References

[1] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "Amass: Archive of motion capture as surface shapes," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5442–5451.

[2] M. Loper, N. Mahmood, and M. J. Black, "Mosh: Motion and shape capture from sparse markers," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–13, 2014.

[3] A. Haque, B. Peng, Z. Luo, A. Alahi, S. Yeung, and L. Fei-Fei, "Towards viewpoint invariant 3d human pose estimation," in *European Conference on Computer Vision*. Springer, 2016, pp. 160–177.

[4] T. Yu, Z. Zheng, K. Guo, J. Zhao, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu, "Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7287–7296.

[5] T. von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, "Recovering accurate 3d human pose in the wild using imus and a moving camera," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 601–617.

[6] M. Trumble, A. Gilbert, C. Malleson, A. Hilton, and J. P. Collomosse, "Total capture: 3d human pose estimation fusing video and inertial sensors." in *BMVC*, vol. 2, no. 5, 2017, pp. 1–13.

[7] M. Dantone, J. Gall, C. Leistner, and L. Van Gool, "Human pose estimation using body parts dependent joint regressors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3041–3048.

[8] Y. Chen, C. Shen, X.-S. Wei, L. Liu, and J. Yang, "Adversarial posenet: A structure-aware convolutional network for human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1212–1221.

[9] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille, and X. Wang, "Multi-context attention for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1831–1840.

[10] X. Sun, J. Shang, S. Liang, and Y. Wei, "Compositional human pose regression," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2602–2611.

[11] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3d human pose," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7025–7034.

[12] H. Yasin, U. Iqbal, B. Kruger, A. Weber, and J. Gall, "A dual-source approach for 3d pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4948–4956.

[13] W. Yang, W. Ouyang, X. Wang, J. Ren, H. Li, and X. Wang, "3d human pose estimation in the wild by adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5255–5264.

[14] B. Wandt and B. Rosenhahn, "Repnet: Weakly supervised training of an adversarial reprojection network for 3d human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7782–7791.

[15] S. Li, L. Ke, K. Pratama, Y.-W. Tai, C.-K. Tang, and K.-T. Cheng, "Cascaded deep monocular 3d human pose estimation with evolutionary training data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6173–6183.

[16] G. Chéron, A. Osokin, I. Laptev, and C. Schmid, "Modeling spatio-temporal human track structure for action localization," *arXiv preprint arXiv:1806.11008*, 2018.

[17] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep bi-directional lstm with cnn features," *IEEE Access*, vol. 6, pp. 1155–1166, 2018.

[18] X. Perez-Sala, S. Escalera, C. Angulo, and J. Gonzalez, "A survey on model based approaches for 2d and 3d visual human pose recovery," *Sensors*, vol. 14, no. 3, pp. 4189–4210, 2014.

[19] Z. Liu, J. Zhu, J. Bu, and C. Chen, "A survey of human pose estimation: the body parts parsing based methods," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 10–19, 2015.

[20] H.-B. Zhang, Q. Lei, B.-N. Zhong, J.-X. Du, and J. Peng, "A survey on human pose estimation," *Intelligent Automation & Soft Computing*, vol. 22, no. 3, pp. 483–489, 2016.

[21] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep learning based 2d human pose estimation: A survey," *Tsinghua Science and Technology*, vol. 24, no. 6, pp. 663–676, 2019.

[22] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[23] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: a review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 13–24, 2009.

[24] H. Jung and Y.-E. Song, "Robotic remote control based on human motion via virtual collaboration system: A survey," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 12, no. 7, pp. JAMDSM0126–JAMDSM0126, 2018.

[25] S. Xia, L. Gao, Y.-K. Lai, M.-Z. Yuan, and J. Chai, "A survey on human performance capture and animation," *Journal of Computer Science and Technology*, vol. 32, no. 3, pp. 536–554, 2017.

[26] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial intelligence review*, vol. 43, no. 1, pp. 1–54, 2015.

[27] F. Noroozi, D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE transactions on affective computing*, 2018.

[28] R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing*, vol. 28, no. 6, pp. 976–990, 2010.

[29] D. D. Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector," *The Visual Computer*, vol. 32, no. 3, pp. 289–306, 2016.

[30] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[31] T. Kawase, T. Sakurada, Y. Koike, and K. Kansaku, "A hybrid bmi-based exoskeleton for paresis: Emg control for assisting arm movements," *Journal of Neural Engineering*, vol. 14, no. 1, p. 016015, 2017.

[32] A. Mehmood, A. Nadeem, M. Ashraf, T. Alghamdi, and M. S. Siddiqui, "A novel fall detection algorithm for elderly using shimmer wearable sensors," *Health and Technology*, vol. 9, no. 4, pp. 631–646, 2019.

[33] X. Song, K. Mann, E. Allison, S.-C. Yoon, H. Hila, A. Muller, and C. Gieder, "A quadcopter controlled by brain concentration and eye blink," in *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2016, pp. 1–4.

[34] D. Tan, Y.-H. Pua, S. Balakrishnan, A. Scully, K. J. Bower, K. M. Prakash, E.-K. Tan, J.-S. Chew, E. Poh, S.-B. Tan *et al.*, "Automated analysis of gait and modified timed up and go using the microsoft kinect in people with parkinson's disease: associations with physical outcome measures," *Medical & biological engineering & computing*, vol. 57, no. 2, pp. 369–377, 2019.

[35] M. Zhang, Z. Zhang, Y. Chang, E.-S. Aziz, S. Esche, and C. Chassapis, "Recent developments in game-based virtual reality educational laboratories using the microsoft kinect," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 1, pp. 138–159, 2018.

[36] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 545–557, 2016.

[37] M. Ariz, A. Villanueva, and R. Cabeza, "Robust and accurate 2d-tracking-based 3d positioning method: Application to head pose estimation," *Computer Vision and Image Understanding*, vol. 180, pp. 13–22, 2019.

[38] S. Kusuma, J. D. Udayan, and A. Sachdeva, "Driver distraction detection using deep learning and computer vision," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1. IEEE, 2019, pp. 289–292.

[39] D. V. McGehee, C. A. Roe, L. N. Boyle, Y. Wu, K. Ebe, J. Foley, and L. Angell, "The wagging foot of uncertainty: data collection and reduction methods for examining foot pedal behavior in naturalistic driving," *SAE International journal of transportation safety*, vol. 4, no. 2, pp. 289–294, 2016.

[40] H.-C. Shih, "A survey of content-aware video analysis for sports," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1212–1231, 2017.

[41] T. Gupta, V. Nunavath, and S. Roy, "Crowdvas-net: A deep-cnn based framework to detect abnormal crowd-motion behavior in videos for predicting crowd disaster," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 2877–2882.

[42] M. Gerber, J. Beck, S. Brand, R. Cody, L. Donath, A. Eckert, O. Faude, X. Fischer, M. Hatzinger, E. Holsboer-Trachsler *et al.*, "The impact of lifestyle physical activity counselling in in-patients with major depressive disorders on physical activity, cardiorespiratory fitness, depression, and cardiovascular health risk markers: study protocol for a randomized controlled trial," *Trials*, vol. 20, no. 1, p. 367, 2019.

[43] J. Klenk, S. Wekenmann, L. Schwickert, U. Lindemann, C. Becker, and K. Rapp, "Change of objectively-measured physical activity during geriatric rehabilitation," *Sensors*, vol. 19, no. 24, p. 5451, 2019.

[44] S. Liu and S. Ostadabbas, "Seeing under the cover: A physics guided learning approach for in-bed pose estimation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 236–245.

[45] D. Ahmedt-Aristizabal, S. Denman, K. Nguyen, S. Sridharan, S. Dionisio, and C. Fookes, "Understanding patients' behavior: Vision-based analysis of seizure disorders," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2583–2591, 2019.

[46] C. Stoll, N. Hasler, J. Gall, H.-P. Seidel, and C. Theobalt, "Fast articulated motion tracking using a sums of gaussians body model," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 951–958.

[47] W. Zhang, Z. Liu, L. Zhou, H. Leung, and A. B. Chan, "Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3d human pose estimation," *Image and Vision Computing*, vol. 61, pp. 22–39, 2017.

[48] Š. Obdržálek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel, "Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 1188–1193.

[49] M. Shin, J. Jang, and J. Paik, "Calibration of a surveillance camera using a pedestrian homology-based rectangular model," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 4, pp. 305–312, 2018.

[50] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Harvesting multiple views for marker-less 3d human pose annotations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6988–6997.

[51] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1145–1153.

[52] H. Rhodin, J. Spörri, I. Katircioglu, V. Constantin, F. Meyer, E. Müller, M. Salzmann, and P. Fua, "Learning monocular 3d human pose estimation from multi-view images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8437–8446.

[53] S. Wu, H.-S. Wong, and S. Wang, "Variant semiboost for improving human detection in application scenes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 7, pp. 1595–1608, 2017.

[54] W. G. Aguilar, M. A. Luna, J. F. Moya, V. Abad, H. Ruiz, H. Parra, and W. Lopez, "Cascade classifiers and saliency maps based people detection," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 2017, pp. 501–510.

[55] V. Gajjar, Y. Khandhediya, and A. Gurnani, "Human detection and tracking for video surveillance: A cognitive science approach," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 2805–2809.

[56] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[57] Y. Liu, L. Liu, H. Rezatofighi, T.-T. Do, Q. Shi, and I. Reid, "Learning pairwise relationship for multi-object detection in crowded scenes," *arXiv preprint arXiv:1901.03796*, 2019.

[58] S. V. Mashak, B. Hosseini, M. Mokji, and S. A. R. Abu-Bakar, "Background subtraction for object detection under varying environments," in *2010 International Conference of Soft Computing and Pattern Recognition*, 2010, pp. 123–126.

[59] R. Zhang and J. Ding, "Object tracking and detecting based on adaptive background subtraction," *Procedia Engineering*, vol. 29, pp. 1351–1355, 2012.

[60] J. Guo, J. Wang, R. Bai, Y. Zhang, and Y. Li, "A new moving object detection method based on frame-difference and background subtraction," in *IOP Conference Series: Materials Science and Engineering*, vol. 242, no. 1. IOP Publishing, 2017, p. 012115.

[61] M. Babaee, D. T. Dinh, and G. Rigoll, "A deep convolutional neural network for video sequence background subtraction," *Pattern Recognition*, vol. 76, pp. 635–649, 2018.

[62] Q. D. Y. Ma, Z. Ma, C. Ji, K. Yin, T. Zhu, and C. Bian, "Artificial object edge detection based on enhanced canny algorithm for high-speed railway apparatus identification," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–6.

[63] K. Zhang, Y. Zhang, P. Wang, Y. Tian, and J. Yang, "An improved sobel edge algorithm and fpga implementation," *Procedia computer science*, vol. 131, pp. 243–248, 2018.

[64] C. Zhang, J. Yan, C. Li, and R. Bie, "Contour detection via stacking random forest learning," *Neurocomput.*, vol. 275, no. C, p. 2702–2715, Jan. 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.11.046

[65] S. K. Choudhury, P. K. Sa, R. P. Padhy, S. Sharma, and S. Bakshi, "Improved pedestrian detection using motion segmentation and silhouette orientation," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 13 075–13 114, 2018.

[66] F. Ebadi and M. Norouzi, "Road terrain detection and classification algorithm based on the color feature extraction," in *2017 Artificial Intelligence and Robotics (IRANOPEN)*, 2017, pp. 139–146.

[67] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[68] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. Ieee, 1999, pp. 1150–1157.

[69] D. Damen, P. Bunnun, A. Calway, and W. W. Mayol-Cuevas, "Real-time learning and detection of 3d texture-less objects: A scalable approach." in *BMVC*, no. 2, 2012.

[70] H. Zhang and Q. Cao, "Texture-less object detection and 6d pose estimation in rgb-d images," *Robotics and Autonomous Systems*, vol. 95, pp. 64–79, 2017.

[71] S. Wang, H. Ai, T. Yamashita, and S. Lao, "Combined top-down/bottom-up human articulated pose estimation using adaboost learning," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 3670–3673.

[72] J. Yang, W. Liang, and Y. Jia, "Face pose estimation with combined 2d and 3d hog features," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 2492–2495.

[73] K. Bhuvaneswari and H. A. Rauf, "Edgelet based human detection and tracking by combined segmentation and soft decision," in *2009 International Conference on Control, Automation, Communication and Energy Conservation*. IEEE, 2009, pp. 1–6.

[74] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1. IEEE, 2005, pp. 90–97.

[75] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[76] S. Chen, L. Liang, W. Liang, and H. Foroosh, "3d pose tracking with multitemplate warping and sift correspondences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 2043–2055, 2015.

[77] T. Pfister, J. Charles, and A. Zisserman, "Flowing convnets for human pose estimation in videos," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1913–1921.

[78] Y. Ding, C. Wang, H. Huang, J. Liu, J. Wang, and L. Wang, "Frame-recurrent video inpainting by robust optical flow inference," *arXiv preprint arXiv:1905.02882*, 2019.

[79] C. Sminchisescu and B. Triggs, "Covariance scaled sampling for monocular 3d body tracking," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[80] Y. Kawana, N. Ukita, J.-B. Huang, and M.-H. Yang, "Ensemble convolutional neural networks for pose estimation," *Computer Vision and Image Understanding*, vol. 169, pp. 62–74, 2018.

[81] A. Jain, J. Tompson, Y. LeCun, and C. Bregler, "Modeep: A deep learning framework using motion features for human pose estimation," in *Asian conference on computer vision*. Springer, 2014, pp. 302–315.

[82] A. M. Lehrmann, P. V. Gehler, and S. Nowozin, "A non-parametric bayesian network prior of human pose," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1281–1288.

[83] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "Vnect: Real-time 3d human pose estimation with a single rgb camera," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–14, 2017.

[84] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab, and S. Ilic, "3d pictorial structures revisited: Multiple human pose estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 10, pp. 1929–1942, 2015.

[85] S. Zuffi, O. Freifeld, and M. J. Black, "From pictorial structures to deformable structures," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3546–3553.

[86] X. Chen and A. L. Yuille, "Articulated pose estimation by a graphical model with image dependent pairwise relations," in *Advances in neural information processing systems*, 2014, pp. 1736–1744.

[87] J. Yu, C. Hong, Y. Rui, and D. Tao, "Multitask autoencoder model for recovering human poses," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 5060–5068, 2017.

[88] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[89] H. Sidenbladh, F. De la Torre, and M. J. Black, "A framework for modeling the appearance of 3d articulated figures," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. IEEE, 2000, pp. 368–375.

[90] H. Rhodin, N. Robertini, D. Casas, C. Richardt, H.-P. Seidel, and C. Theobalt, "General automatic human shape and motion capture using volumetric contour cues," in *European conference on computer vision*. Springer, 2016, pp. 509–526.

[91] J. Delanoy, M. Aubry, P. Isola, A. A. Efros, and A. Bousseau, "3d sketching using multi-view deep volumetric prediction," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 1, no. 1, pp. 1–22, 2018.

[92] Y. Liu, J. Gall, C. Stoll, Q. Dai, H.-P. Seidel, and C. Theobalt, "Markerless motion capture of multiple characters using multiview image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2720–2735, 2013.

[93] X. Wei and J. Chai, "Videomocap: Modeling physically realistic human motion from monocular video sequences," in *ACM SIGGRAPH 2010 papers*, 2010, pp. 1–10.

[94] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll, "Video based reconstruction of 3d people models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8387–8397.

[95] C. Sminchisescu and B. Triggs, "Estimating articulated human motion with covariance scaled sampling," *The International Journal of Robotics Research*, vol. 22, no. 6, pp. 371–391, 2003.

[96] F. Cuzzolin and W. Gong, "A belief-theoretical approach to example-based pose estimation," *IEEE Transactions on Fuzzy Systems (under revision)*, 2012.

[97] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas, "Discriminative density propagation for 3d human motion estimation," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 390–397.

[98] R. Rosales and S. Sclaroff, "Learning body pose via specialized maps," in *Advances in neural information processing systems*, 2002, pp. 1263–1270.

[99] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast pose estimation with parameter-sensitive hashing," in *null*. IEEE, 2003, p. 750.

[100] U. Iqbal, A. Doering, H. Yasin, B. Krüger, A. Weber, and J. Gall, "A dual-source approach for 3d human pose estimation from single images," *Computer Vision and Image Understanding*, vol. 172, pp. 37–49, 2018.

[101] P. Witoonchart and P. Chongstitvatana, "Application of structured support vector machine backpropagation to a convolutional neural network for human pose estimation," *Neural Networks*, vol. 92, pp. 39–46, 2017.

[102] S. Li, W. Zhang, and A. B. Chan, "Maximum-margin structured learning with deep networks for 3d human pose estimation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2848–2856.

[103] A. Agarwal and B. Triggs, "Recovering 3d human pose from monocular images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 1, pp. 44–58, 2005.

[104] ——, "3d human pose from silhouettes by relevance vector regression," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II.

[105] B. Tekin, I. Katircioglu, M. Salzmann, V. Lepetit, and P. Fua, "Structured prediction of 3d human pose with deep neural networks," *arXiv preprint arXiv:1605.05180*, 2016.

[106] C. Hong, J. Yu, J. Wan, D. Tao, and M. Wang, "Multimodal deep autoencoder for human pose recovery," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5659–5670, 2015.

[107] X. Zhou, M. Zhu, S. Leonardos, K. G. Derpanis, and K. Daniilidis, "Sparseness meets deepness: 3d human pose estimation from monocular video," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4966–4975.

[108] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1653–1660.

[109] D. C. Luvizon, H. Tabia, and D. Picard, "Human pose regression by combining indirect part detection and contextual information," *Computers & Graphics*, vol. 85, pp. 15–22, 2019.

[110] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas, "Semantic graph convolutional networks for 3d human pose regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3425–3435.

[111] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*. Springer, 2016, pp. 483–499.

[112] C.-J. Chou, J.-T. Chien, and H.-T. Chen, "Self adversarial training for human pose estimation," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2018, pp. 17–30.

[113] A. Nibali, Z. He, S. Morgan, and L. Prendergast, "3d human pose estimation with 2d marginal heatmaps," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1477–1485.

[114] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2640–2649.

[115] I. Ramírez, A. Cuesta-Infante, E. Schiavi, and J. J. Pantrigo, "Bayesian capsule networks for 3d human pose estimation from single 2d images," *Neurocomputing*, 2019.

[116] F. Flitti, M. Bennamoun, D. Q. Huynh, and R. A. Owens, "Probabilistic human pose recovery from 2d images," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 1517–1520.

[117] A. Tejani, R. Kouskouridas, A. Doumanoglou, D. Tang, and T.-K. Kim, "Latent-class hough forests for 6 dof object pose estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 119–132, 2017.

[118] P. Li, H. Ling, X. Li, and C. Liao, "3d hand pose estimation using randomized decision forest with segmentation index points," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 819–827.

[119] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[120] H. Cho and S. M. Yoon, "Divide and conquer-based 1d cnn human activity recognition using test data sharpening," *Sensors*, vol. 18, no. 4, p. 1055, 2018.

[121] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input cnn-gru based human activity recognition using wearable sensors," *Computing*, pp. 1–18, 2021.

[122] B. Almaslukh, J. AlMuhtadi, and A. Artoli, "An effective deep autoencoder approach for online smartphone-based human activity recognition," *Int. J. Comput. Sci. Netw. Secur*, vol. 17, no. 4, pp. 160–165, 2017.

[123] D. Balabka, "Semi-supervised learning for human activity recognition using adversarial autoencoders," in *Adjunct Proceedings of the 2019 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, 2019, pp. 685–688.

[124] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *International cross-domain conference for machine learning and knowledge extraction*. Springer, 2017, pp. 267–274.

[125] R. Mutegeki and D. S. Han, "A cnn-lstm approach to human activity recognition," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 362–366.

[126] M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari, "Articulated human pose estimation and search in (almost) unconstrained still images," 2010.

[127] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 2014, pp. 3686–3693.

[128] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.

[129] S. ur Rehman, S. Tu, M. Waqas, Y. Huang, O. ur Rehman, B. Ahmad, and S. Ahmad, "Unsupervised pre-trained filter learning approach for efficient convolution neural network," *Neurocomputing*, vol. 365, pp. 171–190, 2019.

[130] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *CVPR 2011*, 2011, pp. 1465–1472.

[131] B. Sapp and B. Taskar, "Modec: Multimodal decomposable models for human pose estimation," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3674–3681.

[132] A. Cherian, J. Mairal, K. Alahari, and C. Schmid, "Mixing body-part sequences for human pose estimation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2361–2368.

[133] L. Sigal, A. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1, pp. 4–27, Mar. 2010.

[134] N. R. Howe, "A recognition-based motion capture baseline on the humaneva ii test data," *Machine Vision and Applications*, vol. 22, pp. 995–1008, 2011.

[135] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, jul 2014.

[136] M. Sun and S. Savarese, "Articulated part-based model for joint object detection and pose estimation," in *2011 International Conference on Computer Vision*, 2011, pp. 723–730.

[137] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[138] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2334–2343.

[139] E. Gärtner, A. Pirinen, and C. Sminchisescu, "Deep reinforcement learning for active human pose estimation," *arXiv preprint arXiv:2001.02024*, 2020.

[140] R. Alp Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7297–7306.

[141] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5693–5703.

[142] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7753–7762.

[143] D. Ramanan, "Learning to parse images of articulated bodies," in *Advances in neural information processing systems*, 2007, pp. 1129–1136.

[144] X. Nie, J. Feng, J. Xing, S. Xiao, and S. Yan, "Hierarchical contextual refinement networks for human pose estimation," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 924–936, 2018.

[145] N. Jammalamadaka, A. Zisserman, and C. Jawahar, "Human pose search using deep networks," *Image and Vision Computing*, vol. 59, pp. 31–43, 2017.

[146] B. Ai, Y. Zhou, Y. Yu, and S. Du, "Human pose estimation using deep structure guided learning," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2017, pp. 1224–1231.

[147] I. Marras, P. Palasek, and I. Patras, "Deep refinement convolutional networks for human pose estimation," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 446–453.

[148] W. Yang, W. Ouyang, H. Li, and X. Wang, "End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3073–3082.

[149] A. Bulat and G. Tzimiropoulos, "Human pose estimation via convolutional part heatmap regression," in *European Conference on Computer Vision*. Springer, 2016, pp. 717–732.

[150] L. Zhao, X. Gao, D. Tao, and X. Li, "A deep structure for human pose estimation," *Signal Processing*, vol. 108, pp. 36–45, 2015.

[151] K. Duan, D. Batra, and D. J. Crandall, "Human pose estimation via multi-layer composite models," *Signal Processing*, vol. 110, pp. 15–26, 2015.

[152] A. Cherian, J. Mairal, K. Alahari, and C. Schmid, "Mixing body-part sequences for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2353–2360.

[153] Z. Zhang, L. Hu, X. Deng, and S. Xia, "Weakly supervised adversarial learning for 3d human pose estimation from point clouds," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1851–1859, 2020.

[154] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7122–7131.

[155] Y. Huang, F. Bogo, C. Lassner, A. Kanazawa, P. V. Gehler, J. Romero, I. Akhter, and M. J. Black, "Towards accurate marker-less human shape and pose estimation over time," in *2017 international conference on 3D vision (3DV)*. IEEE, 2017, pp. 421–430.

[156] E. Jahangiri and A. L. Yuille, "Generating multiple diverse hypotheses for human 3d pose consistent with 2d joint detections," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 805–814.

[157] G. Rogez, P. Weinzaepfel, and C. Schmid, "Lcr-net: Localization-classification-regression for human pose," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1216–1224.

[158] V. Belagiannis and A. Zisserman, "Recurrent human pose estimation," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 468–475.

[159] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 1014–1021.

[160] X. Fan, K. Zheng, Y. Lin, and S. Wang, "Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1347–1355.

[161] Y. Yang, S. Baker, A. Kannan, and D. Ramanan, "Recognizing proxemics in personal photos," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3522–3529.

[162] X. Sun, B. Xiao, F. Wei, S. Liang, and Y. Wei, "Integral human pose regression," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 529–545.

[163] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 567–576.

[164] W. Liu, J. Chen, C. Li, C. Qian, X. Chu, and X. Hu, "A cascaded inception of inception network with attention modulated feature fusion for human pose estimation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[165] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[166] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 4724–4732.

[167] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, pp. 29–60.

[168] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

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

[170] L. Pishchulin, M. Andriluka, P. Gehler, and B. Schiele, "Poselet conditioned pictorial structures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 588–595.

[171] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 466–481.

[172] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards accurate multi-person pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4903–4911.

[173] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[174] Z. Su, M. Ye, G. Zhang, L. Dai, and J. Sheng, "Cascade feature aggregation for human pose estimation," *arXiv preprint arXiv:1902.07837*, 2019.

[175] A. A. Alani, "Arabic handwritten digit recognition based on restricted boltzmann machine and convolutional neural networks," *Information*, vol. 8, no. 4, p. 142, 2017.

[176] X. Yu, F. Zhou, and M. Chandraker, "Deep deformation network for object landmark localization," in *European Conference on Computer Vision*. Springer, 2016, pp. 52–70.

[177] S. Honari, P. Molchanov, S. Tyree, P. Vincent, C. Pal, and J. Kautz, "Improving landmark localization with semi-supervised learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1546–1555.

[178] U. Iqbal, P. Molchanov, T. Breuel Juergen Gall, and J. Kautz, "Hand pose estimation via latent 2.5 d heatmap regression," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 118–134.

[179] Y. Yu and F. Liu, "A two-stream deep fusion framework for high-resolution aerial scene classification," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[180] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *CVPR 2011*. IEEE, 2011, pp. 1465–1472.

[181] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele, "Deepcut: Joint subset partition and labeling for multi person pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4929–4937.

[182] I. Lifshitz, E. Fetaya, and S. Ullman, "Human pose estimation using deep consensus voting," in *European Conference on Computer Vision*. Springer, 2016, pp. 246–260.

[183] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4733–4742.

[184] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *European Conference on Computer Vision*. Springer, 2016, pp. 34–50.

[185] U. Rafi, B. Leibe, J. Gall, and I. Kostrikov, "An efficient convolutional network for human pose estimation." in *BMVC*, vol. 1, 2016, p. 2.

[186] Y. Chen, C. Shen, H. Chen, X.-S. Wei, L. Liu, and J. Yang, "Adversarial learning of structure-aware fully convolutional networks for landmark localization," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[187] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 648–656.

[188] J. Wagner, V. Fischer, M. Herman, and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks." in *ESANN*, vol. 587, 2016, pp. 509–514.

[189] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International journal of computer vision*, vol. 61, no. 1, pp. 55–79, 2005.

[190] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3d human pose annotations," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 1365–1372.

[191] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3d pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2500–2509.

[192] X. Zhou, Q. Huang, X. Sun, X. Xue, and Y. Wei, "Towards 3d human pose estimation in the wild: a weakly-supervised approach," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 398–407.

[193] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp. 1325–1339, 2013.

[194] L. Sigal, A. O. Balan, and M. J. Black, "Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International journal of computer vision*, vol. 87, no. 1-2, p. 4, 2010.

[195] F. Huang, A. Zeng, M. Liu, Q. Lai, and Q. Xu, "Deepfuse: An imu-aware network for real-time 3d human pose estimation from multi-view image," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 429–438.

[196] X. Chen, K.-Y. Lin, W. Liu, C. Qian, and L. Lin, "Weakly-supervised discovery of geometry-aware representation for 3d human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 895–10 904.

[197] M. Rayat Imtiaz Hossain and J. J. Little, "Exploiting temporal information for 3d pose estimation," *arXiv*, pp. arXiv–1711, 2017.

[198] H.-S. Fang, Y. Xu, W. Wang, X. Liu, and S.-C. Zhu, "Learning pose grammar to encode human body configuration for 3d pose estimation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[199] K. Lee, I. Lee, and S. Lee, "Propagating lstm: 3d pose estimation based on joint interdependency," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 119–135.

[200] G. Pavlakos, X. Zhou, and K. Daniilidis, "Ordinal depth supervision for 3d human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7307–7316.

[201] G. Gkioxari, A. Toshev, and N. Jaitly, "Chained predictions using convolutional neural networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 728–743.

[202] M. Rayat Imtiaz Hossain and J. J. Little, "Exploiting temporal information for 3d human pose estimation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 68–84.

[203] X. Zhang, Z. Tang, J. Hou, and Y. Hao, "3d human pose estimation via human structure-aware fully connected network," *Pattern Recognition Letters*, vol. 125, pp. 404–410, 2019.

[204] I. Habibie, W. Xu, D. Mehta, G. Pons-Moll, and C. Theobalt, "In the wild human pose estimation using explicit 2d features and intermediate 3d representations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 905–10 914.

[205] F. Moreno-Noguer, "3d human pose estimation from a single image via distance matrix regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2823–2832.

[206] D. Liu, Z. Zhao, X. Wang, Y. Hu, L. Zhang, and T. Huang, "Improving 3d human pose estimation via 3d part affinity fields," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1004–1013.

[207] K. Wang, L. Lin, C. Jiang, C. Qian, and P. Wei, "3d human pose machines with self-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[208] B. Tekin, A. Rozantsev, V. Lepetit, and P. Fua, "Direct prediction of 3d body poses from motion compensated sequences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 991–1000.

[209] M. Lin, L. Lin, X. Liang, K. Wang, and H. Cheng, "Recurrent 3d pose sequence machines," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 810–819.

[210] P. Verma, A. Sah, and R. Srivastava, "Deep learning-based multi-modal approach using rgb and skeleton sequences for human activity recognition," *Multimedia Systems*, vol. 26, no. 6, pp. 671–685, 2020.

[211] S. K. Tripathy and R. Srivastava, "A real-time two-input stream multi-column multi-stage convolution neural network (tis-mcms-cnn) for efficient crowd congestion-level analysis," *Multimedia Systems*, vol. 26, no. 5, pp. 585–605, 2020.

[212] L. Bo, C. Sminchisescu, A. Kanaujia, and D. Metaxas, "Fast algorithms for large scale conditional 3d prediction," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[213] G. Mori and J. Malik, "Recovering 3d human body configurations using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1052–1062, 2006.

[214] I. Katircioglu, B. Tekin, M. Salzmann, V. Lepetit, and P. Fua, "Learning latent representations of 3d human pose with deep neural networks," *International Journal of Computer Vision*, vol. 126, no. 12, pp. 1326–1341, 2018.

[215] A.-I. Popa, M. Zanfir, and C. Sminchisescu, "Deep multitask architecture for integrated 2d and 3d human sensing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6289–6298.

[216] J. C. Núñez, R. Cabido, J. F. Vélez, A. S. Montemayor, and J. J. Pantrigo, "Multiview 3d human pose estimation using improved least-squares and lstm networks," *Neurocomputing*, vol. 323, pp. 335–343, 2019.

[217] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[218] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[219] B. Tekin, P. Márquez-Neila, M. Salzmann, and P. Fua, "Learning to fuse 2d and 3d image cues for monocular body pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3941–3950.

[220] C. Hong, J. Yu, D. Tao, and M. Wang, "Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3742–3751, 2014.

[221] C. Hong, X. Chen, X. Wang, and C. Tang, "Hypergraph regularized autoencoder for image-based 3d human pose recovery," *Signal Processing*, vol. 124, pp. 132–140, 2016.

[222] S. Ershadi-Nasab, E. Noury, S. Kasaei, and E. Sanaei, "Multiple human 3d pose estimation from multiview images," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 15 573–15 601, 2018.

[223] P. Verma and R. Srivastava, "Three stage deep network for 3d human pose reconstruction by exploiting spatial and temporal data via its 2d pose," *Journal of Visual Communication and Image Representation*, vol. 71, p. 102866, 2020.

[224] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[225] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[226] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[227] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[228] G. Varol, I. Laptev, and C. Schmid, "Long-term temporal convolutions for action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1510–1517, 2018.

[229] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[230] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras, "Two-person interaction detection using body-pose features and multiple instance learning," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2012, pp. 28–35.

[231] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from rgbd images," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 842–849.

[232] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 9–14.

[233] C. Chen, R. Jafari, and N. Kehtarnavaz, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *2015 IEEE International conference on image processing (ICIP)*. IEEE, 2015, pp. 168–172.

[234] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley mhad: A comprehensive multimodal human action database," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 2013, pp. 53–60.

[235] J. Liu, A. Shahroudy, M. L. Perez, G. Wang, L.-Y. Duan, and A. K. Chichung, "Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[236] P. Khaire, P. Kumar, and J. Imran, "Combining cnn streams of rgb-d and skeletal data for human activity recognition," *Pattern Recognition Letters*, vol. 115, pp. 107–116, 2018.

[237] M. F. Bulbul, S. Islam, and H. Ali, "Human action recognition using mhi and shi based glac features and collaborative representation classifier," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 4, pp. 3385–3401, 2019.

[238] M. R. e Souza and H. Pedrini, "Motion energy image for evaluation of video stabilization," *The Visual Computer*, vol. 35, no. 12, pp. 1769–1781, 2019.

[239] F. Jiang, S. Zhang, S. Wu, Y. Gao, and D. Zhao, "Multi-layered gesture recognition with kinect," in *Gesture Recognition*. Springer, 2017, pp. 387–416.

[240] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, and Z. Tang, "Robust cnn-based gait verification and identification using skeleton gait energy image," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018, pp. 1–7.

[241] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568–576.

[242] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.

[243] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1725–1732.

[244] C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of depth, skeleton, and inertial data for human action recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 2712–2716.

[245] E. Escobedo and G. Camara, "A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes," in *2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*. IEEE, 2016, pp. 209–216.

[246] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-d posture data," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 586–597, 2015.

[247] E. K. Zavadskas, Z. Turskis, J. Antucheviciene, and A. Zakarevicius, "Optimization of weighted aggregated sum product assessment," *Elektronika ir elektrotechnika*, vol. 122, no. 6, pp. 3–6, 2012.

[248] M. Velasquez and P. T. Hester, "An analysis of multi-criteria decision making methods," *International journal of operations research*, vol. 10, no. 2, pp. 56–66, 2013.

[249] V. V. Dhanisetty, W. Verhagen, and R. Curran, "Multi-criteria weighted decision making for operational maintenance processes," *Journal of Air Transport Management*, vol. 68, pp. 152–164, 2018.

[250] C. Caetano, J. Sena, F. Brémond, J. A. Dos Santos, and W. R. Schwartz, "Skelemotion: A new representation of skeleton joint sequences based on motion information for 3d action recognition," in *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2019, pp. 1–8.

[251] Y. Annadani, D. L. Rakshith, and S. Biswas, "Sliding dictionary based sparse representation for action recognition," *CoRR*, vol. abs/1611.00218, 2016.

[252] M. F. Bulbul, Y. Jiang, and J. Ma, "Dmms-based multiple features fusion for human action recognition," *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, vol. 6, no. 4, pp. 23–39, 2015.

[253] N. E. D. Elmadany, Y. He, and L. Guan, "Human gesture recognition via bag of angles for 3d virtual city planning in cave environment," in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2016, pp. 1–5.

[254] Y. Zhu, W. Chen, and G. Guo, "Evaluating spatiotemporal interest point features for depth-based action recognition," *Image and Vision Computing*, vol. 32, no. 8, pp. 453–464, 2014.

[255] G. I. Parisi, C. Weber, and S. Wermter, "Self-organizing neural integration of pose-motion features for human action recognition," *Frontiers in neurorobotics*, vol. 9, p. 3, 2015.

[256] E. Cippitelli, E. Gambi, S. Spinsante, and F. Flórez-Revuelta, "Evaluation of a skeleton-based method for human activity recognition on a large-scale rgb-d dataset," 2016.

[257] J. Liu, A. Shahroudy, D. Xu, A. C. Kot, and G. Wang, "Skeleton-based action recognition using spatio-temporal lstm network with trust gates," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 3007–3021, 2017.

[258] H. Yang, D. Yan, L. Zhang, D. Li, Y. Sun, S. You, and S. J. Maybank, "Feedback graph convolutional network for skeleton-based action recognition," *arXiv preprint arXiv:2003.07564*, 2020.

[259] Q. Ke, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, "A new representation of skeleton sequences for 3d action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3288–3297.

[260] ——, "Learning clip representations for skeleton-based 3d action recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2842–2855, 2018.

List of Publications

List of papers published/accepted in journals:

1. P. Verma and R. Srivastava, “Three stage deep network for 3D human pose reconstruction by exploiting spatial and temporal data via its 2D pose,” in Journal of Visual Communication and Image Representation, Elsvier, vol. 71, pp. 102–866, 1 Aug. 2020. (SCIE-2.259)

2. P. Verma and R. Srivastava, “Two-stage multi-view deep network for 3D human pose reconstruction using images and its 2D joint heatmaps through enhanced stack-hourglass approach” in The Visual Computer, Springer, vol. 62, pp. 1-14, May. 2021. (SCIE-2.6)

3. P. Verma and R. Srivastava, “Deep learning-based multi-modal approach using RGB and skeleton sequences for human activity recognition,” in Multimedia Systems, vol. 25, pp. 671-685, Dec. 2020. (SCI-1.93)

4. P. Verma and R. Srivastava, ”Reconsideration of multi-stage deep network for human pose estimation”, in Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, Taylor and Fransic, pp. 1-13, Apr. 2021.(ESCI)

List of papers communicated in journals:

1. P. Verma and R. Srivastava, “3D Human pose estimation using estimated 2d joint locations and densely connected deep network” is under review in Multimedia Tools and Applications, Springer, 2021. (Impact Factor – 2.7).

2. Pratishtha Verma, R Srivastava, “A Survey on Human Pose Estimation using RGB image: techniques, datasets and challenges” Multimedia Systems, Springer, (Impact Factor – 1.93)

List of papers presented in conferences:

1. Pratishtha Verma, R Srivastava, “Detect and Estimate: Simple and Efficient baseline for Human Pose Estimation”, paper was presented in “ INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND SPEECH TECHNOLOGY (AIST 2019) @ Indira Gandhi Delhi Technical University for Women” and the extended version of this paper is currently in revision in “International Journal of Information Technology, Springer”. [scopus]