

## REFERENCES:

---

- [1] R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010.
- [2] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates BT - Pattern Analysis and Machine Intelligence, IEEE Transactions on," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, 2001.
- [3] I. Laptev, "On space-time interest points," *Int. J. Comput. Vis.*, vol. 64, no. 2–3, pp. 107–123, 2005.
- [4] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis," *ACM Comput. Surv.*, vol. 43, no. 3, pp. 1–43, 2011.
- [5] P. N. Bodor R, Jackson B, "Vision-Based Human Tracking and Activity Recognition," in *Proc. of 11th Mediterranean Conf. on Control and Automation*, 2003.
- [6] Lawtomated, "A.I. Technical: Machine vs Deep Learning," *Medium*, 2019. .
- [7] S. Samanta and B. Chanda, "Space-Time Facet Model for Human Activity Classification," vol. 16, no. 6, pp. 1525–1535, 2014.
- [8] K. Mozafari, J. A. Nasiri, N. M. Charkari, and S. Jalili, "Action recognition by local space-time features and least square twin SVM (LS-TSVM)," *Proc. - 1st Int. Conf. Informatics Comput. Intell. ICI 2011*, pp. 287–292, 2011.
- [9] M. S. Bella, R. Bhanumathi, and G. R. Suresh, "Human Action Recognition Using Local Space Time Features and Adaboost Svm," *Int. J. Res. Eng. Technol.*, vol. 3, no. 4, pp. 139–142, 2014.
- [10] S. Ma, L. Sigal, and S. Sclaroff, "Space-Time Tree Ensemble for Action Recognition - TechTalks.tv," pp. 5024–5032, 2015.
- [11] T. Lan, Y. Zhu, A. R. Zamir, and S. Savarese, "Action recognition by hierarchical mid-level action elements," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 4552–4560, 2015.
- [12] C. Chen and K. Grauman, "Efficient Activity Detection with Max-Subgraph Search," 2012.
- [13] E. Vig, M. Dorr, and D. Cox, "Space-variant descriptor sampling for action recognition based on saliency and eye movements," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7578 LNCS, no. PART 7, pp. 84–97, 2012.
- [14] C. Wang, Y. Wang, and A. L. Yuille, "An approach to pose-based action recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 915–922, 2013.
- [15] H. Wang *et al.*, "Action Recognition by Dense Trajectories To cite this version :," 2011.
- [16] X. Peng, "Exploring Motion Boundary based Sampling and Spatial-Temporal Context Descriptors for Action Recognition," *Bmvc2013*, no. 1, pp. 1–11, 2013.
- [17] Y. G. Jiang, Q. Dai, X. Xue, W. Liu, and C. W. Ngo, "Trajectory-based modeling of human actions with motion reference points," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7576 LNCS, no. PART 5, pp. 425–438, 2012.
- [18] M. Jain, H. Jegou, and P. Bouthemy, "Better exploiting motion for better action recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2555–2562, 2013.

- [19] B. Fernando, E. Gavves, M. José Oramas, A. Ghodrati, and T. Tuytelaars, “Modeling video evolution for action recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 5378–5387, 2015.
- [20] L. Wang and D. Suter, “Recognizing human activities from silhouettes: Motion subspace and factorial discriminative graphical model,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2007.
- [21] H. Zhao and Z. Liu, “Shape-Based Human Activity Recognition Using Edit Distance,” *Image Signal Process. 2009. CISP '09. 2nd Int. Congr.*, no. 2006, pp. 1–4, 2009.
- [22] N. Ikizler and P. Duygulu, “Histogram of oriented rectangles: A new pose descriptor for human action recognition,” *Image Vis. Comput.*, vol. 27, no. 10, pp. 1515–1526, 2009.
- [23] F. Niu and M. Abdel-Mottaleb, “View-Invariant Human Activity Recognition Based on Shape and Motion Features,” *Int. Symp. Multimed. Softw. Eng.*, pp. 546–556, 2004.
- [24] M. Chen, L. Gong, T. Wang, and Q. Feng, “Action recognition using lie algebrized gaussians over dense local spatio-temporal features,” *Multimed. Tools Appl.*, vol. 74, no. 6, pp. 2127–2142, 2015.
- [25] X. Wu, D. Xu, L. Duan, and J. Luo, “Action recognition using context and appearance distribution features,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 489–496, 2011.
- [26] L. M. Wang, Y. Qiao, and X. Tang, “Motionlets: Mid-level 3D parts for human motion recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2674–2681, 2013.
- [27] L. Jingen, L. Jiebo, and S. Mubarak, “Recognizing Realistic Actions from Videos &quot;in the Wild&quot;,” pp. 1–8, 2009.
- [28] D. Xing, X. Wang, and H. Lu, “Action Recognition Using Hybrid Feature Descriptor and VLAD Video Encoding,” *Accv*, pp. 1–15, 2014.
- [29] H. Qian, Y. Mao, W. Xiang, and Z. Wang, “Recognition of human activities using SVM multi-class classifier,” *Pattern Recognit. Lett.*, vol. 31, no. 2, pp. 100–111, 2010.
- [30] O. Kliper-Gross, Y. Gurovich, T. Hassner, and L. Wolf, “Motion interchange patterns for action recognition in unconstrained videos,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7577 LNCS, no. PART 6, pp. 256–269, 2012.
- [31] C. Chen, R. Jafari, and N. Kehtarnavaz, “Action recognition from depth sequences using depth motion maps-based local binary patterns,” *Proc. - 2015 IEEE Winter Conf. Appl. Comput. Vision, WACV 2015*, pp. 1092–1099, 2015.
- [32] S. M. M. Ahsan, J. K. Tan, H. Kim, and S. Ishikawa, “Histogram of spatio temporal local binary patterns for human action recognition,” *2014 Jt. 7th Int. Conf. Soft Comput. Intell. Syst. 15th Int. Symp. Adv. Intell. Syst.*, pp. 1007–1011, 2014.
- [33] A. K. S. Kushwaha, S. Srivastava, and R. Srivastava, “Multi-view human activity recognition based on silhouette and uniform rotation invariant local binary patterns,” *Multimed. Syst.*, vol. 23, no. 4, pp. 451–467, 2017.
- [34] V. Kellokumpu, G. Zhao, and M. Pietikäinen, “Recognition of human actions using texture descriptors,” *Mach. Vis. Appl.*, vol. 22, no. 5, pp. 767–780, 2011.
- [35] T. P. Nguyen, A. Manzanera, N. S. Vu, and M. Garrigues, “Revisiting LBP-based texture models for human action recognition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8259 LNCS, no. PART 2, pp. 286–293, 2013.

- [36] S. K. Dhar, M. M. Hasan, and S. A. Chowdhury, "Human activity recognition based on Gaussian mixture model and directive local binary pattern," *2016 2nd Int. Conf. Electr. Comput. Telecommun. Eng.*, no. December, pp. 1–4, 2016.
- [37] E. Chen, S. Zhang, and C. Liang, "Action Recognition Using Motion History Image and Static History Image-based Local Binary Patterns," *Int. J. Multimed. Ubiquitous Eng.*, vol. 12, no. 1, pp. 203–214, 2017.
- [38] M. Helmi and S. M. T. Almodarresi, "Human Activity Recognition Using a Fuzzy Inference System," pp. 0–5, 2009.
- [39] A. O. Extraction, "Fuzzy Rule Inference Based Human Activity Recognition," pp. 211–215, 2009.
- [40] M. Nii, K. Nakai, T. Fujita, and Y. Takahashi, "Action Estimation from Human Activity Monitoring Data using Soft Computing Approach," pp. 434–439, 2010.
- [41] P. T. Chinimilli, S. Redkar, and W. Zhang, "Human Activity Recognition Using Inertial Measurement Units and Smart Shoes," pp. 1462–1467, 2017.
- [42] R. Vezzani, M. Piccardi, and R. Cucchiara, "An efficient Bayesian framework for on-line action recognition," *IEEE Int. Conf. Image Process. ICIP*, pp. 3553–3556, 2009.
- [43] J. Liu, B. Kuipers, and S. Savarese, "Recognizing human actions by attributes," *Cypr*, pp. 3337–3344, 2011.
- [44] H. Pirsiavash and D. Ramanan, "Parsing videos of actions with segmental grammars," 2014.
- [45] A. K. Singh, "Rule based Human Activity Recognition for Surveillance System," pp. 28–29, 2012.
- [46] S. Okour, A. Maeder, and J. Basilakis, "An adaptive rule-based approach to classifying activities of daily living," pp. 5–8, 2015.
- [47] S. Sadanand and J. J. Corso, "e xt e T xt e e xt xt e xt e xt Action Bank : A High-Level Representation of Activity in Video," *Comput. Vis. Pattern Recognit. (CVPR), 2012 IEEE Conf.*, no. May, pp. 1234–1241, 2012.
- [48] A. Alfaro, D. Mery, and A. Soto, "Action Recognition in Video Using Sparse Coding and Relative Features," 2016.
- [49] J. Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 4694–4702, 2015.
- [50] A. Karpathy *et al.*, "Large-scale Video Classification with Convolutional Neural Networks," *Www*, 2014. [Online]. Available: <https://cs.stanford.edu/people/karpathy/deepvideo/>. [Accessed: 25-Mar-2019].
- [51] T. Guha, S. Member, and R. K. Ward, "Sequential deep learning for human action recognition," *Hum. Behav. Underst.*, no. November, pp. 29–39, 2011.
- [52] S. Ji, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition.," *Pami*, vol. 35, no. 1, pp. 221–31, 2013.
- [53] B. Singh, T. K. Marks, M. Jones, O. Tuzel, and M. Shao, "A Multi-stream Bi-directional Recurrent Neural Network for Fine-Grained Action Detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 1961–1970, 2016.
- [54] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, "ACTIVITYNET: Video Benchmark for Human Action Understanding," 2015.
- [55] M. Panwar *et al.*, "CNN based approach for activity recognition using a wrist-worn accelerometer," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 2438–2441, 2017.

- [56] W. Zhu, J. Hu, G. Sun, X. Cao, and Y. Qiao, "A Key Volume Mining Deep Framework for Action Recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. Mi, pp. 1991–1999, 2016.
- [57] H. Bilen, B. Fernando, E. Gavves, and A. Vedaldi, "Action Recognition with Dynamic Image Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8828, no. c, pp. 1–14, 2017.
- [58] M. Hasan and A. K. Roy-Chowdhury, "Continuous Learning of Human Activity Models Using Deep Nets," pp. 705–720, 2014.
- [59] Z. Luo, B. Peng, D.-A. Huang, A. Alahi, and L. Fei-Fei, "Unsupervised Learning of Long-Term Motion Dynamics for Videos," *arXiv Prepr. arXiv1701.01821*, 2017.
- [60] J. Monteiro, R. Granada, R. C. Barros, and F. Meneguzzi, "Deep neural networks for kitchen activity recognition," *2017 Int. Jt. Conf. Neural Networks*, pp. 2048–2055, 2017.
- [61] "Combing Multiple Sources of Knowledge in Deep CNNs for Action Recognition."
- [62] J. Lei, J. Zhang, G. Li, Q. Guo, and D. Tu, "Continuous action segmentation and recognition using hybrid convolutional neural network-hidden Markov model model," *IET Comput. Vis.*, vol. 10, no. 6, pp. 537–544, 2016.
- [63] T. Dobhal, V. Shitole, G. Thomas, and G. Navada, "Human Activity Recognition using Binary Motion Image and Deep Learning," *Procedia Comput. Sci.*, vol. 58, pp. 178–185, 2015.
- [64] E. P. Ijjina and C. K. Mohan, "Human Action Recognition based on Motion Capture Information using Fuzzy Convolution Neural Networks," pp. 2–7, 2014.
- [65] S. Yu, Y. Cheng, L. Xie, and S.-Z. Li, "Fully convolutional networks for action recognition," *IET Comput. Vis.*, vol. 11, no. 8, pp. 744–749, 2017.
- [66] V. Veeriah, N. Zhuang, and G.-J. Qi, "Differential Recurrent Neural Networks for Action Recognition," 2015.
- [67] L. Sun, K. Jia, D. Yeung, and B. E. Shi, "Human Action Recognition Using Factorized Spatio-Temporal Convolutional Networks," *2015 IEEE Int. Conf. Comput. Vis.*, pp. 4597–4605, 2015.
- [68] G. Ch, I. Laptev, C. Schmid, G. Ch, and G. Ch, "P-CNN : Pose-based CNN Features for Action Recognition To cite this version : P-CNN : Pose-based CNN Features for Action Recognition," *Iccv*, pp. 3218–3226, 2015.
- [69] L. Sun, K. Jia, T. H. Chan, Y. Fang, G. Wang, and S. Yan, "DL-SFA: Deeply-learned slow feature analysis for action recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2625–2632, 2014.
- [70] S. Shrestha, "Learning Deep Representations for Human Activity in Video," pp. 225–229.
- [71] A. Mehrjou, R. Hosseini, and B. N. Araabi, "Combining CNN streams of RGB-D and skeletal data for human activity recognition," *Pattern Recognit. Lett.*, pp. 2–8, 2018.
- [72] Z. Tu *et al.*, "Multi-stream CNN: Learning representations based on human-related regions for action recognition," *Pattern Recognit.*, vol. 79, pp. 32–43, 2018.
- [73] P. Wang, W. Li, Z. Gao, Y. Zhang, C. Tang, and P. Ogunbona, "Scene flow to action map: A new representation for RGB-D based action recognition with convolutional neural networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 416–425, 2017.
- [74] T. Guha, S. Member, and R. K. Ward, "Sequential deep learning for human action recognition," *Hum. Behav. Underst.*, pp. 29--39, 2011.

- [75] Y. Li, D. Shi, B. Ding, and D. Liu, "Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6067–6074, 2014.
- [76] M. S. Ryoo and L. Matthies, "Video-based convolutional neural networks for activity recognition from robot-centric videos," *Proc. SPIE 9837, Unmanned Syst. Technol. XVIII*, vol. Proc. SPIE, no. doi: 10.1117/12.2229531, p. 98370R, 2016.
- [77] B. Fernando, E. Gavves, M. Jose Oramas, A. Ghodrati, and T. Tuytelaars, "Rank Pooling for Action Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 773–787, 2017.
- [78] E. P. Ijjina and C. K. Mohan, "Human action recognition using action bank features and convolutional neural networks," *2014 Asian Conf. Comput. Vis.*, vol. 59, pp. 178–182, 2014.
- [79] L. Cheng, Y. Guan, Kecheng Zhu, and Yiyang Li, "Recognition of human activities using machine learning methods with wearable sensors," *2017 IEEE 7th Annu. Comput. Commun. Work. Conf.*, pp. 1–7, 2017.
- [80] P. Wang, W. Li, Z. Gao, C. Tang, and P. O. Ogunbona, "Depth Pooling Based Large-Scale 3-D Action Recognition with Convolutional Neural Networks," *IEEE Trans. Multimed.*, vol. 20, no. 5, pp. 1051–1061, 2018.
- [81] N. Jaouedi, N. Boujnah, and M. S. Bouhleb, "A new hybrid deep learning model for human action recognition," *J. King Saud Univ. - Comput. Inf. Sci.*, 2019.
- [82] J. Li, R. Wu, J. Zhao, and Y. Ma, "Convolutional Neural Networks ( CNN ) for Indoor Human Activity Recognition Using Ubisense System," *Proc. 29th Chinese Control Decis. Conf.*, pp. 2068–2072, 2017.
- [83] A. Jayabalan, T. , Harish Karunakaran, Shravan Murlidharan, and Shizume, "Dynamic Action Recognition: A convolutional neural network model for temporally organized joint location data," *arXiv:1612.06703 [cs.CV]*, no. May, pp. 0–0, 2017.
- [84] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional Two-Stream Network Fusion for Video Action Recognition," *IEEE Conf. Comput. Vis. Pattern Recognit.*, no. i, 2016.
- [85] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," *Adv. neural Inf. Process. Syst.*, pp. 568–576, 2014.
- [86] E. P. Ijjina and K. M. Chalavadi, "Human action recognition using genetic algorithms and convolutional neural networks," *Pattern Recognit.*, vol. 59, pp. 199–212, 2016.
- [87] G. Varol, I. Laptev, and C. Schmid, "Long-term Temporal Convolutions for Action Recognition," *arXiv Prepr. arXiv1604.04494*, pp. 1–8, 2016.
- [88] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "A.: Sequential deep learning for human action recognition," no. October 2016, 2011.
- [89] L. Mo, F. Li, Y. Zhu, and A. Huang, "Human physical activity recognition based on computer vision with deep learning model," *Conf. Rec. - IEEE Instrum. Meas. Technol. Conf.*, vol. 2016-July, 2016.
- [90] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, 2016.
- [91] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. O. Ogunbona, "Action Recognition from Depth Maps Using Deep Convolutional Neural Networks," *IEEE Trans. Human-Machine Syst.*, vol. 46, no. 4, pp. 498–509, 2016.
- [92] S. U. Park, J. H. Park, M. A. Al-Masni, M. A. Al-Antari, M. Z. Uddin, and T. S. Kim, "A Depth Camera-based Human Activity Recognition via Deep Learning

- Recurrent Neural Network for Health and Social Care Services,” *Procedia Comput. Sci.*, vol. 100, pp. 78–84, 2016.
- [93] A. Montes, A. Salvador, S. Pascual, and X. Giro-i-Nieto, “Temporal Activity Detection in Untrimmed Videos with Recurrent Neural Networks,” no. Lscvs 2016, pp. 1–5, 2016.
- [94] Z. Liu, C. Zhang, and Y. Tian, “3D-based Deep Convolutional Neural Network for action recognition with depth sequences,” *Image Vis. Comput.*, vol. 55, pp. 93–100, 2016.
- [95] V. Sonone, “Notes on deep learning-Advanced CNN,” *mc.ai.* .
- [96] M. Asadi-Aghbolaghi *et al.*, “A Survey on Deep Learning Based Approaches for Action and Gesture Recognition in Image Sequences,” *Proc. - 12th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2017 - 1st Int. Work. Adapt. Shot Learn. Gesture Underst. Prod. ASLAGUP 2017, Biometrics Wild, Bwild 2017, Heteroge.*, pp. 476–483, 2017.
- [97] S. Ramasamy Ramamurthy and N. Roy, “Recent trends in machine learning for human activity recognition-A survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, no. February, pp. 1–11, 2018.
- [98] S. Herath, M. Harandi, and F. Porikli, “Going deeper into action recognition: A survey,” *Image Vis. Comput.*, vol. 60, pp. 4–21, 2017.
- [99] M. Ramezani and F. Yaghmaee, “A review on human action analysis in videos for retrieval applications,” *Artif. Intell. Rev.*, vol. 46, no. 4, pp. 485–514, 2016.
- [100] N. Zhang, Y. Wang, and P. Yu, “A Review of Human Action Recognition in Video,” in *Proceedings - 17th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2018*, 2018, pp. 57–62.
- [101] L. Zelnik-Manor and M. Irani, “Event-based analysis of video,” *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*, no. 1229, pp. II-123-II-130, 2005.
- [102] C. Schüldt, I. Laptev, and B. Caputo, “Recognizing human actions: A local SVM approach,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 3, pp. 32–36, 2004.
- [103] C. Gu *et al.*, “AVA: A Video Dataset of Spatio-Temporally Localized Atomic Visual Actions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 6047–6056, 2018.
- [104] J. Carreira, E. Noland, A. Banki-Horvath, C. Hillier, and A. Zisserman, “A Short Note about Kinetics-600,” 2018.
- [105] Y. Yoshikawa, J. Lin, and A. Takeuchi, “STAIR Actions: A Video Dataset of Everyday Home Actions,” 2018.
- [106] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, “A survey of video datasets for human action and activity recognition,” *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, 2013.
- [107] D. F. Fouhey, W. C. Kuo, A. A. Efros, and J. Malik, “From Lifestyle Vlogs to Everyday Interactions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4991–5000, 2018.
- [108] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, “ActivityNet: A large-scale video benchmark for human activity understanding,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 961–970, 2015.
- [109] K. Soomro, A. R. Zamir, and M. Shah, “UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild,” *CoRR*, no. November, 2012.
- [110] W. Kay *et al.*, “The Kinetics Human Action Video Dataset,” *arXiv:1705.06950 [cs.CV]*, 2017.
- [111] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB: A large video

- database for human motion recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2556–2563, 2011.
- [112] Y. W. Chao, Z. Wang, Y. He, J. Wang, and J. Deng, “HICO: A benchmark for recognizing human-object interactions in images,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 1017–1025, 2015.
- [113] D.-T. Le, J. Uijlings, and R. Bernardi, “TUHOI: Trento Universal Human Object Interaction Dataset,” 2015, pp. 17–24.
- [114] S. Blunsden and R. B. Fisher, “The BEHAVE video dataset: ground truthed video for multi-person behavior classification,” *Ann. BMVA*, vol. 2010, no. 4, pp. 1–11, 2010.
- [115] M. Monfort *et al.*, “Moments in Time Dataset: one million videos for event understanding,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2019.
- [116] D. Moltisanti *et al.*, “Scaling Egocentric Vision: The Dataset,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11208 LNCS, pp. 753–771.
- [117] H. Kuehne, A. Iqbal, A. Richard, and J. Gall, “Mining YouTube - A dataset for learning fine-grained action concepts from webly supervised video data,” 2019.
- [118] S. Gella and F. Keller, “An analysis of action recognition datasets for language and vision tasks,” *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.)*, vol. 2, no. c, pp. 64–71, 2017.
- [119] K. Soomro and A. R. Zamir, “Computer Vision in Sports,” *Comput. Vis. Sport.*, 2014.
- [120] M. D. Rodriguez, J. Ahmed, and M. Shah, “Action MACH: A spatio-temporal maximum average correlation height filter for action recognition,” *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2008.
- [121] J. C. Niebles, C. W. Chen, and L. Fei-Fei, “Modeling temporal structure of decomposable motion segments for activity classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6312 LNCS, no. PART 2, pp. 392–405, 2010.
- [122] M. S. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori, “A Hierarchical Deep Temporal Model for Group Activity Recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 1971–1980, 2016.
- [123] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, “SoccerNet: A scalable dataset for action spotting in soccer videos,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2018-June, pp. 1792–1802, 2018.
- [124] V. Bloom, D. Makris, and V. Argyriou, “G3D: A gaming action dataset and real time action recognition evaluation framework,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 7–12, 2012.
- [125] V. Bloom, V. Argyriou, and D. Makris, “G3di: A gaming interaction dataset with a real time detection and evaluation framework,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8925, pp. 698–712, 2015.
- [126] W. Li, Z. Zhang, and Z. Liu, “Action recognition based on a bag of 3D points,” *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010*, vol. 2010, pp. 9–14, 2010.
- [127] L. Rybok, S. Friedberger, U. D. Hanebeck, and R. Stiefelhagen, “The KIT Robo-kitchen data set for the evaluation of view-based activity recognition systems,” in *IEEE-RAS International Conference on Humanoid Robots*, 2011, pp. 128–133.
- [128] F. De La Torre *et al.*, “Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database,” *C. Robot. Inst.*, no. April, p. 19, 2008.

- [129] M. Tenorth, J. Bandouch, and M. Beetz, “The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition,” in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, 2009, pp. 1089–1096.
- [130] Y. Li, M. Liu, and J. M. Rehg, “In the Eye of Beholder: Joint Learning of Gaze and Actions in First Person Video,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11209 LNCS, pp. 639–655, 2018.
- [131] V. Fontana, G. Singh, S. Akrigg, M. Di Maio, S. Saha, and F. Cuzzolin, “Action Detection from a Robot-Car Perspective,” 2018.
- [132] Oxford Mobile Robotics Group, “Oxford RobotCar Dataset,” *WWW*, 2016. [Online]. Available: <http://robotcar-dataset.robots.ox.ac.uk>. [Accessed: 25-Mar-2019].
- [133] R. B. Fisher, “The PETS04 surveillance ground-truth data sets,” *Eval. Track. Surveill.*, pp. 1–5, 2004.
- [134] T. List, J. Bins, J. Vazquez, and R. B. Fisher, “Performance evaluating the evaluator,” *Proc. - 2nd Jt. IEEE Int. Work. Vis. Surveill. Perform. Eval. Track. Surveillance, VS-PETS*, vol. 2005, pp. 129–136, 2005.
- [135] A. Nghiem *et al.*, “ETISEO , performance evaluation for video surveillance systems,” 2010.
- [136] X. Dai, B. Singh, G. Zhang, L. S. Davis, and Y. Q. Chen, “Temporal Context Network for Activity Localization in Videos,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, pp. 5727–5736.
- [137] Y. Liu, L. Nie, L. Liu, and D. S. Rosenblum, “From action to activity: Sensor-based activity recognition,” *Neurocomputing*, vol. 181, pp. 108–115, 2016.
- [138] J. Zhang, W. Li, P. O. Ogunbona, P. Wang, and C. Tang, “RGB-D-based action recognition datasets: A survey,” *Pattern Recognit.*, vol. 60, pp. 86–105, 2016.
- [139] R. Chavarriaga *et al.*, “The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition,” *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 2033–2042, 2013.
- [140] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, “Sensor-based datasets for human activity recognition - A systematic review of literature,” *IEEE Access*, vol. 6, pp. 59192–59210, 2018.
- [141] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional neural networks for human action recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, 2013.
- [142] T. Hassner, “A critical review of action recognition benchmarks,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 245–250.
- [143] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, “A database for fine grained activity detection of cooking activities,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, no. June, pp. 1194–1201, 2012.
- [144] H. Pirsiavash and D. Ramanan, “Detecting activities of daily living in first-person camera views,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012. [Online]. Available: <https://www.csee.umbc.edu/~hpirsiav/papers/ADLdataset/>. [Accessed: 10-Aug-2019].
- [145] P. Weinzaepfel, X. Martin, and C. Schmid, “Human Action Localization with Sparse Spatial Supervision,” 2016.



- [146] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, “Hollywood in homes: Crowdsourcing data collection for activity understanding,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 510–526, 2016.
- [147] R. Goyal *et al.*, “The ‘Something Something’ Video Database for Learning and Evaluating Visual Common Sense,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, no. March 2018, pp. 5843–5851.
- [148] C. R. de Souza, A. Gaidon, Y. Cabon, and A. M. L. Peña, “Procedural Generation of Videos to Train Deep Action Recognition Networks,” in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4757–4767.
- [149] G. Paolacci, J. Chandler, and P. Mueller, “Online Experimentation: Amazon Mechanical Turk,” *Four Essays Consum. Decis.*, vol. 5, no. 5, p. 59, 2012.
- [150] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, pp. 1–8, 2008.
- [151] N. Ikizler-cinbis, R. G. Cinbis, and S. Sclaroff, “Learning Actions From the Web,” no. Iccv, 2010.
- [152] A. Torralba and A. A. Efros, “Unbiased look at dataset bias,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016, no. 28, pp. 1521–1528, 2011.
- [153] G. A. Sigurdsson, O. Russakovsky, and A. Gupta, “What Actions are Needed for Understanding Human Actions in Videos?,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, pp. 2156–2165.
- [154] S. Nowozin and J. Shotton, “Action Points: A Representation for Low-latency Online Human Action Recognition,” *Microsoft Res. Cambridge Technical Rep.*, no. MSR-TR-2012-68, pp. 1–18, 2012.
- [155] F. Heidariwincheh, M. Mirmehdi, and D. Damen, “Action Completion: A Temporal Model for Moment Detection,” pp. 1–15, 2018.
- [156] H. Idrees *et al.*, “The THUMOS challenge on action recognition for videos ‘in the wild,’” *Comput. Vis. Image Underst.*, vol. 155, pp. 1–23, 2017.
- [157] C. Liu, Y. Hu, Y. Li, S. Song, and J. Liu, “PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding,” 2017.
- [158] P. Bojanowski *et al.*, “Weakly Supervised Action Labeling in Videos under Ordering Constraints Lecture Notes in Computer Science,” *Comput. Vis. – ECCV 2014*, vol. 8693, no. Chapter 41, pp. 628–643, 2014.
- [159] D. Minnen, T. L. Westeyn, T. Starner, J. a Ward, and P. Lukowicz, “Performance Metrics and Evaluation Issues for Continuous Activity Recognition,” *Proc. Int. Work. Perform. Metrics Intell. Syst.*, pp. 141–148, 2006.
- [160] R. Krishna, K. Hata, F. Ren, L. Fei-Fei, and J. C. Niebles, “Dense-Captioning Events in Videos,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2017-October, pp. 706–715, 2017.
- [161] O. Kliper-Gross, T. Hassner, and L. Wolf, “The action similarity labeling challenge,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 615–621, 2012.
- [162] L. Cao, Z. Liu, and T. S. Huang, “Cross-dataset action detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1998–2005, 2010.
- [163] L. Zelnik-Manor and M. Irani, “Event-based analysis of video,” *WWW*, 2005. [Online]. Available:

- <http://www.wisdom.weizmann.ac.il/~vision/VideoAnalysis/Demos/EventDetection/EventDetection.html>. [Accessed: 25-Mar-2019].
- [164] Ivan Laptev and Barbara Caputo, "Recognition of human actions," 2005. [Online]. Available: <http://www.nada.kth.se/cvap/actions/>. [Accessed: 25-Mar-2019].
- [165] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007.
- [166] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes," 2007. [Online]. Available: <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>. [Accessed: 25-Mar-2019].
- [167] CAVIAR, "CAVIAR: Context Aware Vision using Image-based Active Recognition, European Commission project IST 2001 37540," *WWW*, 2001. [Online]. Available: <http://groups.inf.ed.ac.uk/vision/CAVIAR/>. [Accessed: 25-Mar-2019].
- [168] A. T. Nghiem, F. Bremond, M. Thonnat, and R. Ma, "A new evaluation approach for video processing algorithms," *2007 IEEE Work. Motion Video Comput. WMVC 2007*, 2007.
- [169] A. T. Nghiem, F. Bremond, M. Thonnat, and R. Ma, "ETISEO dataset," *WWW*, 2007. [Online]. Available: <https://www-sop.inria.fr/orion/ETISEO/>. [Accessed: 25-Mar-2019].
- [170] R. Vezzani, R. Cucchiara, and I. Dipartimento, "ViSOR: Video Surveillance Online Repository," vol. 2010, no. 2, pp. 1–13, 2010.
- [171] R. Vezzani and R. Cucchiara, "Video surveillance online repository (ViSOR)," *WWW*, 2013. [Online]. Available: <http://www.openvisor.org/index.asp>. [Accessed: 25-Mar-2019].
- [172] "INRIA Xmas Motion Acquisition Sequences (IXMAS)," *WWW*, 2006. [Online]. Available: <http://4drepository.inrialpes.fr/public/viewgroup/>. [Accessed: 25-Mar-2019].
- [173] D. Z. S. Professor Tieniu Tan, "Center for Biometrics and Security Research," *WWW*. [Online]. Available: <http://www.cbsr.ia.ac.cn/china/Iris Databases CH.asp>. [Accessed: 25-Mar-2019].
- [174] "UCF Aerial Action Data Set," *WWW*, 2009. [Online]. Available: [http://cvcv.ucf.edu/data/UCF\\_Aerial\\_Action.php](http://cvcv.ucf.edu/data/UCF_Aerial_Action.php). [Accessed: 25-Mar-2019].
- [175] M. Rodriguez, Mikel D. Ahmed, Javed, Shah, "UCF-ARG Data Set," *WWW*, 2011. [Online]. Available: <http://cvcv.ucf.edu/data/UCF-ARG.php>. [Accessed: 25-Mar-2019].
- [176] A. R. Z. Khurram Soomro, "UCF Sports Action Data Set," *WWW*, 2008. [Online]. Available: [http://cvcv.ucf.edu/data/UCF\\_Sports\\_Action.php](http://cvcv.ucf.edu/data/UCF_Sports_Action.php).
- [177] R. Fu, Y. Song, and W. Zhao, "Human Activity Recognition with Smartphones," *Proc. 10th Eur. Conf. Comput. Vis. - ECCV '08*, no. Section 2, pp. 548–561, 2008.
- [178] D. Tran and A. Sorokin, "Human activity recognition with metric learning," *WWW*, 2008. [Online]. Available: <http://vision.cs.uiuc.edu/projects/activity/>. [Accessed: 25-Mar-2019].
- [179] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, "The i3DPost multi-view and 3D human action/interaction database," *CVMP 2009 - 6th Eur. Conf. Vis. Media Prod.*, pp. 159–168, 2009.
- [180] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, "i3DPost Multi-view Human Action Datasets," *WWW*, 2009. [Online]. Available: [http://kahlan.eps.surrey.ac.uk/i3dpost\\_action/](http://kahlan.eps.surrey.ac.uk/i3dpost_action/). [Accessed: 25-Mar-2019].
- [181] R. Messing, C. Pal, and H. Kautz, "Activity recognition using the velocity histories

- of tracked keypoints,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 104–111, 2009.
- [182] R. Messing, C. Pal, and H. Kautz, “University of Rochester Activities of Daily Living Dataset,” *Www*, 2009. [Online]. Available: <http://www.cs.rochester.edu/~rmessing/uradl/>. [Accessed: 25-Mar-2019].
- [183] W. Choi, K. Shahid, and S. Savarese, “What are they doing?: Collective activity classification using spatio-temporal relationship among people,” *2009 IEEE 12th Int. Conf. Comput. Vis. Work. ICCV Work. 2009*, vol. 24, pp. 1282–1289, 2009.
- [184] W. Choi, K. Shahid, and S. Savarese, “Collective Activity Dataset .,” *Www*, 2009. [Online]. Available: <http://vhosts.eecs.umich.edu/vision/activity-dataset.html>. [Accessed: 25-Mar-2019].
- [185] B. Project, “Computer-assisted prescreening of video streams for unusual activities,” *Www*, 2004. [Online]. Available: <http://homepages.inf.ed.ac.uk/rbf/BEHAVE/>. [Accessed: 25-Mar-2019].
- [186] S. Singh, S. A. Velastin, and H. Ragheb, “MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods,” in *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 2010, pp. 48–55.
- [187] F. Murtaza, M. H. Yousaf, and S. A. Velastin, “Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description,” *IET Comput. Vis.*, vol. 10, no. 7, pp. 758–767, 2016.
- [188] S. Singh, S. A. Velastin, and H. Ragheb, “MuHAVi: Multicamera Human Action Video Data.” [Online]. Available: <http://dipersec.king.ac.uk/MuHAVi-MAS/>. [Accessed: 25-Mar-2019].
- [189] M. S. Ryoo, J. K. Aggarwal, C. Chen, and A. Roy-chowdhury, “An Overview of Contest on Semantic Description of Human Activities (SDHA),” *ICPR contests*. 2010.
- [190] M. S. Ryoo and J. K. Aggarwal, “Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities,” *Proc. IEEE Int. Conf. Comput. Vis.*, no. Iccv, pp. 1593–1600, 2009.
- [191] M. S. Ryoo, J. K. Aggarwal, C. Chen, and A. Roy-chowdhury, “ICPR 2010 Contest on Semantic Description of Human Activities (SDHA 2010),” *Www*, 2010. [Online]. Available: <http://cvrc.ece.utexas.edu/SDHA2010/>. [Accessed: 25-Mar-2019].
- [192] C. Chen and J. K. Aggarwal, “Recognizing Human Action from a Far Field of View The University of Texas at Austin,” *Analysis*. 2009.
- [193] C.-C. Chen, M. S. Ryoo, and J. K. Aggarwal, “UT-Tower Dataset: Aerial View Activity Classification Challenge.” 2010.
- [194] S. Oh *et al.*, “AVSS 2011 demo session: A large-scale benchmark dataset for event recognition in surveillance video,” *2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011*, no. 2, pp. 527–528, 2011.
- [195] S. Oh *et al.*, “VIRAT Video Dataset,” *Www*, 2011. [Online]. Available: <http://www.viratdata.org/>. [Accessed: 25-Mar-2019].
- [196] D. Gehrig *et al.*, “Combined intention, activity, and motion recognition for a humanoid household robot,” in *IEEE International Conference on Intelligent Robots and Systems*, 2011, pp. 4819–4825.
- [197] D. Gehrig *et al.*, “The Karlsruhe Motion, Intention, and Activity Data set (MINTA),” *Www*, 2011. [Online]. Available: <https://cvhci.anthropomatik.kit.edu/~lrybok/projects/minta/>. [Accessed: 25-Mar-2019].
- [198] L. Rybok, S. Friedberger, U. D. Hanebeck, and R. Stiefelhagen, “The KIT Robo-

- Kitchen Activity Data Set,” *Www*, 2011. [Online]. Available: <https://cvhci.anthropomatik.kit.edu/~lrybok/projects/kitchen/>. [Accessed: 25-Mar-2019].
- [199] L. Ivan, M. Marszałek, C. Schmid, and B. Rozenfeld, “IRISA/INRIA Rennes France : Learning Human Actions from Movies,” *Www*, 2009. [Online]. Available: <http://www.irisa.fr/vista/actions/hollywood>. [Accessed: 25-Mar-2019].
- [200] M. Marszałek, I. Laptev, and C. Schmid, “Actions in context,” *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, vol. 2009 IEEE, no. i, pp. 2929–2936, 2009.
- [201] L. Ivan, M. Marszałek, C. Schmid, and B. Rozenfeld, “IRISA/INRIA Rennes France : Learning Human Actions from Movies,” *Www*, 2008. [Online]. Available: <http://www.irisa.fr/vista/actions/hollywood2>. [Accessed: 25-Mar-2019].
- [202] J. Liu, J. Luo, and M. Shah, “Recognizing realistic actions from videos in the Wild,” *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, vol. 2009 IEEE, pp. 1996–2003, 2009.
- [203] J. L. and M. S. Jingen Liu, “UCF YouTube Action Data Set,” *Www*, 2009. [Online]. Available: [http://crcv.ucf.edu/data/UCF\\_YouTube\\_Action.php](http://crcv.ucf.edu/data/UCF_YouTube_Action.php). [Accessed: 25-Mar-2019].
- [204] A. Patron-Perez, M. Marszałek, A. Zisserman, and I. Reid, “High Five: Recognising human interactions in TV shows,” *Br. Mach. Vis. Conf. BMVC 2010 - Proc.*, pp. 50.1-50.11, 2010.
- [205] A. Patron-Perez, “TV Human Interaction Dataset,” *Www*, 2010. [Online]. Available: [http://www.robots.ox.ac.uk/~alonso/tv\\_human\\_interactions.html](http://www.robots.ox.ac.uk/~alonso/tv_human_interactions.html). [Accessed: 25-Mar-2019].
- [206] “Olympic sports dataset,” *Stanford University*, 2010. [Online]. Available: <http://vision.stanford.edu/Datasets/OlympicSports/>.
- [207] K. K. Reddy and M. Shah, “Recognizing 50 human action categories of web videos,” *Mach. Vis. Appl.*, vol. 24, no. 5, pp. 971–981, 2013.
- [208] K. K. Reddy and M. Shah, “UCF50 - Action Recognition Data Set,” *Www*, 2010. [Online]. Available: <https://www.crcv.ucf.edu/data/UCF50.php>. [Accessed: 25-Mar-2019].
- [209] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB : A Large Human Motion Database,” *Www*, 2014. [Online]. Available: <http://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/>. [Accessed: 25-Mar-2019].
- [210] A. R. Z. and M. S. Khurram Soomro, “UCF101 - Action Recognition Data Set,” *Www*, 2012. [Online]. Available: <http://crcv.ucf.edu/data/UCF101.php>. [Accessed: 25-Mar-2019].
- [211] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, “MPII Cooking Activities Dataset,” *Www*, 2013. [Online]. Available: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/human-activity-recognition/mpii-cooking-activities-dataset/>. [Accessed: 25-Mar-2019].
- [212] M. Rohrbach, M. Regneri, M. Andriluka, S. Amin, M. Pinkal, and B. Schiele, “Script data for attribute-based recognition of composite activities,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7572 LNCS, no. PART 1, pp. 144–157, 2012.
- [213] M. Rohrbach, M. Regneri, M. Andriluka, S. Amin, M. Pinkal, and B. Schiele, “MPII Cooking Composite Activities Dataset,” *Www*, 2012. [Online]. Available: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/human-activity-recognition/mpii-cooking-activities-dataset/>.

- [Accessed: 10-Aug-2019].
- [214] O. Kliper-gross and T. Hassner, “The Action Similarity Labeling Challenge,” *Www*, 2012. [Online]. Available: <http://www.openu.ac.il/home/hassner/data/ASLAN/ASLAN.html>. [Accessed: 25-Mar-2019].
- [215] P. Mettes, J. C. van Gemert, and C. G. M. Snoek, “Spot on: Action localization from pointly-supervised proposals,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9909 LNCS, no. Mil, pp. 437–453, 2016.
- [216] P. Mettes, J. C. van Gemert, and C. G. M. Snoek, “Hollywood2Tubes,” *Www*, 2016. [Online]. Available: <https://staff.fnwi.uva.nl/p.s.m.mettes/codedata.html>. [Accessed: 25-Mar-2019].
- [217] P. Das, C. Xu, R. F. Doell, and J. J. Corso, “A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2634–2641, 2013.
- [218] P. Das, C. Xu, R. F. Doell, and J. J. Corso, “U of Michigan, YouCook: An Annotated Data Set of Unconstrained Third-Person Cooking Videos,” *Www*, 2013. [Online]. Available: <http://web.eecs.umich.edu/~jjcorso/r/youcook/>. [Accessed: 25-Mar-2019].
- [219] Y.-G. Jiang *et al.*, “THUMOS: ECCV Workshop on Action Recognition with a Large Number of Classes,” 2014.
- [220] Y.-G. Jiang *et al.*, “THUMOS 13: The First International Workshop on Action Recognition with a Large Number of Classes, in conjunction with ICCV ’13, Sydney, Australia,” *Www*, 2013. [Online]. Available: <http://csrc.ucf.edu/ICCV13-Action-Workshop/>. [Accessed: 25-Mar-2019].
- [221] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, “Towards understanding action recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3192–3199, 2013.
- [222] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, “Joint-annotated Human Motion Data Base,” *Www*, 2013. [Online]. Available: <http://jhmdb.is.tue.mpg.de>. [Accessed: 25-Mar-2019].
- [223] H. Kuehne, A. Arslan, and T. Serre, “The language of actions: Recovering the syntax and semantics of goal-directed human activities,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 780–787, 2014.
- [224] H. Kuehne, A. Arslan, and T. Serre, “The Breakfast Actions Dataset,” *Www*, 2014. [Online]. Available: <http://serre-lab.clps.brown.edu/resource/breakfast-actions-dataset/>. [Accessed: 10-Aug-2019].
- [225] Y.-G. Jiang *et al.*, “THUMOS Challenge 2014,” *Www*, 2014. [Online]. Available: <http://csrc.ucf.edu/THUMOS14/>. [Accessed: 25-Mar-2019].
- [226] A. Karpathy and T. Leung, “Large-scale Video Classification with Convolutional Neural Networks,” *Www*, 2014. [Online]. Available: <https://cs.stanford.edu/people/karpathy/deepvideo/>. [Accessed: 25-Mar-2019].
- [227] H. Idrees *et al.*, “THUMOS Challenge 2015 In conjunction with CVPR’15,” *Www*, 2015. [Online]. Available: <http://www.thumos.info/home.html>. [Accessed: 25-Mar-2019].
- [228] K. Lee, D. Ognibene, H. J. Chang, T. K. Kim, and Y. Demiris, “STARE: Spatiooral Attention Relocation for Multiple Structured Activities Detection,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5916–5927, 2015.
- [229] K. Lee, D. Ognibene, H. J. Chang, T. K. Kim, and Y. Demiris, “Crêpe Dataset,” *Www*, 2013. [Online]: <https://osf.io/d5k38/wiki/home/>. [Accessed: 25-Mar-2019].

- [230] S. M. Safdarnejad, X. Liu, L. Udpa, B. Andrus, J. Wood, and D. Craven, “Sports Videos in the Wild (SVW): A video dataset for sports analysis,” *2015 11th IEEE Int. Conf. Work. Autom. Face Gesture Recognition, FG 2015*, 2015.
- [231] S. M. Safdarnejad, X. Liu, L. Udpa, B. Andrus, J. Wood, and D. Craven, “Sports Videos in the Wild (SVW),” *Www*, 2015. [Online]. Available: <http://cvlab.cse.msu.edu/project-svw.html>. [Accessed: 25-Mar-2019].
- [232] F. C. Heilbron, V. Escorcia, B. Ghanem, and J. C. Niebles, “ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding,” *Www*, 2015. [Online]. Available: <http://www.activity-net.org>. [Accessed: 25-Mar-2019].
- [233] M. Rohrbach *et al.*, “Recognizing Fine-Grained and Composite Activities Using Hand-Centric Features and Script Data,” *Int. J. Comput. Vis.*, vol. 119, no. 3, pp. 346–373, 2016.
- [234] M. Rohrbach *et al.*, “MPII Cooking 2 Dataset,” *Www*, 2015. [Online]. Available: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/human-activity-recognition/mpii-cooking-2-dataset/>. [Accessed: 10-Aug-2019].
- [235] B. Singh, T. K. Marks, M. Jones, O. Tuzel, and M. Shao, “MERL Shopping Dataset,” *Www*, 2016. [Online]. Available: <http://www.merl.com/demos/merl-shopping-dataset>. [Accessed: 25-Mar-2019].
- [236] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev, and A. Gupta, “Charades and Charades-Ego Datasets,” *Www*, 2016. [Online]. Available: <http://allenai.org/plato/charades/>. [Accessed: 25-Mar-2019].
- [237] G. Mori, M. Andriluka, O. Russakovsky, N. Jin, L. Fei-Fei, and S. Yeung, “Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos,” *Int. J. Comput. Vis.*, vol. 126, no. 2–4, pp. 375–389, 2017.
- [238] G. Mori, M. Andriluka, O. Russakovsky, N. Jin, L. Fei-Fei, and S. Yeung, “Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos,” *Www*, 2018. [Online]. Available: <http://ai.stanford.edu/~syyeung/everymoment.html>. [Accessed: 25-Mar-2019].
- [239] R. Goyal *et al.*, “The 20BN Something Something Dataset,” *Www*, 2017. [Online]. Available: <https://20bn.com/datasets/something-something>. [Accessed: 10-Aug-2019].
- [240] P. Weinzaepfel, X. Martin, and C. Schmid, “Daily Action Localization in Youtube videos,” *Www*, 2017. [Online]. Available: <http://pascal.inrialpes.fr/data/daly/>. [Accessed: 25-Mar-2019].
- [241] C. Gu *et al.*, “AVA Actions Dataset,” *Www*, 2017. [Online]. Available: <https://research.google.com/ava/>. [Accessed: 25-Mar-2019].
- [242] C. Xu, S. H. Hsieh, C. Xiong, and J. J. Corso, “Can humans fly? Action understanding with multiple classes of actors,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 2264–2273, 2015.
- [243] C. Xu, S. H. Hsieh, C. Xiong, and J. J. Corso, “A2D: A Dataset and Benchmark for Action Recognition and Segmentation with Multiple Classes of Actors,” *Www*, 2015. [Online]. Available: <http://web.eecs.umich.edu/~jjcorso/r/a2d/>. [Accessed: 25-Mar-2019].
- [244] W. Kay *et al.*, “Kinetics Datasets,” *Www*, 2017. [Online]. Available: <http://deepmind.com/kinetics>. [Accessed: 25-Mar-2019].
- [245] D. F. Fouhey, W. Kuo, A. A. Efros, and J. Malik, “From Lifestyle VLOGs to Everyday Interactions: The VLOG Dataset,” *Www*, 2017. [Online]. Available: <http://web.eecs.umich.edu/~fouhey/2017/VLOG/>. [Accessed: 10-Aug-2019].
- [246] L. Zhou, C. Xu, and J. J. Corso, “Towards automatic learning of procedures from

- web instructional videos,” in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2018, pp. 7590–7598.
- [247] L. Zhou, C. Xu, and J. J. Corso, “YouCook2 Dataset,” *Www*, 2018. [Online]. Available: <http://youcook2.eecs.umich.edu/>.
- [248] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, “SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos,” *Www*, 2018. [Online]. Available: <https://silviogiancola.github.io/SoccerNet>. [Accessed: 25-Mar-2019].
- [249] A. J. Piergiovanni and M. S. Ryoo, “Fine-grained activity recognition in baseball videos,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2018-June, pp. 1821–1830, 2018.
- [250] A. J. Piergiovanni and M. S. Ryoo, “MLB-YouTube Dataset,” *Www*, 2018. [Online]. Available: <https://github.com/piergiaj/mlb-youtube/>. [Accessed: 25-Mar-2019].
- [251] Y. Yoshikawa, J. Lin, and A. Takeuchi, “STAIR Actions: A Large-Scale Video Dataset of Everyday Human Actions,” *Www*, 2018. [Online]. Available: <http://actions.stair.center>. [Accessed: 25-Mar-2019].
- [252] M. Shim, Y. H. Kim, K. Kim, and S. J. Kim, “Teaching machines to understand baseball games: Large-scale baseball video database for multiple video understanding tasks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11219 LNCS, pp. 420–437, 2018.
- [253] M. Shim, Y. H. Kim, K. Kim, and S. J. Kim, “Teaching Machines to Understand Baseball Games: Large-Scale Baseball Video Database for Multiple Video Understanding Tasks,” *Www*, 2018. [Online]. Available: <https://sites.google.com/site/eccv2018bbdb/>. [Accessed: 25-Mar-2019].
- [254] M. Monfort *et al.*, “Moments in Time Dataset: A large-scale dataset for recognizing and understanding action in videos,” *Www*, 2019. [Online]. Available: <http://moments.csail.mit.edu>. [Accessed: 25-Mar-2019].
- [255] H. Zhao, Z. Yan, L. Torresani, and A. Torralba, “HACS: Human Action Clips and Segments Dataset for Recognition and Temporal Localization,” 2017.
- [256] H. Zhao, Z. Yan, L. Torresani, and A. Torralba, “HACS: Human Action Clips and Segments Dataset for Recognition and Temporal Localization,” *Www*, 2019. [Online]. Available: <http://hacs.csail.mit.edu>. [Accessed: 25-Mar-2019].
- [257] Y. Tang *et al.*, “COIN: A Large-scale Dataset for Comprehensive Instructional Video Analysis,” pp. 1207–1216, 2019.
- [258] Y. Tang *et al.*, “COIN: A Large-scale Dataset for Comprehensive Instructional Video Analysis,” *Www*, 2019. [Online]. Available: <https://coin-dataset.github.io/>. [Accessed: 10-Aug-2019].
- [259] A. Fathi, X. Ren, and J. Rehg, “Georgia Tech Egocentric Activity Datasets,” *Www*, 2011. [Online]. Available: <http://www.cbi.gatech.edu/fpv/>. [Accessed: 10-Aug-2019].
- [260] A. Fathi, Y. Li, and J. M. Rehg, “Learning to recognize daily actions using gaze,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7572 LNCS, no. PART 1, pp. 314–327, 2012.
- [261] A. Calway, W. Mayol-Cuevas, D. Damen, O. Haines, and T. Leelasawassuk, “Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video,” *Bmvc*, pp. 30.1-30.13, 2015.
- [262] A. C. and W. M.-C. Dima Damen, Teesid Leelasawassuk, Osian Haines, Michael Wray, Davide Moltisanti, “Bristol Egocentric Object Interactions Dataset,” *Www*, 2014. [Online]. <http://data.bris.ac.uk/data/dataset/o4hx7jnmfqt01lyzf2n4rchg6>. [Accessed: 10-Aug-2019].

- [263] Y. Iwashita, A. Takamine, R. Kurazume, and M. S. Ryoo, “First-person animal activity recognition from egocentric videos,” *Proc. - Int. Conf. Pattern Recognit.*, no. i, pp. 4310–4315, 2014.
- [264] Y. Iwashita, A. Takamine, R. Kurazume, and M. S. Ryoo, “DogCentric Activity Dataset,” *Www*, 2014. [Online]. Available: [http://robotics.ait.kyushu-u.ac.jp/~yumi/db/first\\_dog.html](http://robotics.ait.kyushu-u.ac.jp/~yumi/db/first_dog.html). [Accessed: 10-Aug-2019].
- [265] Y. Li, Z. Ye, and J. M. Rehg, “Delving into egocentric actions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, no. 1, pp. 287–295, 2015.
- [266] K. Ohnishi, A. Kanehira, A. Kanazaki, and T. Harada, “Recognizing Activities of Daily Living with a Wrist-Mounted Camera,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, no. c, pp. 3103–3111, 2016.
- [267] K. Ohnishi, A. Kanehira, A. Kanazaki, and T. Harada, “Recognizing Activities of Daily Living with a Wrist-mounted Camera,” *Www*, 2015. [Online]. Available: <https://www.mi.t.u-tokyo.ac.jp/static/projects/miladl/>. [Accessed: 25-Mar-2019].
- [268] S. Singh, C. Arora, and C. V. Jawahar, “Trajectory aligned features for first person action recognition,” *Pattern Recognit.*, vol. 62, pp. 45–55, 2017.
- [269] S. Singh, C. Arora, and C. V. Jawahar, “PR 2016 Paper - Trajectory Aligned Features For First Person Action Recognition,” *Www*, 2017. [Online]. Available: <http://cvit.iit.ac.in/research/projects/cvit-projects/first-person-action-recognition>. [Accessed: 10-Aug-2019].
- [270] G. A. Sigurdsson, A. Gupta, C. Schmid, A. Farhadi, and K. Alahari, “Charades-Ego: A Large-Scale Dataset of Paired Third and First Person Videos,” pp. 1–3, 2018.
- [271] D. Moltisanti *et al.*, “EPIC-Kitchens 2019,” *Www*, 2019. [Online]. Available: <https://epic-kitchens.github.io/2018>. [Accessed: 25-Mar-2019].
- [272] G. Abebe, A. Catala, and A. Cavallaro, “A first-person vision dataset of office activities,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11377 LNAI, pp. 27–37.
- [273] G. Abebe, A. Catala, and A. Cavallaro, “A first-person vision dataset of office activities,” *www*, 2019. [Online]. Available: <http://www.eecs.qmul.ac.uk/~andrea/fpvo.html>. [Accessed: 10-Aug-2019].
- [274] V. Delaitre, I. Laptev, and J. Sivic, “Recognizing human actions in still images: A study of bag-of-features and part-based representations,” in *British Machine Vision Conference, BMVC 2010 - Proceedings*, 2010, pp. 97.1-97.11.
- [275] G. Guo and A. Lai, “A survey on still image based human action recognition,” *Pattern Recognit.*, vol. 47, no. 10, pp. 3343–3361, 2014.
- [276] V. Delaitre, L. Ivan, and J. Sivic, “Human Action Classification in Still Images,” *Www*, 2010. [Online]. Available: <https://www.di.ens.fr/willow/research/stillactions>. [Accessed: 25-Mar-2019].
- [277] B. Yao and L. Fei-Fei, “Grouplet: A structured image representation for recognizing human and object interactions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 9–16, 2010.
- [278] B. Yao, F. F. Li, and A. Khosla, “People Playing Musical Instrument (PPMI),” *Www*, 2010. [Online]. Available: <http://ai.stanford.edu/~bangpeng/ppmi.html>. [Accessed: 25-Mar-2019].
- [279] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei, “Human action recognition by learning bases of action attributes and parts,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1331–1338, 2011.



- [280] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-fei, “Stanford 40 Actions dataset,” *Www*, 2011. [Online]. Available: <http://vision.stanford.edu/Datasets/40actions.html>. [Accessed: 25-Mar-2019].
- [281] Y. W. Chao, Z. Wang, Y. He, J. Wang, and J. Deng, “HICO & HICO-DET Benchmarks for Recognizing Human-Object Interactions in Images,” *Www*, 2015. [Online]. Available: <http://www-personal.umich.edu/~ywchao/hico/>. [Accessed: 25-Mar-2019].
- [282] S. Ma, S. A. Bargal, J. Zhang, L. Sigal, and S. Sclaroff, “Do less and achieve more: Training CNNs for action recognition utilizing action images from the Web,” *Pattern Recognit.*, vol. 68, pp. 334–345, 2017.
- [283] S. Ma, S. A. Bargal, J. Zhang, L. Sigal, and S. Sclaroff, “BU-action Datasets,” *Www*, 2017. [Online]. Available: <http://cs-people.bu.edu/sbargal/BU-action/>. [Accessed: 25-Mar-2019].
- [284] Y. W. Chao, Y. Liu, X. Liu, H. Zeng, and J. Deng, “Learning to Detect Human-Object Interactions,” *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 381–389, 2018.
- [285] B. C. I. Laptev, “Recognition of human actions,” *November*, 2011. .
- [286] University of Surrey and CErTH-ITI, “i3dpost multi-view human actiondatasets,” *January*, 2012. .
- [287] Z. Liu, “MSR Action Dataset,” *Www*, 2010. [Online]. Available: <https://www.uow.edu.au/~wanqing/#Datasets>. [Accessed: 25-Mar-2019].
- [288] V. Kulathumani, “WVU Multi-View Action Recognition Dataset.” .
- [289] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, “NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis,” in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1010–1019.
- [290] J. Wang, Z. Liu, Y. Wu, and J. Yuan, “Mining actionlet ensemble for action recognition with depth cameras,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1290–1297, 2012.
- [291] 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,” *Proc. - Int. Conf. Image Process. ICIP*, 2015.
- [292] J. Sung, C. Ponce, B. Selman, and A. Saxena, “Unstructured Human Activity Detection from RGBD Images,” 2011.
- [293] M. B. Holte, C. Tran, M. M. Trivedi, and T. B. Moeslund, “Human pose estimation and activity recognition from multi-view videos: Comparative explorations of recent developments,” *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 5, pp. 538–552, 2012.
- [294] X. Ji and H. Liu, “Advances in view-invariant human motion analysis: A review,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 1, pp. 13–24, 2010.
- [295] B. C. I. Laptev, “Recognition of human actions,” *November*, 2011. [Online]. Available: <http://www.nada.kth.se/cvap/actions/>.
- [296] University of Surrey and CErTH-ITI, “i3dpost multi-view human actiondatasets,” *January*, 2012. [Online]. Available: <http://kahlan.eps.surrey.ac.uk/i3dpostaction>. [Accessed: 20-Jul-2006].
- [297] Y. W. J. Yuan, Z. Liu, “Discriminative video pattern search for efficient action detection,” *January*, 2012. [Online]. Available: [http://users.ece.northwestern.edu/~jyu410/index\\_files/actiondetection.html](http://users.ece.northwestern.edu/~jyu410/index_files/actiondetection.html). [Accessed: 20-Aug-2003].
- [298] V. Kulathumani, “WVU Multi-View Action Recognition Dataset.” [Online]. Available: <http://csee.wvu.edu/~vkkulathumani/wvu-action.html#download2>.

- [299] S. Sadek, A. Al-Hamadi, G. Krell, and B. Michaelis, “Affine-Invariant Feature Extraction for Activity Recognition,” *ISRN Mach. Vis.*, vol. 2013, pp. 1–7, 2013.
- [300] N. Ikizler-Cinbis and S. Sclaroff, “Object, scene and actions: Combining multiple features for human action recognition,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6311 LNCS, no. PART 1, pp. 494–507, 2010.
- [301] M. Ahmad and S. W. Lee, “Human action recognition using shape and CLG-motion flow from multi-view image sequences,” *Pattern Recognit.*, vol. 41, no. 7, pp. 2237–2252, 2008.
- [302] Y. Dedeoğlu, B. U. Töreyn, U. Güdükbay, and A. Enis Çetin, “Silhou[1] Y. Dedeoğlu, B. U. Töreyn, U. Güdükbay, and A. Enis Çetin, ‘Silhouette-based method for object classification and human action recognition in video,’ *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformat)*,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3979 LNCS, pp. 64–77, 2006.
- [303] S. Suzuki and K. A. be, “Topological structural analysis of digitized binary images by border following,” *Comput. Vision, Graph. Image Process.*, vol. 30, no. 1, pp. 32–46, 1985.
- [304] A. Bruhn, J. Weickert, and C. Schnörr, “Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods,” *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 1–21, 2005.
- [305] M. Pietikäinen, *Computer vision using local binary patterns*, vol. 40. 2011.
- [306] M. Pietikäinen, *Computer vision using local binary patterns*, vol. 40. 2011.
- [307] C. Bregler, “Learning and recognizing human dynamics in video sequences,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 568–574, 1997.
- [308] Y. Nam and K. Wohn, “Recognition of space-time hand-gestures using hidden Markov model,” pp. 51–58, 1996.
- [309] J. Yang, Y. Xu, and C. S. Chen, “Human action learning via hidden Markov model,” *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans.*, vol. 27, no. 1, pp. 34–44, 1997.
- [310] Y. W. J. Yuan, Z. Liu, “Discriminative video pattern search for efficient action detection,” *January*, 2012. .
- [311] H. Wang, M. M. Ullah, A. Kläser, I. Laptev, and C. Schmid, “Evaluation of local spatio-temporal features for action recognition,” *Br. Mach. Vis. Conf. BMVC 2009 - Proc.*, 2009.
- [312] N. Srivastava, E. Mansimov, and R. Salakhutdinov, “Unsupervised learning of video representations using LSTMs,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 843–852, 2015.
- [313] O. Oreifej and Z. Liu, “HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 716–723, 2013.
- [314] L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, “Towards Good Practices for Very Deep Two-Stream ConvNets,” pp. 1–5, 2015.
- [315] J. Shotton *et al.*, “Real-Time human pose recognition in parts from single depth images,” *Commun. ACM*, vol. 56, no. 1, pp. 116–124, 2013.
- [316] X. Yang and Y. Tian, “Super normal vector for activity recognition using depth sequences,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 804–811, 2014.
- [317] L. Wang, Y. Qiao, and X. Tang, “Latent hierarchical model of temporal structure for complex activity classification,” *IEEE Trans. Image Process.*, vol. 23, no. 2,

- pp. 810–822, 2014.
- [318] G. E. H. A. Krizhevsky, I. Sutskever, “ImageNet Classification with Deep Convolutional Neural Networks,” *Commun. ACM*, vol. May, 2017.
- [319] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning Spatiotemporal Features with 3D Convolutional Networks,” *Iccv*, 2015.
- [320] H. Wang, C. Schmid, A. Recognition, and T. Iccv, “Action Recognition with Improved Trajectories To cite this version : Action Recognition with Improved Trajectories,” 2013.
- [321] Li Fei-Fei, Wei Dong, Jia Deng, Kai Li, R. Socher, and Li-Jia Li, “ImageNet: A large-scale hierarchical image database,” *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 248–255, 2009.
- [322] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, “Learning deep features for scene recognition using places database,” *Adv. Neural Inf. Process. Syst.*, vol. 1, no. January, pp. 487–495, 2014.
- [323] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu, “Robust 3D action recognition with random occupancy patterns,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7573 LNCS, no. PART 2, pp. 872–885, 2012.
- [324] Y. Du, W. Wang, and L. Wang, “Hierarchical recurrent neural network for skeleton based action recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1110–1118, 2015.
- [325] E. Ohn-Bar and M. M. Trivedi, “Joint angles similarities and HOG2 for action recognition,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 465–470, 2013.
- [326] R. Vemulapalli, F. Arrate, and R. Chellappa, “Human action recognition by representing 3D skeletons as points in a lie group,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 588–595, 2014.
- [327] G. Evangelidis, G. Singh, and R. Horaud, “Skeletal quads: Human action recognition using joint quadruples,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 4513–4518, 2014.
- [328] J. Hu, W. Zheng, J. Lai, and J. Zhang, “Jointly learning heterogeneous features for rgb-d activity Recognition,” vol. 39, no. 11, pp. 2186–2200, 2017.
- [329] W. Li, Z. Zhang, and Z. Liu, “Expandable data-driven graphical modeling of human actions based on salient postures,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1499–1510, 2008.
- [330] D. Kim, W. H. Yun, H. S. Yoon, and J. Kim, “Action recognition with depth maps using HOG descriptors of multi-view motion appearance and history,” *UBICOMM 2014 - 8th Int. Conf. Mob. Ubiquitous Comput. Syst. Serv. Technol.*, no. c, pp. 126–130, 2014.
- [331] and Y. Y. Chen Chen, Zhenjie Hou, Baochang Zhang, Junjun Jiang, “Gradient Local Auto-Correlations and Extreme Learning Machine for Depth-Based Activity Recognition Chen,” vol. 2, no. December, pp. 857–866, 2015.
- [332] X. Yang, C. Zhang, and Y. Tian, “Recognizing actions using depth motion maps-based histograms of oriented gradients,” *MM 2012 - Proc. 20th ACM Int. Conf. Multimed.*, no. c, pp. 1057–1060, 2012.
- [333] R. C. O’Reilly, D. Wyatte, S. Herd, B. Mingus, and D. J. Jilk, “Deep Residual Learning for Image Recognition Kaiming,” *Front. Psychol.*, vol. 4, no. APR, pp. 428–429, 2013.
- [334] C. Szegedy *et al.*, “Going deeper with convolutions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015.

- [335] J. Wang, A. Cherian, and F. Porikli, "Ordered pooling of optical flow sequences for action recognition," *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017*, pp. 168–176, 2017.
- [336] K. Simonyan, "Two-Stream Convolutional Networks for Action Recognition in Videos," pp. 1–9.
- [337] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, "Convolutional learning of spatio-temporal features," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6316 LNCS, no. PART 6, pp. 140–153, 2010.
- [338] G. E. H. A. Krizhevsky, I. Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. (2012)*, pp., vol. 12, pp. 1097–1105, 2012.
- [339] J. K. Aggarwal and M. S. Ryoo, "Human Activity Analysis : A Review," 2007.
- [340] J. Zhang, P. Wang, C. Tang, W. Li, Z. Gao, and P. Ogunbona, "ConvNets-Based Action Recognition from Depth Maps through Virtual Cameras and Pseudocoloring," pp. 1119–1122, 2016.
- [341] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. Ogunbona, "Deep Convolutional Neural Networks for Action Recognition Using Depth Map Sequences," 2015.
- [342] P. Wang *et al.*, "Action Recognition From Depth Maps Using Deep Convolutional Neural Networks," *IEEE Trans. Human-Machine Syst.*, vol. 46, no. 4, pp. 498–509, 2016.
- [343] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the Devil in the Details : Delving Deep into Convolutional Nets," pp. 1–11.
- [344] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
- [345] E. Triantaphyllou, B. Shu, S. N. Sanchez, and T. Ray, "Multi-Criteria Decision Making : An Operations Research Approach," vol. 15, pp. 175–186, 1998.
- [346] S. Althloothi, M. H. Mahoor, X. Zhang, and R. M. Voyles, "Human activity recognition using multi-features and multiple kernel learning," *Pattern Recognit.*, vol. 47, no. 5, pp. 1800–1812, 2014.
- [347] M. Li, H. Leung, and H. P. H. Shum, "Human Action Recognition via Skeletal and Depth based Feature Fusion," pp. 123–132, 2016.
- [348] J. Wang, S. Member, Z. Liu, S. Member, Y. Wu, and S. Member, "Learning Actionlet Ensemble for 3D Human Action Recognition," vol. 36, no. 5, pp. 914–927, 2014.
- [349] P. Khaire, P. Kumar, and J. Imran, "Combining CNN streams of RGB-D and skeletal data for human activity recognition," *Pattern Recognit. Lett.*, vol. 115, pp. 107–116, 2018.
- [350] S. Liu and H. Wang, "Human Activities Recognition Based on Skeleton Information via Sparse Representation," vol. 12, no. 1, pp. 1–11, 2018.
- [351] C. Li, Y. Hou, P. Wang, and S. Member, "With Convolutional Neural Networks," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 624–628, 2017.
- [352] S. Gaglio, G. Lo Re, S. Member, and M. Morana, "Human Activity Recognition Process Using 3-D Posture Data," *IEEE Trans. Human-Machine Syst.*, vol. 45, no. 5, pp. 586–597, 2015.
- [353] J.-F. Hu, W.-S. Zheng, J. Lai, and J. Zhang, "Jointly learning heterogeneous features for RGB-D activity recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 5344–5352, 2015.

## LIST OF PUBLICATIONS

---

### Published

#### In SCI/SCIE Journals

1. Roshan Singh, AKS Kushwaha, R Srivastava, “Multi-view recognition system for human activity based on multiple features for video surveillance system”, **Multimedia Tools and Applications, Springer**, 78 (12), 17165-17196, 2019. **(Impact Factor – 2.101)**
2. Roshan Singh, JK Dhillon, AKS Kushwaha, R Srivastava, “Depth based enlarged temporal dimension of 3D deep convolutional network for activity recognition”, **Multimedia Tools and Applications, Springer**, 1-16, 2018. **(Impact Factor – 2.101)**
3. Roshan Singh, A Sonawane, R Srivastava, “Recent evolution of modern datasets for human activity recognition: a deep survey”, **Multimedia Systems, Springer**, 1-24, 2019. **(Impact Factor – 1.956)**
4. Roshan Singh, Rajat Khurana, AKS Kushwaha, R Srivastava, “Combing CNN Streams of Dynamic Image and Depth Data for Action Recognition”. **Multimedia Systems, Springer**, Jan 2020. **(Impact Factor – 1.956)**
5. Roshan Singh, Rajat Khurana, AKS Kushwaha, R Srivastava, “A Dual Stream Model for Activity Recognition: Exploiting Residual- CNN with Transfer Learning”, **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, Taylor & Francis**, August 2020.

#### In International Conferences

1. Roshan Singh, AKS Kushwaha, R Khurana, R Srivastava, “Activity Recognition by Delving deeper using CNN and RNN”, ISCON 2019, GLA University, Mathura, UP, 21-22 Nov. 2019.
2. Roshan Singh, Rajeev Srivastava, “Some Contemporary Approaches for Human Activity Recognition: A Survey”, PARC 2020, GLA University, Mathura, UP, 28-29 Feb 2020

#### Communicated:

1. Recent Trends in Human Activity Recognition- A Comparative Study”, has been communicated in Cognitive Systems Research Journal, Elsevier- **Under Review**

