo, G. J. Giezeman, R. T. Tan, and R. C. Veltkamp. 2011. UMPM benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction. In *IEEE International Conference on Computer Vision*. IEEE, 1264–1269. <https://doi.org/10.1109/ICCVW.2011.6130396>
- [619] Coert van Gemeren, Robby T Tan, Ronald Poppe, and Remco C Veltkamp. 2014. Dyadic Interaction Detection from Pose and Flow. In *Human Behavior Understanding*, Hyun Soo Park, Albert Ali Salah, Yong Jae Lee, Louis-Philippe Morency, Yaser Sheikh, and Rita Cucchiara (Eds.). Springer International Publishing, Cham, 101–115.
- [620] Jan Vanrie and Karl Verfaillie. 2004. Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers* 36, 4 (2004), 625–629.
- [621] Geoffrey Vaquette, Astrid Orcesi, Laurent Lucat, and Catherine Achard. 2017. The daily home life activity dataset: a high semantic activity dataset for online recognition. In *12th IEEE International Conference on Automatic Face Gesture Recognition (FG)*. IEEE, 497–504.
- [622] Yuri Vladimir Huallpa Vargas, Naysha Naydu Diaz Ccasa, and Lauro Enciso Rodas. 2019. Peruvian Sign Language Recognition Using a Hybrid Deep Neural Network. In *Information Management and Big Data*, Juan Antonio Lössio-Ventura, Nelly Condori-Fernandez, and Jorge Carlos Valverde-Rebaza (Eds.). Springer International Publishing, Cham, 165–172.
- [623] Ashok Veeraraghavan, Rama Chellappa, and Amit K. Roy-Chowdhury. 2006. The function space of an activity. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 959–966. <https://doi.org/10.1109/CVPR.2006.304>

- [624] Sergio A. Velastin and Diego A. Gómez-Lira. 2017. People detection and pose classification inside a moving train using computer vision. In *Lecture Notes in Computer Science*. 319–330. [https://doi.org/10.1007/978-3-319-70010-6\\_30](https://doi.org/10.1007/978-3-319-70010-6_30)
- [625] Andreas Verleysen, Matthijs Biondina, and Francis Wyffels. 2020. Video dataset of human demonstrations of folding clothing for robotic folding. *The International Journal of Robotics Research* 39, 9 (2020), 1031–1036.
- [626] Visintdataset [n.d.]. Mind’s Eye Year 1 2 Videos. <https://sites.google.com/site/visintorg/datasets>.
- [627] Ekaterina Volkova, Stephan De La Rosa, Heinrich H. Bühlhoff, and Betty Mohler. 2014. The MPI emotional body expressions database for narrative scenarios. *PLoS ONE* 9, 12 (2014), 1–28. <https://doi.org/10.1371/journal.pone.0113647>
- [628] Gualtiero Volpe, Ksenia Kolykhalova, Erica Volta, Simone Ghisio, George Waddell, Paolo Albornò, Stefano Piana, Corrado Canepa, and Rafael Ramirez-Melendez. 2017. A multimodal corpus for technology-enhanced learning of violin playing. <https://doi.org/10.1145/3125571.3125588>
- [629] Timo von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. 2018. Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera. In *ECCV*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 614–631.
- [630] Timo von Marcard, Gerard Pons-Moll, and Bodo Rosenhahn. 2016. *Multimodal motion capture dataset TNT15*. Technical Report.
- [631] Michalis Vrigkas, Christophoros Nikou, and Ioannis A Kakadiadis. 2014. Classifying Behavioral Attributes Using Conditional Random Fields. In *Artificial Intelligence: Methods and Applications*, Aristidis Likas, Konstantinos Blekas, and Dimitris Kalles (Eds.). Springer International Publishing, Cham, 95–104.
- [632] Priscilla K. Wagner, Sarajane M. Peres, Clodoaldo A.M. Lima, Fernando A. Freitas, and Renata Cristina Barros Madeo. 2014. Gesture unit segmentation using spatial-temporal information and Machine Learning. In *27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014*. 101–106.
- [633] Christian Wallraven, Michael Schultze, Betty Mohler, Argiro Vatakis, and Katerina Pastra. 2011. The POETICON enacted scenario corpus A tool for human and computational experiments on action understanding. In *IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG*. 484–491. <https://doi.org/10.1109/FG.2011.5771446>
- [634] Jun Wan, Stan Z. Li, Yibing Zhao, Shuai Zhou, Isabelle Guyon, and Sergio Escalera. 2016. ChaLearn Looking at People RGB-D Isolated and Continuous Datasets for Gesture Recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 761–769. <https://doi.org/10.1109/CVPRW.2016.100>
- [635] Isaac Wang, Mohtadi Ben Fraj, Pradyumna Narayana, Dhruva Patil, Gururaj Mulay, Rahul Bangar, J. Ross Beveridge, Bruce A. Draper, and Jaime Ruiz. 2017. EGGNOG: A Continuous, Multi-modal Data Set of Naturally Occurring Gestures with Ground Truth Labels. In *12th IEEE International Conference on Automatic Face and Gesture Recognition, FG*. IEEE, 414–421. <https://doi.org/10.1109/FG.2017.145>
- [636] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. 2012. Mining actionlet ensemble for action recognition with depth cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1290–1297. <https://doi.org/10.1109/CVPR.2012.6247813>
- [637] Jiang Wang, Xiaohan Nie, Yin Xia, Ying Wu, and Song Chun Zhu. 2014. Cross-view action modeling, learning, and recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2649–2656. <https://doi.org/10.1109/CVPR.2014.339>
- [638] Jian-ping Wang, Shi-hua Wang, Xuan Zhao, Hai Hu, Yan-qing Wang, Jin-lai Liu, Xu Chen, and Yu Li. 2020. A Data Process of Human Knee Joint Kinematics Obtained by Motion-Capture Measurement. *BMC Medical Informatics and Decision Making* 21, 1 (2020), 1–18.
- [639] Keze Wang, Xiaolong Wang, Liang Lin, Meng Wang, and Wangmeng Zuo. 2015. 3D Human Activity Recognition with Reconfigurable Convolutional Neural Networks. In *22nd ACM International conference on Multimedia*. 97–106. <https://doi.org/10.1145/2647868.2654912>
- [640] Kaiye Wang, Zhang Zhang, and Liang Wang. 2012. Violence Video Detection by Discriminative Slow Feature Analysis. In *Pattern Recognition*, Cheng-Lin Liu, Changshui Zhang, and Liang Wang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 137–144.
- [641] Lan Wang, Chenqiang Gao, Luyu Yang, Yue Zhao, Wangmeng Zuo, and Deyu Meng. 2018. PM-GANs: Discriminative Representation Learning for Action Recognition Using Partial-Modalities. In *ECCV*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 389–406.
- [642] Lichen Wang, Bin Sun, Joseph Robinson, Taotao Jing, and Yun Fu. 2020. EV-Action: Electromyography-Vision multi-modal action dataset. *arXiv* (2020).
- [643] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* 25, 12 (2003), 1505–1518.
- [644] Pichao Wang, Wanqing Li, Zhimin Gao, Jing Zhang, Chang Tang, and Philip O. Ogunbona. 2016. Action Recognition from Depth Maps Using Deep Convolutional Neural Networks. *IEEE Transactions on Human-Machine Systems* 46, 4 (2016), 498–509. <https://doi.org/10.1109/THMS.2015.2504550>
- [645] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2014. Person Re-identification by Video Ranking. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 688–703.
- [646] Xiaolong Wang, Ali Farhadi, and Abhinav Gupta. 2016. Actions ~ Transformations. In *IEEE conference on Computer Vision and Pattern Recognition*. 2658–2667.
- [647] Xiaogang Wang, Xiaoxu Ma, and W. Eric L. Grimson. 2009. Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 3 (2009), 539–555. <https://doi.org/10.1109/TPAMI.2008.87>
- [648] Zhe Wang, Liyan Chen, Shaurya Rathore, Daeyun Shin, and Charless Fowlkes. 2019. Geometric Pose Affordance: 3D Human Pose with Scene Constraints. *arXiv* (2019).

- [649] Haolin Wei, David S Monaghan, Noel E O'Connor, and Patricia Scanlon. 2014. A New Multi-modal Dataset for Human Affect Analysis. In *International Workshop on Human Behavior Understanding*. Springer, 42–51.
- [650] Daniel Weinland, Mustafa Özyusal, and Pascal Fua. 2010. Making Action Recognition Robust to Occlusions and Viewpoint Changes. In *ECCV*, Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 635–648.
- [651] Daniel Weinland, Remi Ronfard, and Edmond Boyer. 2006. Free viewpoint action recognition using motion history volumes. *Computer vision and image understanding* 104, 2-3 (2006), 249–257.
- [652] Philippe Weinzaepfel, Xavier Martin, and Cordelia Schmid. 2017. Towards Weakly-Supervised Action Localization. *arXiv* (2017).
- [653] Anjana Wijekoon, Nirmalie Wiratunga, Sadiq Sani, Stewart Massie, and Kay Cooper. 2018. Improving kNN for Human Activity Recognition with Privileged Learning Using Translation Models. In *Case-Based Reasoning Research and Development*, Michael T Cox, Peter Funk, and Shahina Begum (Eds.). Springer International Publishing, Cham, 448–463.
- [654] Wijekoon2019dataset [n.d.]. MEX - Multi-modal Exercise Dataset for Human Activity Recognition. <https://data.mendeley.com/datasets/p89fwbzmkd/2>.
- [655] Kamil Wnuk and Stefano Soatto. 2011. Analyzing Diving: A Dataset for Judging Action Quality. In *ACCV*, Reinhard Koch and Fay Huang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 266–276.
- [656] Christian Wojek, Stefan Walk, and Bernt Schiele. 2009. Multi-Cue onboard pedestrian detection. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. IEEE, 794–801. <https://doi.org/10.1109/CVPRW.2009.5206638>
- [657] Christian Wolf, Eric Lombardi, Julien Mille, Oya Celiktutan, Mingyuan Jiu, Emre Dogan, Gonen Eren, Moez Baccouche, Emmanuel Dellandréa, Charles-Edmond Bichot, Christophe Garcia, and Bülent Sankur. 2014. Evaluation of video activity localizations integrating quality and quantity measurements. *Computer Vision and Image Understanding* 127 (2014), 14–30. <https://doi.org/10.1016/j.cviu.2014.06.014>
- [658] Chenxia Wu, Jiemi Zhang, Silvio Savarese, and Ashutosh Saxena. 2015. Watch-n-Patch : Unsupervised Understanding of Actions and Relations. In *CVPR*. 4362–4370.
- [659] Yi Wu, Jongwoo Lim, and Ming Hsuan Yang. 2013. Online object tracking: A benchmark. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2411–2418. <https://doi.org/10.1109/CVPR.2013.312>
- [660] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. 2020. Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
- [661] Lu Xia, Chia Chih Chen, and J. K. Aggarwal. 2012. View invariant human action recognition using histograms of 3D joints. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 20–27. <https://doi.org/10.1109/CVPRW.2012.6239233>
- [662] Chi Xu, Yasushi Makihara, Gakuto Ogi, Xiang Li, Yasushi Yagi, and Jianfeng Lu. 2017. The ou-isir gait database comprising the large population dataset with age and performance evaluation of age estimation. *IPSJ Transactions on Computer Vision and Applications* 9, 1 (2017), 24.
- [663] Jun Xu, Tao Mei, Ting Yao, and Yong Rui. 2016. MSR-VTT: A Large Video Description Dataset for Bridging Video and Language. In *IEEE conference on computer vision and pattern recognition*. 5288–5296. <https://doi.org/10.1109/cvpr.2016.571>
- [664] Ning Xu, Anan Liu, Weizhi Nie, Yongkang Wong, Fuwu Li, and Yuting Su. 2015. Multi-modal multi-view interactive benchmark dataset for human action recognition. In *Proceedings of the 23rd ACM international conference on Multimedia*. 1195–1198.
- [665] Santosh Kumar Yadav, Amitojdeep Singh, Abhishek Gupta, and Jagdish Lal Raheja. 2019. Real-time Yoga recognition using deep learning. *Neural Computing and Applications* 31, 12 (2019), 9349–9361. <https://doi.org/10.1007/s00521-019-04232-7>
- [666] Seiichi Yamamoto, Keiko Taguchi, Koki Ijuin, Ichiro Umata, and Masafumi Nishida. 2015. Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it. *Language Resources and Evaluation* 49, 4 (2015), 857–882. <https://doi.org/10.1007/s10579-015-9299-2>
- [667] Allen Y Yang, Philip Kuryloski, and Ruzena Bajcsy. 2009. WARD : A Wearable Action Recognition Database. In *CHI LBR*.
- [668] Jinlong Yang, Jean-Sébastien Franco, Franck Hérouy-Wheeler, and Stefanie Wuhrer. 2016. Estimation of Human Body Shape in Motion with Wide Clothing. In *ECCV*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 439–454.
- [669] Seunghan Yang, Seungjun Jung, Heekwang Kang, and Changick Kim. 2020. The Korean Sign Language Dataset for Action Recognition. In *MultiMedia Modeling*, Yong Man Ro, Wen-Huang Cheng, Junmo Kim, Wei-Ta Chu, Peng Cui, Jung-Woo Choi, Min-Chun Hu, and Wesley De Neve (Eds.). Springer International Publishing, Cham, 532–542.
- [670] Serena Yeung, Olga Russakovsky, Ning Jin, Mykhaylo Andriluka, Greg Mori, and Li Fei-Fei. 2017. Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos. *International Journal of Computer Vision* 126, 2-4 (2017), 375–389. <https://doi.org/10.1007/s11263-017-1013-y>
- [671] Kristina Yordanova, Albert Hein, and Thomas Kirste. 2020. Kitchen task assessment dataset for measuring errors due to cognitive impairments. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 1–6.
- [672] Yuya Yoshikawa, Jiaqing Lin, and Akikazu Takeuchi. 2018. Stair actions: A video dataset of everyday home actions. *arXiv* (2018).
- [673] Gang Yu, Zicheng Liu, and Junsong Yuan. 2015. Discriminative Orderlet Mining for Real-Time Recognition of Human-Object Interaction. In *ACCV*, Daniel Cremers, Ian Reid, Hideo Saito, and Ming-Hsuan Yang (Eds.). Springer International Publishing, Cham, 50–65.
- [674] J Yu, A Lei, Z Song, T Wang, H Cai, and N Feng. 2018. Comprehensive Dataset of Broadcast Soccer Videos. *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (2018). <https://doi.org/10.1109/MIPR.2018.00090>
- [675] Shiqi Yu, Daoliang Tan, and Tieniu Tan. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *International Conference on Pattern Recognition*. 441–444. <https://doi.org/10.1109/ICPR.2006.67>
- [676] Shiqi Yu, Qing Wang, and Yongzhen Huang. 2013. A large RGB-D gait dataset and the baseline algorithm. In *Chinese Conference on Biometric Recognition*. Springer, 417–424.

- [677] Tsz Ho Yu, Tae Kyun Kim, and Roberto Cipolla. 2013. Unconstrained monocular 3D human pose estimation by action detection and cross-modality regression forest. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3642–3649. <https://doi.org/10.1109/CVPR.2013.467>
- [678] Junsong Yuan, Zicheng Liu, and Ying Wu. 2009. Discriminative subvolume search for efficient action detection. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2442–2449. <https://doi.org/10.1109/cvpr.2009.5206671>
- [679] Kiwon Yun, Jean Honorio, Debaleena Chattopadhyay, Tamara L. Berg, and Dimitris Samaras. 2012. Two-person interaction detection using body-pose features and multiple instance learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (2012)*, 28–35. <https://doi.org/10.1109/CVPRW.2012.6239234>
- [680] Xenophon Zabulis, Dimitris Grammenos, Thomas Sarmis, Konstantinos Tzevanidis, Pashalis Paderleris, Panagiotis Koutlemanis, and Antonis A Argyros. 2013. Multicamera human detection and tracking supporting natural interaction with large-scale displays. *Machine Vision and Applications* 24, 2 (2013), 319–336. <https://doi.org/10.1007/s00138-012-0408-6>
- [681] Janez Zaletelj and Andrej Košir. 2017. Predicting students’ attention in the classroom from Kinect facial and body features. *EURASIP Journal on Image and Video Processing* 2017, 1 (2017), 80. <https://doi.org/10.1186/s13640-017-0228-8>
- [682] Cemil Zalluhoglu and Nazli Ikizler-Cinbis. 2020. Collective Sports: A multi-task dataset for collective activity recognition. *Image and Vision Computing* 94 (2020), 103870.
- [683] Aurélie Zara, Valérie Maffiolo, Jean Claude Martin, and Laurence Devillers. 2007. Collection and annotation of a corpus of human-human multimodal interactions: Emotion and others anthropomorphic characteristics. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 464–475.
- [684] Gloria Zen, Bruno Lepri, Elisa Ricci, and Oswald Lanz. 2010. Space speaks - Towards socially and personality aware visual surveillance. In *ACM Workshop on Multimodal Pervasive Video Analysis*. 37–42. <https://doi.org/10.1145/1878039.1878048>
- [685] Jing Zhang, Wanqing Li, Pichao Wang, Philip Ogunbona, Song Liu, and Chang Tang. 2016. A large scale rgb-d dataset for action recognition. In *International Workshop on Understanding Human Activities through 3D Sensors*. Springer, 101–114.
- [686] Mingming Zhang, Lu Yu, Keye Zhang, Bixuan Du, Bin Zhan, Shaohua Chen, Xiuhao Jiang, Shuai Guo, Jiafeng Zhao, and Yang Wang. 2020. Kinematic dataset of actors expressing emotions. *Scientific data* 7, 1 (2020), 1–8.
- [687] Weichen Zhang, Zhiguang Liu, Liuyang Zhou, Howard Leung, and Antoni B Chan. 2017. Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation. *Image and Vision Computing* 61 (2017), 22–39.
- [688] Weiyu Zhang, Menglong Zhu, and Konstantinos G. Derpanis. 2013. From actemes to action: A strongly-supervised representation for detailed action understanding. In *IEEE International Conference on Computer Vision*. 2248–2255. <https://doi.org/10.1109/ICCV.2013.280>
- [689] Xinfeng Zhang, Su Yang, Yuan Yan Tang, and Weishan Zhang. 2016. A thermodynamics-inspired feature for anomaly detection on crowd motions in surveillance videos. *Multimedia Tools and Applications* 75, 14 (2016), 8799–8826. <https://doi.org/10.1007/s11042-015-3101-8>
- [690] Yuting Zhang, Gang Pan, Kui Jia, Minlong Lu, Yueming Wang, and Zhaohui Wu. 2015. Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points with Clusters. *IEEE Transactions on Cybernetics* 45, 9 (2015), 1864–1875. <https://doi.org/10.1109/TCYB.2014.2361287>
- [691] Zhong Zhang, Christopher Conly, and Vassilis Athitsos. 2014. Evaluating Depth-Based Computer Vision Methods for Fall Detection under Occlusions. In *International Symposium on Visual Computing*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ryan McMahan, Jason Jerald, Hui Zhang, Steven M Drucker, Chandra Kambhampettu, Maha El Choubassi, Zhigang Deng, and Mark Carlson (Eds.). Springer International Publishing, Cham, 196–207.
- [692] Zhong Zhang, Weihua Liu, Vangelis Metsis, and Vassilis Athitsos. 2012. A Viewpoint-Independent Statistical Method for Fall Detection. In *International Conference on Pattern Recognition*. IEEE, 3626–3630.
- [693] Hang Zhao, Antonio Torralba, Lorenzo Torresani, and Zhicheng Yan. 2019. Hacs: Human action clips and segments dataset for recognition and temporal localization. In *IEEE International Conference on Computer Vision*. 8668–8678.
- [694] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. 2016. MARS: A Video Benchmark for Large-Scale Person Re-Identification. In *ECCV*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.). Springer International Publishing, Cham, 868–884.
- [695] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *IEEE international conference on computer vision*. 1116–1124.
- [696] Bolei Zhou, Xiaoou Tang, and Xiaogang Wang. 2013. Measuring crowd collectiveness. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3049–3056. <https://doi.org/10.1109/CVPR.2013.392>
- [697] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. 2011. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 3441–3448. <https://doi.org/10.1109/CVPR.2011.5995459>
- [698] Feng Zhu, Xiaogang Wang, and Nenghai Yu. 2014. Crowd Tracking with Dynamic Evolution of Group Structures. In *ECCV*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 139–154.
- [699] Aneeq Zia, Yachna Sharma, Vinay Bettadapura, Eric L Sarin, and Irfan Essa. 2018. Video and accelerometer-based motion analysis for automated surgical skills assessment. *International Journal of Computer Assisted Radiology and Surgery* 13, 3 (2018), 443–455. <https://doi.org/10.1007/s11548-018-1704-z>
- [700] Athanasia Zlatintsi, Petros Koutras, Georgios Evangelopoulos, Nikolaos Malandrakis, Niki Efthymiou, Katerina Pastra, Alexandros Potamianos, and Petros Maragos. 2017. COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization. *EURASIP Journal on Image and Video Processing* 2017, 1 (2017), 54. <https://doi.org/10.1186/s13640-017-0194-1>

- [701] Andreas Zweng, Sebastian Zambanini, and Martin Kampel. 2010. Introducing a statistical behavior model into camera-based fall detection. In *International Symposium on Visual Computing*. Springer, 163–172.