List of Publications

1. Jain, Hiteshi, and Gaurav Harit. "A framework to assess sun salutation videos." *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*. ACM, 2016.

2. Jain, Hiteshi, and Gaurav Harit. "Detecting missed and anomalous action segments using approximate string matching algorithm." *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics.* Springer, Singapore, 2017.

3. Jain, Hiteshi, and Gaurav Harit. Leveraging information from imperfect examples : Common action sequence mining from a mix of incorrect performances. Accepted in *The Eleventh Indian Conference on Computer Vision, Graphics and Image Processing*, 2018.

4. Jain, Hiteshi, and Gaurav Harit. "Unsupervised Temporal Segmentation of Human Action Using Community Detection." *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018.

5. Jain, Hiteshi, and Gaurav Harit. An unsupervised sequence-to-sequence autoencoder based human action scoring model. Accepted in *7th IEEE Global Conference on Signal and Information Processing*, 2019.

6. Jain, Hiteshi, Gaurav Harit and Sharma, Avinash, "Action Quality Assessment using Siamese Network Based Deep Metric Learning," in *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2020.3017727.

...

References

Aggarwal, J. K. and Ryoo, M. S. (2011), "Human activity analysis: A review", *ACM Computing Surveys (CSUR)*, Vol. 43, No. 3, p. 16 2011

Alexiadis, D. S. and Daras, P. (2014), "Quaternionic signal processing techniques for automatic evaluation of dance performances from MoCap data", *IEEE Transactions on Multimedia*, Vol. 16, No. 5, pp. 1391–1406 2014

Alexiadis, D. S., Kelly, P., Daras, P., O'Connor, N. E., Boubekeur, T., and Moussa, M. B. (2011), "Evaluating a dancer's performance using kinect-based skeleton tracking", *in Proceedings of the 19th ACM international conference on Multimedia*, pp. 659–662, ACM 2011

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

Ansorge, C. J. and Scheer, J. K. (1988), "International bias detected in judging gymnastic competition at the 1984 Olympic Games", *Research quarterly for exercise and sport*, Vol. 59, No. 2, pp. 103–107 1988

Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., and Baskurt, A. (2010), "Action classification in soccer videos with long short-term memory recurrent neural networks", *in International Conference on Artificial Neural Networks*, pp. 154–159, Springer 2010

Badler, N. I., Hollick, M. J., and Granieri, J. P. (1993), "Real-time control of a virtual human using minimal sensors", *Presence: Teleoperators & Virtual Environments*, Vol. 2, No. 1, pp. 82–86 1993

Belongie, S., Malik, J., and Puzicha, J. (2001), "Shape context: A new descriptor for shape matching and object recognition", *in Advances in neural information processing systems*, pp. 831–837 2001

Belongie, S., Malik, J., and Puzicha, J. (2002), "Shape matching and object recognition using shape contexts", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 24, No. 4, pp. 509–522 2002

Blank, M., Gorelick, L., Shechtman, E., Irani, M., and Basri, R. (2005), "Actions as space-time shapes", *in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 2, pp. 1395–1402, IEEE 2005

Bobick, A. F. and Davis, J. W. (2001), "The Recognition of Human Movement Using Temporal Templates", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol. 23, pp. 257–267 2001

Borzeshi, E. Z., Concha, O. P., Da Xu, R. Y., and Piccardi, M. (2013a), "Joint action segmentation and classification by an extended hidden Markov model", *IEEE Signal Processing Letters*, Vol. 20, No. 12, pp. 1207–1210 2013a

Borzeshi, E. Z., Concha, O. P., Da Xu, R. Y., and Piccardi, M. (2013b), "Joint action segmentation and classification by an extended hidden Markov model", *IEEE Signal Processing Letters*, Vol. 20, No. 12, pp. 1207–1210 2013b

Brand, M., Oliver, N., and Pentland, A. (1997), "Coupled hidden Markov models for complex action recognition", *in Computer vision and pattern recognition, 1997. proceedings., 1997 ieee computer society conference on*, pp. 994–999, IEEE 1997

Caba Heilbron, F., Escorcia, V., Ghanem, B., and Carlos Niebles, J. (2015), "Activitynet: A large-scale video benchmark for human activity understanding", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 961–970 2015

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

Chang, C.-Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., Sawchuk, A. A., and Rizzo, A. A. (2012), "Towards pervasive physical rehabilitation using Microsoft Kinect", *in 2012 6th international*

conference on pervasive computing technologies for healthcare (PervasiveHealth) and workshops, pp. 159–162, IEEE 2012

Chang, Y.-J., Chen, S.-F., and Huang, J.-D. (2011), "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities", *Research in developmental disabilities*, Vol. 32, No. 6, pp. 2566–2570 2011

Chaquet, J. M., Carmona, E. J., and Fernández-Caballero, A. (2013), "A survey of video datasets for human action and activity recognition", *Computer Vision and Image Understanding*, Vol. 117, No. 6, pp. 633–659 2013

Chopra, S., Hadsell, R., LeCun, Y. *et al.* (2005), "Learning a similarity metric discriminatively, with application to face verification", *in CVPR (1)*, pp. 539–546 2005

Cohen, I., Garg, A., Huang, T. S. *et al.* (2000), "Emotion recognition from facial expressions using multilevel HMM", *in NIPS*, Vol. 2, Citeseer 2000

Concha, O. P., Da Xu, R. Y., Moghaddam, Z., and Piccardi, M. (2011a), "HMM-MIO: an enhanced hidden Markov model for action recognition", *in CVPR 2011 WORKSHOPS*, pp. 62–69, IEEE 2011a

Concha, O. P., Da Xu, R. Y., Moghaddam, Z., and Piccardi, M. (2011b), "HMM-MIO: an enhanced hidden Markov model for action recognition", *in CVPR 2011 WORKSHOPS*, pp. 62–69, IEEE 2011b

Costa, L., Trigueiros, P., and Cunha, A. (2016), "Automatic Meal Intake Monitoring Using Hidden Markov Models", *Procedia Computer Science*, Vol. 100, pp. 110–117 2016

Dalal, N. and Triggs, B. (2005), "Histograms of oriented gradients for human detection", *in 2005 CVPR*, Vol. 1, pp. 886–893, IEEE 2005

De la Torre, F., Campoy, J., Ambadar, Z., and Cohn, J. F. (2007), "Temporal segmentation of facial behavior", *in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE 2007

De la Torre, F., Hodgins, J. K., Montano, J., and Valcarcel, S. (2009), "Detailed Human Data Acquisition of Kitchen Activities: the CMU-Multimodal Activity Database (CMU-MMAC)", *in Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, in conjuction with CHI 2009* 2009

Derpanis, K. G., Lecce, M., Daniilidis, K., and Wildes, R. P. (2012), "Dynamic scene understanding: The role of orientation features in space and time in scene classification", *in 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1306–1313, IEEE 2012

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

Doughty, H., Damen, D., and Mayol-Cuevas, W. (2018a), "Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6057–6066 2018a

Doughty, H., Mayol-Cuevas, W., and Damen, D. (2018b), "The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos", *arXiv preprint arXiv:1812.05538* 2018b

Du, Y., Wang, W., and Wang, L. (2015), "Hierarchical recurrent neural network for skeleton based action recognition", *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1110–1118 2015

Duchenne, O., Laptev, I., Sivic, J., Bach, F., and Ponce, J. (2009a), "Automatic annotation of human actions in video", *in 2009 IEEE 12th International Conference on Computer Vision*, pp. 1491–1498, IEEE 2009a

Duchenne, O., Laptev, I., Sivic, J., Bach, F., and Ponce, J. (2009b), "Automatic annotation of human actions in video", *in 2009 IEEE 12th ICCV*, pp. 1491–1498, IEEE 2009b

Efros, A. A., Berg, A. C., Mori, G., and Malik, J. (2003a), "Recognizing Action at a Distance.", *in ICCV*, pp. 726–733, IEEE Computer Society 2003a

Efros, A. A., Berg, A. C., Mori, G., Malik, J. *et al.* (2003b), "Recognizing Action at a Distance.", *in ICCV*, Vol. 3, pp. 726–733 2003b

Emerson, J. W., Seltzer, M., and Lin, D. (2009), "Assessing judging bias: An example from the 2000

Olympic Games", *The American Statistician*, Vol. 63, No. 2, pp. 124–131 2009
Fallahzadeh, R. and Ghasemzadeh, H. (2017), "Personalization without user interruption: boosting activity recognition in new subjects using unlabeled data", *in Proceedings of the 8th International Conference on Cyber-Physical Systems*, pp. 293–302, ACM 2017
Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2018), "Evaluating surgical skills from kinematic data using convolutional neural networks", *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 214–221, Springer 2018
Felzenszwalb, P. F. and Huttenlocher, D. P. (2005), "Pictorial structures for object recognition", *International Journal of Computer Vision*, Vol. 61, No. 1, pp. 55–79 2005
Ferrigno, G. and Pedotti, A. (1985), "ELITE: a digital dedicated hardware system for movement analysis via real-time TV signal processing", *IEEE Transactions on Biomedical Engineering*, , No. 11, pp. 943–950 1985
Forney, G. D. (1973), "The viterbi algorithm", *Proceedings of the IEEE*, Vol. 61, No. 3, pp. 268–278 1973
Forsyth, D. A., Arikan, O., Ikemoto, L., O'Brien, J., Ramanan, D. *et al.* (2006), "Computational studies of human motion: Part 1, tracking and motion synthesis", *Foundations and Trends® in Computer Graphics and Vision*, Vol. 1, No. 2–3, pp. 77–254 2006
Fortunato, S. (2010), "Community detection in graphs", *Physics reports*, Vol. 486, No. 3, pp. 75–174 2010
Gavrila, D. M. (1999), "The visual analysis of human movement: A survey", *Computer vision and image understanding*, Vol. 73, No. 1, pp. 82–98 1999
Gilbert, A., Illingworth, J., and Bowden, R. (2009), "Fast realistic multi-action recognition using mined dense spatio-temporal features", *in 2009 IEEE 12th ICCV*, pp. 925–931, IEEE 2009
Gordon, A. S. (1995), "Automated video assessment of human performance", *in Proceedings of AI-ED*, pp. 16–19 1995
Gorelick, L., Blank, M., Shechtman, E., Irani, M., and Basri, R. (2007), "Actions as Space-Time Shapes", *Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 12, pp. 2247–2253 December 2007
Grushin, A., Monner, D. D., Reggia, J. A., and Mishra, A. (2013), "Robust human action recognition via long short-term memory", *in The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE 2013
Hassner, T. (2013), "A critical review of action recognition benchmarks", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 245–250 2013
He, K., Zhang, X., Ren, S., and Sun, J. (2016), "Deep residual learning for image recognition", *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 2016
Herda, L., Fua, P., Plankers, R., Boulic, R., and Thalmann, D. (2000), "Skeleton-based motion capture for robust reconstruction of human motion", *in Computer Animation 2000. Proceedings*, pp. 77–83, IEEE 2000
Hoai, M., Lan, Z.-Z., and De la Torre, F. (2011), "Joint segmentation and classification of human actions in video", *in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011*, pp. 3265–3272, IEEE 2011
Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 1997
Hu, M.-C., Chen, C.-W., Cheng, W.-H., Chang, C.-H., Lai, J.-H., and Wu, J.-L. (2015), "Real-time human movement retrieval and assessment with kinect sensor", *IEEE transactions on cybernetics*, Vol. 45, No. 4, pp. 742–753 2015
Jain, H. and Harit, G. (2016), "A framework to assess Sun salutation videos", *in Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, p. 29, ACM 2016
Jain, H. and Harit, G. (2018), "Detecting Missed and Anomalous Action Segments Using Approximate String Matching Algorithm", *in Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pp. 101–111, Springer Singapore, Singapore 2018
Johansson, G. (1973), "Visual perception of biological motion and a model for its analysis", *Perception & psychophysics*, Vol. 14, No. 2, pp. 201–211 1973

Jug, M., Perš, J., Dežman, B., and Kovačič, S. (2003), "Trajectory based assessment of coordinated human activity", *in ICCV*, pp. 534–543, Springer 2003

Kliper-Gross, O., Hassner, T., and Wolf, L. (2012), "The action similarity labeling challenge", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 615–621 2012

Koppula, H. S., Gupta, R., and Saxena, A. (2013), "Learning human activities and object affordances from rgb-d videos", *The International Journal of Robotics Research*, Vol. 32, No. 8, pp. 951–970 2013

Kosmopoulos, D., Papoutsakis, K., and Argyros, A. (2016), "A framework for online segmentation and classification of modeled actions performed in the context of unmodeled ones", *IEEE Transactions on Circuits and Systems for Video Technology* 2016

Kosmopoulos, D. I., Papoutsakis, K., and Argyros, A. A. (2011), "Online segmentation and classification of modeled actions performed in the context of unmodeled ones", *Trans. PAMI*, Vol. 33, No. 11, pp. 2188–2202 2011

Krüger, B., Vögele, A., Willig, T., Yao, A., Klein, R., and Weber, A. (2017), "Efficient unsupervised temporal segmentation of motion data", *IEEE Transactions on Multimedia*, Vol. 19, No. 4, pp. 797–812 2017

Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., and Serre, T. (2011), "HMDB: a large video database for human motion recognition", *in 2011 International Conference on Computer Vision*, pp. 2556–2563, IEEE 2011

Kyan, M., Sun, G., Li, H., Zhong, L., Muneesawang, P., Dong, N., Elder, B., and Guan, L. (2015), "An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment", *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 6, No. 2, p. 23 2015

Lancichinetti, A. and Fortunato, S. (2009), "Community detection algorithms: a comparative analysis", *Physical review E*, Vol. 80, No. 5, p. 056117 2009

Lange, B., Chang, C.-Y., Suma, E., Newman, B., Rizzo, A. S., and Bolas, M. (2011), "Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor", *in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1831–1834, IEEE 2011

Laptev, I. (2005a), "On space-time interest points", *International Journal of Computer Vision*, Vol. 64, No. 2-3, pp. 107–123 2005a

Laptev, I. (2005b), "On space-time interest points", *IJCV*, Vol. 64, No. 2-3, pp. 107–123 2005b

Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B. (2008a), "Learning realistic human actions from movies", *in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE 2008a

Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B. (2008b), "Learning realistic human actions from movies", *in CVPR*, pp. 1–8, IEEE 2008b

Li, Y., Chai, X., and Chen, X. (2018), "End-To-End Learning for Action Quality Assessment", *in Pacific Rim Conference on Multimedia*, pp. 125–134, Springer 2018

Liu, H., Feris, R., and Sun, M.-T. (2011), "Benchmarking datasets for human activity recognition", in Visual Analysis of Humanspp. 411–427, Springer 2011

Lv, F. and Nevatia, R. (2007), "Single view human action recognition using key pose matching and viterbi path searching", *in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8, IEEE 2007

Marr, D. and Nishihara, H. K. (1978), "Representation and recognition of the spatial organization of three-dimensional shapes", *Proceedings of the Royal Society of London B: Biological Sciences*, Vol. 200, No. 1140, pp. 269–294 1978

Marszalek, M., Laptev, I., and Schmid, C. (2009), "Actions in context", *in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 2929–2936, IEEE 2009

Martin, J. H. and Jurafsky, D. (2000), "Speech and language processing", *International Edition*, Vol. 710 2000

Mazurova, E. and Penttinen, E. (2020), "Probing Athletes' Perceptions Towards Electronic Judging Systems–A Case Study in Gymnastics", *in Proceedings of the 53rd Hawaii International Conference on System Sciences* 2020

Mendoza, M. Á. and De La Blanca, N. P. (2008), "Applying space state models in human action recognition: a comparative study", *in International Conference on Articulated Motion and Deformable Objects*, pp. 53–62, Springer 2008

Moeslund, T. B., Hilton, A., and Krüger, V. (2006), "A Survey of Advances in Vision-based Human Motion Capture and Analysis", *Comput. Vis. Image Underst.*, Vol. 104, No. 2, pp. 90–126 Nov 2006, URL `http://dx.doi.org/10.1016/j.cviu.2006.08.002`

Molet, T., Boulic, R., and Thalmann, D. (1999), "Human motion capture driven by orientation measurements", *Presence: Teleoperators and Virtual Environments*, Vol. 8, No. 2, pp. 187–203 1999

Morgan, H. N. and Rotthoff, K. W. (2014), "The harder the task, the higher the score: Findings of a difficulty bias", *Economic Inquiry*, Vol. 52, No. 3, pp. 1014–1026 2014

Natarajan, P. and Nevatia, R. (2008), "View and scale invariant action recognition using multiview shape-flow models", *in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, IEEE 2008

Newell, A., Yang, K., and Deng, J. (2016), "Stacked hourglass networks for human pose estimation", *in European Conference on Computer Vision*, pp. 483–499, Springer 2016

Newman, M. E. (2004), "Fast algorithm for detecting community structure in networks", *Physical review E*, Vol. 69, No. 6, p. 066133 2004

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002), "On spectral clustering: Analysis and an algorithm", *in Advances in neural information processing systems*, pp. 849–856 2002

Niebles, J. C., Wang, H., and Fei-Fei, L. (2008a), "Unsupervised learning of human action categories using spatial-temporal words", *International journal of computer vision*, Vol. 79, No. 3, pp. 299–318 2008a

Niebles, J. C., Wang, H., and Fei-Fei, L. (2008b), "Unsupervised learning of human action categories using spatial-temporal words", *IJCV*, Vol. 79, No. 3, pp. 299–318 2008b

Niebles, J. C., Chen, C.-W., and Fei-Fei, L. (2010), "Modeling temporal structure of decomposable motion segments for activity classification", *in European conference on computer vision*, pp. 392–405, Springer 2010

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

Omkar, S. (2012), "Surya Namaskaar for holistic well being: A comprehensive review of Surya Namaskaar", *Journal of Yoga & Physical Therapy*, Vol. 2012 2012

Oneata, D., Verbeek, J., and Schmid, C. (2013), "Action and event recognition with fisher vectors on a compact feature set", *in Proceedings of the IEEE International Conference on Computer Vision*, pp. 1817–1824 2013

Ozkan, D., Scherer, S., and Morency, L.-P. (2012), "Step-wise emotion recognition using concatenated-hmm", *in Proceedings of the 14th ACM ICMI*, pp. 477–484, ACM 2012

Pachoulakis, I., Xilourgos, N., Papadopoulos, N., and Analyti, A. (2016), "A Kinect-Based Physiotherapy and Assessment Platform for Parkinson's Disease Patients", *Journal of medical engineering*, Vol. 2016 2016

Palma, C., Salazar, A., and Vargas, F. (2016), "HMM and DTW for evaluation of therapeutical gestures using kinect", *arXiv preprint arXiv:1602.03742* 2016

Papoutsakis, K., Panagiotakis, C., and Argyros, A. A. (2017), "Temporal Action Co-segmentation in 3D Motion Capture Data and Videos", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6827–6836 2017

Parisi, G. I., Magg, S., and Wermter, S. (2016), "Human motion assessment in real time using recurrent self-organization", *in Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 71–76, IEEE 2016

Parmar, P. and Morris, B. T. (2019), "What and how well you performed? A multitask learning approach to action quality assessment", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 304–313 2019

Parmar, P. and Tran Morris, B. (2017), "Learning to score olympic events", *in proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition Workshops, pp. 20–28 2017

Pirsiavash, H., Vondrick, C., and Torralba, A. (2014), "Assessing the quality of actions", *in European Conference on Computer Vision*, pp. 556–571, Springer 2014

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

Poppe, R. (2007), "Vision-based human motion analysis: An overview", *Computer vision and image understanding*, Vol. 108, No. 1, pp. 4–18 2007

Poppe, R. (2010), "A survey on vision-based human action recognition", *Image and vision computing*, Vol. 28, No. 6, pp. 976–990 2010

Qiu, Z., Yao, T., and Mei, T. (2017), "Learning spatio-temporal representation with pseudo-3d residual networks", *in proceedings of the IEEE International Conference on Computer Vision*, pp. 5533–5541 2017

Rabiner, L. R. (1989), "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286 1989

Rao, C., Gritai, A., Shah, M., and Syeda-Mahmood, T. (2003), "View-invariant alignment and matching of video sequences", *in null*, p. 939, IEEE 2003

Raptis, M., Kirovski, D., and Hoppe, H. (2011), "Real-time classification of dance gestures from skeleton animation", *in Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation*, pp. 147–156, ACM 2011

Reddy, K. K. and Shah, M. (2013), "Recognizing 50 human action categories of web videos", *Machine Vision and Applications*, Vol. 24, No. 5, pp. 971–981 2013

Rodriguez, M. D., Ahmed, J., and Shah, M. (2008), "Action MACH: a spatio-temporal maximum average correlation height filter for action recognition", *in In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* 2008

Sakoe, H. and Chiba, S. (1978), "Dynamic programming algorithm optimization for spoken word recognition", *IEEE transactions on acoustics, speech, and signal processing*, Vol. 26, No. 1, pp. 43–49 1978

Sapp, B. and Taskar, B. (2013), "Modec: Multimodal decomposable models for human pose estimation", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3674–3681 2013

Schuldt, C., Laptev, I., and Caputo, B. (2004a), "Recognizing human actions: a local SVM approach", *in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, Vol. 3, pp. 32–36, IEEE 2004a

Schuldt, C., Laptev, I., and Caputo, B. (2004b), "Recognizing Human Actions: A Local SVM Approach", *in Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Volume 03*, ICPR '04, pp. 32–36, IEEE Computer Society, Washington, DC, USA 2004b, URL `http://dx.doi.org/10.1109/ICPR.2004.747`

Shi, Q., Wang, L., Cheng, L., and Smola, A. (2008), "Discriminative human action segmentation and recognition using semi-Markov model", *in IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.*, pp. 1–8, IEEE 2008

Shi, Y., Tian, Y., Wang, Y., and Huang, T. (2017), "Sequential deep trajectory descriptor for action recognition with three-stream CNN", *IEEE Transactions on Multimedia*, Vol. 19, No. 7, pp. 1510–1520 2017

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A. *et al.* (2013), "Efficient human pose estimation from single depth images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 12, pp. 2821–2840 2013

Soomro, K., Zamir, A. R., and Shah, M. (2012), "UCF101: A dataset of 101 human actions classes from videos in the wild", *arXiv preprint arXiv:1212.0402* 2012

Soran, B., Farhadi, A., and Shapiro, L. (2015), "Generating Notifications for Missing Actions: Don't Forget to Turn the Lights Off!", *in Proceedings of the IEEE International Conference on Computer Vision*, pp. 4669–4677 2015

Starner, T. and Pentland, A. (1997), "Real-time american sign language recognition from video using

hidden markov models", in Motion-Based Recognitionpp. 227–243, Springer 1997
Su, C.-J. (2013), "Personal rehabilitation exercise assistant with kinect and dynamic time warping", *International Journal of Information and Education Technology*, Vol. 3, No. 4, p. 448 2013
Su, C.-J., Chiang, C.-Y., and Huang, J.-Y. (2014), "Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic", *Applied Soft Computing*, Vol. 22, pp. 652–666 2014
Sutskever, I., Vinyals, O., and Le, Q. V. (2014), "Sequence to sequence learning with neural networks", *in Advances in neural information processing systems*, pp. 3104–3112 2014
Sutton, C. and McCallum, A., *An introduction to conditional random fields for relational learning*, Vol. 2, Introduction to statistical relational learning. MIT Press 2006
Toshev, A. and Szegedy, C. (2014), "Deeppose: Human pose estimation via deep neural networks", *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1653–1660 2014
Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015), "Learning spatiotemporal features with 3d convolutional networks", *in Proceedings of the IEEE international conference on computer vision*, pp. 4489–4497 2015
Turaga, P., Chellappa, R., Subrahmanian, V. S., and Udrea, O. (2008), "Machine recognition of human activities: A survey", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 11, pp. 1473–1488 2008
Turaga, P., Veeraraghavan, A., and Chellappa, R. (2009), "Unsupervised view and rate invariant clustering of video sequences", *Computer Vision and Image Understanding*, Vol. 113, No. 3, pp. 353–371 2009
Veeraraghavan, A., Chellappa, R., and Roy-Chowdhury, A. K. (2006), "The function space of an activity", *in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 1, pp. 959–968, IEEE 2006
Veeriah, V., Zhuang, N., and Qi, G.-J. (2015), "Differential recurrent neural networks for action recognition", *in Proceedings of the IEEE international conference on computer vision*, pp. 4041–4049 2015
Venkataraman, V., Vlachos, I., and Turaga, P. K. (2015), "Dynamical Regularity for Action Analysis.", *in BMVC*, pp. 67–1 2015
Wang, J., Liu, Z., Wu, Y., and Yuan, J. (2012), "Mining actionlet ensemble for action recognition with depth cameras", *in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1290–1297, IEEE 2012
Weinland, D., Ronfard, R., and Boyer, E. (2006), "Free Viewpoint Action Recognition Using Motion History Volumes", *Comput. Vis. Image Underst.*, Vol. 104, No. 2, pp. 249–257 Nov 2006, URL `http://dx.doi.org/10.1016/j.cviu.2006.07.013`
Weinland, D., Ronfard, R., and Boyer, E. (2011), "A survey of vision-based methods for action representation, segmentation and recognition", *Computer vision and image understanding*, Vol. 115, No. 2, pp. 224–241 2011
Willems, G., Tuytelaars, T., and Van Gool, L. (2008), "An efficient dense and scale-invariant spatio-temporal interest point detector", *in European conference on computer vision*, pp. 650–663, Springer 2008
Wu, C., Zhang, J., Savarese, S., and Saxena, A. (2015a), "Watch-n-patch: Unsupervised understanding of actions and relations", *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4362–4370 2015a
Wu, C., Zhang, J., Savarese, S., and Saxena, A. (2015b), "Watch-n-Patch: Unsupervised Understanding of Actions and Relations", *in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* June 2015b
Xiang, X., Tian, Y., Reiter, A., Hager, G. D., and Tran, T. D. (2018), "S3d: Stacking segmental p3d for action quality assessment", *in 2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 928–932, IEEE 2018
Yamato, J., Ohya, J., and Ishii, K. (1992a), "Recognizing human action in time-sequential images using hidden markov model", *in Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992*

IEEE Computer Society Conference on, pp. 379–385, IEEE 1992a

Yamato, J., Ohya, J., and Ishii, K. (1992b), "Recognizing human action in time-sequential images using hidden markov model", *in CVPR*, pp. 379–385, IEEE 1992b

Yang, Y. and Ramanan, D. (2011), "Articulated pose estimation with flexible mixtures-of-parts", *in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1385–1392, IEEE 2011

Yi, D., Lei, Z., and Li, S. (2014), "Deep metric learning for practical person re-identification (2014)", *ArXiv e-prints* 2014

Yucer, S. and Akgul, Y. S. (2018), "3D Human Action Recognition with Siamese-LSTM Based Deep Metric Learning", *arXiv preprint arXiv:1807.02131* 2018

Zhang, H., Zhou, W., and Parker, L. E. (2014), "Fuzzy segmentation and recognition of continuous human activities", *in 2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6305–6312, IEEE 2014

Zhao, W., Feng, H., Lun, R., Espy, D. D., and Reinthal, M. A. (2014), "A Kinect-based rehabilitation exercise monitoring and guidance system", *in Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on*, pp. 762–765, IEEE 2014

Zhou, F., De la Torre, F., and Hodgins, J. K. (2008), "Aligned cluster analysis for temporal segmentation of human motion", *in 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG'08.*, pp. 1–7, IEEE 2008

Zhou, F., De la Torre, F., and Hodgins, J. K. (2013), "Hierarchical aligned cluster analysis for temporal clustering of human motion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 3, pp. 582–596 2013

Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., and Xie, X. (2016), "Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks", *in Thirtieth AAAI Conference on Artificial Intelligence* 2016

Zia, A., Sharma, Y., Bettadapura, V., Sarin, E. L., Clements, M. A., and Essa, I. (2015), "Automated assessment of surgical skills using frequency analysis", *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 430–438, Springer 2015

Zia, A., Sharma, Y., Bettadapura, V., Sarin, E. L., Ploetz, T., Clements, M. A., and Essa, I. (2016), "Automated video-based assessment of surgical skills for training and evaluation in medical schools", *International journal of computer assisted radiology and surgery*, Vol. 11, No. 9, pp. 1623–1636 2016