

Michael S. Ryoo

231 New Computer Science Building,
Stony Brook University, Stony Brook, NY 11794, USA
<http://michaelryoo.com/>
<https://www.cs.stonybrook.edu/people/faculty/MichaelRyoo>

e-mails: mryoo@cs.stonybrook.edu
mryoo@google.com

Education

The University of Texas at Austin	2004~2008
Ph.D. in Electrical and Computer Engineering, August 2008	
M.S. in Electrical and Computer Engineering, August 2006	
Korea Advanced Institute of Science and Technology (KAIST)	2000~2004
B.S. in Computer Science, <i>magna cum laude</i> , August 2004	

Professional Appointments

<i>SUNY Empire Innovation Associate Professor</i>	Stony Brook, NY
Department of Computer Science, Stony Brook University	(2019.09 ~)
<i>Staff Research Scientist</i>	NYC, NY
Robotics at Google -> Google DeepMind	(2019.08 ~)
<i>Visiting Faculty</i>	Mountain View, CA
Google Brain -> Robotics at Google	(2018.09 ~ 2019.08)
<i>Founder and CTO</i>	South Korea
EgoVid Inc.	(2016.08 ~2019.08)
<i>Assistant Professor</i>	Bloomington, IN
Department of Computer Science (CS), Indiana University Bloomington	(2015.08 ~ 2019.08)
<i>Research Affiliate (adjunct)</i>	Pasadena, CA
NASA's Jet Propulsion Laboratory (NASA-JPL)	(2015.08 ~2019.08)
<i>Research Technologist</i>	Pasadena, CA
NASA's Jet Propulsion Laboratory (NASA-JPL)	(2011.10 ~ 2015.07)
<i>Research Scientist (military duty for South Korea)</i>	South Korea
Electronics and Telecommunications Research Institute (ETRI) - a national lab	(2008.09 ~ 2011.09)

Selected Awards and Honors

- **Best Paper Award in Robot Vision**
IEEE International Conference on Robotics and Automation (ICRA), 2016.
Gori, Aggarwal, Matthies & Ryoo, "Multi-Type Activity Recognition in Robot-Centric Scenarios"
- **Best Paper Award Finalist**
ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2015.
Ryoo et al., "Robot-Centric Activity Prediction from First-Person Videos: What Will They Do to Me?"

Research Funding

- **(PI)** IITP grant by South Korean Ministry of Science and ICT, "Semantic Action Policy Learning and State Inference for Mobile Robot Intelligence," 2020.01~2022.12, ~\$200,000 for 36 months, with Electronics and Telecommunications Research Institute (ETRI).
- **(PI)** NSF Information and Intelligent Systems (IIS): Core Programs, "RI: Small: Collaborative: Understanding Human-Object Interactions from First-person and Third-person Videos," 2018.08~2022.08, \$250,000 for 36 months, with Y. J. Lee (UC Davis).

- **(PI)** NSF Computer and Network Systems (CNS): Core Programs, “CSR: Small: Collaborative: Decentralized Real-Time Machine Learning Systems on Near-User Edge Devices,” 2018.08~2022.08, \$250,000 for 36 months, with H. Kim (Gatech).
- **(PI)** IITP grant by South Korean Ministry of Science and ICT, “Semantic Action Policy Learning and State Inference for Mobile Robot Intelligence,” 2018.01~2019.12, ~\$130,000 for 24 months, with Electronics and Telecommunications Research Institute (ETRI).
- **(PI)** ARL’s Robotics Collaborative Technology Alliance (RCTA), Task P5-5 “Human Activity Recognition with Context Learning” (2016) and T2C1S2D “Predicting Human Intent and Activity Possibilities” (2017-2018), 2016.01~ 2018.12, \$220,000 for 36 months.
- **(PI)** ICT R&D program of South Korean Ministry of Science, “Recognizing Objects and Events from Videos for XD-Media Special Effects,” 2016.01~2017.12, ~\$260,000 for 24 months, with Electronics and Telecommunications Research Institute (ETRI).
- **(co-PI)** DARPA’s Simplifying Complexity in Scientific Discovery (SIMPLEX), Task “Action Recognition and Learning from a First-Person View,” 2015.03~2016.03, \$90,000 for 12 months, with S.-C. Zhu (UCLA).
- **(PI)** NVIDIA hardware donation program, 2015, 2016.
- **(co-PI, subtask-PI)** ARL’s Robotics Collaborative Technology Alliance (RCTA), Task P5-2 “Understanding of Human Interactions and Reactions,” Phase1: 2012.04~2014.12, “Semantic Understanding of Human Activities,” Phase2: 2015.01~ 2015.12, ~\$500,000, with L. Matthies (JPL).
- **(PI)** NASA-JPL B&P Funding, “Group Activity Recognition from Aerial Videos,” etc., 2013~2014, \$17,000.
- **(PI)** Otis Elevator Korea, “Detection of Abnormal Activities in Elevators Using Cameras,” 2011, \$60,000.

Talks

Tutorials

- *Vision-based Robot Learning*
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, June 2022.
(Speakers: Michael S. Ryoo, Andy Zeng, Pete Florence, Shuran Song, Samir Gadre)
- *Unifying Human Activity Understanding*
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, June 2019.
(Speakers: Gunnar Sigurdsson, Michael S. Ryoo)
- *Human Activity Recognition*
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, June 2018.
(Speakers: Michael S. Ryoo, Greg Mori, Kris Kitani)
- *Emerging Topics in Human Activity Recognition*
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, June 2014.
(Speakers: Michael S. Ryoo, Ivan Laptev, Greg Mori, Sangmin Oh)
- *Activity Recognition for Visual Surveillance*
IEEE Conference on Advanced Video and Signal-based Surveillance (AVSS), Beijing, China, Sep. 2012.
(Speakers: Michael S. Ryoo, Anthony Hoogs, Arslan Basharat, Sangmin Oh)
- *Frontiers of Human Activity Analysis*
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, June 2011.
(Speakers: J. K. Aggarwal, Michael S. Ryoo, Kris Kitani)
- *Understanding Videos – Human Activity Analysis*
11th Pacific Rim International Conference on Artificial Intelligent (PRICAI), Daegu, Korea, August 2010.
(Speakers: Michael S. Ryoo, Kris Kitani)

Selected panels and talks

- *Panel: Embodied Computer Vision*, IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR), June 2022 (with Martial Hebert, Kristen Grauman, and Nick Roy).
- *Representing Longer Videos – TokenLearner*, CVPR Tutorial on 2nd Comprehensive Tutorial on Video Modeling, June 2021.
- *Representing Visual Information in Time*, ECCV Workshop on 4D Vision, August 2020.
- *Video Architecture Search*, CVPR Workshop on Neural Architecture Search and Beyond for Representation Learning, June 2020.
- *Representing Motion in Videos*, ICCV Workshop on Moving Cameras, Seoul, Korea, October 2019.
- *Robots Anticipating Future Scene*, ECCV Workshop on Anticipating Human Behavior, Munich, Germany, September 2018.
- *Human Activity Recognition from Anonymized Videos*, Joint BMTT-PETS Workshop on Tracking and Surveillance (PETS), in conjunction with CVPR, Honolulu, HI, July 2017.
- *Activity Recognition from Persons' Viewpoint and Robots' Viewpoint*, International Workshop on Human Activity Analysis with Highly Diverse Cameras, in conjunction with WACV, Santa Rosa, CA, March 2017.
- *First-Person Activity Recognition: What Are They Doing and What Will They Do to Me?* The 4th International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI), in conjunction with UbiComp, Seattle, WA, September 2014.
- *First-Person Activity Recognition: Understanding Human Interactions from Egocentric Videos*, ICCV Workshop on Understanding Human Activities: Context and Interaction, Sydney, Australia, December 2013.

Selected seminars

- *Machine Learning Research*
Department Seminar (AI), KAIST, Daejeon, Korea, October 2019.
- *Representation Learning for Video Understanding*
Department Seminar (CS), Yale University, New Haven, CT, February 2019.
- *Deep Learning with Human/Robot Activity Videos*
AI and Machine Learning Seminar (ECE), Purdue University, West Lafayette, IN, September 2018.
- *Robot Perception and Action Using Convolutional Human Activity Models*
Department Seminar (CS), University of North Carolina, Chapel Hill, NC, February 2018.
- *Human Activity Recognition from a Robot's Viewpoint*
ARO Workshop on Multimodal Data Analysis for Human Activity Detection and Understanding, Marina del Rey, CA, August 2016.
- *Human Activity Recognition from a Robot's Viewpoint*
VASC Seminar (RI), Carnegie Mellon University, Pittsburgh, PA, February 2016.
- *First-Person Activity Prediction*
Department Seminar (CS), University of Central Florida, Orlando, FL, February 2014.
- *First-Person Computer Vision – Understanding Egocentric Video Observation*
Department Seminar (CSE), Seoul National University, Seoul, Korea, May 2013.
- *Human Activity Recognition for Real-World Scenarios: Prediction and Cross-Domain Composition*
Institute Seminar (IRIS), University of Southern California, Los Angeles, CA, March 2012.
- *Computer Vision for Videos – From Objects to Events and Activities*
Department Seminar (CS), KAIST, Daejeon, Korea, May 2011.
- *Stochastic Representation and Recognition of High-level Group Activities*
International Workshop on Stochastic Image Grammars (SIG) with CVPR, Miami, FL, June 2009.

Articles/media

- *Improving Vision Transformer Efficiency and Accuracy by Learning to Tokenize*
Article, Google AI Blog, December 7, 2021.
- *Video Architecture Search*
Article, Google AI Blog, October 17, 2019.
- *Decoding the Language of Human Movements*
Interview, Communications of the ACM, Vol. 57, Issue 12, pages 12-14, December 2014.

Publications

Preprints

- [1] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, T. Jackson, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, K. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada, J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. Ryoo, G. Salazar, P. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, B. Zitkovich, “RT-1: Robotics Transformer for Real-World Control at Scale,” arXiv:2211.13224.
- [2] R. Burgert, K. Ranasinghe, X. Li, and M. S. Ryoo, “Peekaboo: Text to Image Diffusion Models Are Zero-Shot Segmentors,” arXiv:2211.13224.
- [3] J. Park, K. Kahatapitiya, D. Kim, S. Sudalairaj, Q. Fan, and M. S. Ryoo, “Grafting Vision Transformers,” arXiv:2210.15943.

Refereed conference publications

- [4] M. S. Ryoo, K. Gopalakrishnan, K. Kahatapitiya, T. Xiao, K. Rao, A. Stone, Y. Lu, J. Ibarz, and A. Arnab, “Token Turing Machines,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023. [*acceptance rate: 25.8%*]
- [5] B. Chen, F. Xia, B. Ichter, K. Rao, K. Gopalakrishnan, M. S. Ryoo, A. Stone, and D. Kappler, “Open-vocabulary Queryable Scene Representations for Real World Planning,” *IEEE International Conference on Robotics and Automation (ICRA)*, May 2023.
- [6] A. Wu and M. S. Ryoo, “Energy-Based Models for Cross-Modal Localization using Convolutional Transformers,” *IEEE International Conference on Robotics and Automation (ICRA)*, May 2023.
- [7] A. Zeng, M. Attarian, B. Ichter, K. Choromanski, A. Wong, S. Welker, F. Tombari, A. Purohit, M. Ryoo, V. Sindhwani, J. Lee, V. Vanhoucke, and P. Florence, “Socratic Models: Composing Zero-Shot Multimodal Reasoning with Language,” *International Conference on Learning Representations (ICLR)*, May 2023. [*acceptance rate: ~30%*]
- [8] K. Kahatapitiya, Z. Ren, H. Li, Z. Wu, M. S. Ryoo, “Self-supervised pretraining with classification labels for temporal activity detection,” *AAAI Conference on Artificial Intelligence (AAAI)*, February 2023. [*acceptance rate: 19.6%*]
- [9] S. Das and M. S. Ryoo, “ViewCLR: Learning Self-Supervised Video Representation for Unseen Viewpoints,” *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023.
- [10] R. Burgert, J. Shang, X. Li, and M. S. Ryoo, “Neural Neural Textures Make Sim2Real Consistent,” *Conference on Robot Learning (CoRL)*, December 2022.
- [11] J. Shang, S. Das, and M. S. Ryoo, “Learning Viewpoint-Agnostic Visual Representations by Recovering Tokens in 3D Space,” *Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS)*, December 2022. [*acceptance rate: 25.6%*]
- [12] X. Li, J. Shang, S. Das, and M. S. Ryoo, “Does Self-supervised Learning Really Improve Reinforcement Learning from Pixels?,” *Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS)*, December 2022. [*acceptance rate: 25.6%*]

- [13] J. Shang and M. S. Ryoo, “StARformer: Transformer with State-Action-Reward Representations,” *European Conference on Computer Vision (ECCV)*, October 2022. [*acceptance rate: 28%*]
- [14] A. Piergiovanni, K. Morton, W. Kuo, M. S. Ryoo, and A. Angelova, “Video Question Answering with Iterative Video-Text Co-Tokenization,” *European Conference on Computer Vision (ECCV)*, October 2022. [*acceptance rate: 28%*]
- [15] K. Ranasinghe, M. Naseer, S. Khan, F. S. Khan and M. S. Ryoo, “Self-supervised Video Transformer,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022. [*oral acceptance rate: ~4%*]
- [16] R. Dai, S. Das, K. Kahatapitiya, M. S. Ryoo, and F. Bremond, “MS-TCT: Multi-Scale Temporal ConvTransformer for Action Detection,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022. [*acceptance rate: 25.3%*]
- [17] K. Choromanski, H. Chen, H. Lin, Y. Ma, A. Sehanobish, D. Jain, M. S. Ryoo, J. Varley, A. Zeng, V. Likhoshesterov, D. Kalashnikov, V. Sindhvani, A. Weller, “Hybrid Random Features,” *International Conference on Learning Representations (ICLR)*, April 2022. [*acceptance rate: 32.3%*]
- [18] M. S. Ryoo, A. Piergiovanni, A. Arnab, M. Dehghani, and A. Angelova, “TokenLearner: Adaptive Space-Time Tokenization for Videos,” *Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS)*, December 2021. [*acceptance rate: 26%*]
- [19] A. Piergiovanni, V. Casser, M. S. Ryoo, and A. Angelova, “4D-Net for Learned Multi-Modal Alignment,” *International Conference on Computer Vision (ICCV)*, October 2021. [*acceptance rate: 25.9%*]
- [20] J. Shang and M. S. Ryoo, “Self-Supervised Disentangled Representation Learning for Third-Person Imitation Learning,” *International Conference on Intelligent Robots and Systems (IROS)*, September 2021.
- [21] K. Kahatapitiya and M. S. Ryoo, “Coarse-Fine Networks for Temporal Activity Detection in Videos,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021. [*acceptance rate: 23.7%*]
- [22] A. Piergiovanni and M. S. Ryoo, “Recognizing Actions in Videos from Unseen Viewpoints,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021. [*acceptance rate: 23.7%*]
- [23] I. Akinola, A. Angelova, Y. Lu, Y. Chebotar, D. Kalashnikov, J. Varley, J. Ibarz, and M. S. Ryoo, “Visionary: Vision Architecture Discovery for Robot Learning,” *IEEE International Conference on Robotics and Automation (ICRA)*, May 2021.
- [24] A. Piergiovanni and M. S. Ryoo, “AViD Dataset: Anonymized Videos from Diverse Countries,” *Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS)*, December 2020. [*acceptance rate: 20.1%*]
- [25] A. Piergiovanni, A. Angelova, A. Toshev, and M. S. Ryoo, “Adversarial Generative Grammars for Human Activity Prediction,” *European Conference on Computer Vision (ECCV)*, August 2020. [*oral acceptance rate: 2.8%*]
- [26] M. S. Ryoo, A. Piergiovanni, J. Kangaspunta, and A. Angelova, “AssembleNet++: Assembling Modality Representations via Attention Connections,” *European Conference on Computer Vision (ECCV)*, August 2020. [*acceptance rate: 27.1%*]
- [27] X. Gu, W. Luo, M. S. Ryoo, and Y. J. Lee, “Password-conditioned Anonymization and Deanonimization with Face Identity Transformers,” *European Conference on Computer Vision (ECCV)*, August 2020. [*acceptance rate: 27.1%*]
- [28] X. Wang, X. Xiong, M. Neumann, A. Piergiovanni, M. S. Ryoo, A. Angelova, K. M. Kitani, and W. Hua, “AttentionNAS: Spatiotemporal Attention Cell Search for Video Classification,” *European Conference on Computer Vision (ECCV)*, August 2020. [*acceptance rate: 27.1%*]
- [29] A. Piergiovanni, A. Angelova, and M. S. Ryoo, “Evolving Losses for Unlabeled Video Representation Learning,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. [*oral acceptance rate: 5.7%*]

- [30] M. S. Ryoo, A. Piergiovanni, M. Tan, and A. Angelova, “AssembleNet: Searching for Multi-Stream Neural Connectivity in Video Architectures,” *International Conference on Learning Representations (ICLR)*, April 2020. [*acceptance rate: 26.5%*]
- [31] A. Piergiovanni and M. S. Ryoo, “Unseen Action Recognition with Multimodal Learning,” *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020.
- [32] A. Piergiovanni, A. Angelova, and M. S. Ryoo, “Differentiable Grammars for Videos,” *AAAI Conference on Artificial Intelligence (AAAI)*, February 2020. [*oral acceptance rate: ~5%*]
- [33] A. Piergiovanni, A. Wu, and M. S. Ryoo, “Learning Real-World Robot Policies by Dreaming,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, November 2019.
- [34] M. U. Kim, H. Lee, H. J. Yang, and M. S. Ryoo, “Privacy-Preserving Robot Vision with Anonymized Faces by Extreme Low Resolution,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, November 2019.
- [35] A. Wu, A. Piergiovanni, and M. S. Ryoo, “Model-based Behavioral Cloning with Future Image Similarity Learning,” *Conference on Robot Learning (CoRL)*, October 2019. [*acceptance rate: 27.6%*]
- [36] A. Piergiovanni, A. Angelova, A. Toshev, and M. S. Ryoo, “Evolving Space-Time Neural Architectures for Videos,” *International Conference on Computer Vision (ICCV)*, October 2019. [*acceptance rate: 25.0%*]
- [37] A. Piergiovanni and M. S. Ryoo, “Temporal Gaussian Mixture Layer for Videos,” *International Conference on Machine Learning (ICML)*, Long Island, CA, June 2019. [*acceptance rate: 22.6%*]
- [38] A. Piergiovanni and M. S. Ryoo, “Early Detection of Injuries in MLB Pitchers from Video,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Island, CA, June 2019.
- [39] A. Piergiovanni and M. S. Ryoo, “Representation Flow for Action Recognition,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Island, CA, June 2019. [*acceptance rate: 25.2%*]
- [40] Z. Ren, Y. J. Lee, and M. S. Ryoo, “Learning to Anonymize Faces for Privacy Preserving Action Detection,” *European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018. [*acceptance rate: 31.8%*]
- [41] M. Xu, C. Fan, Y. Wang, M. S. Ryoo, and D. J. Crandall, “Joint Person Segmentation and Identification in Synchronized First- and Third-person Videos,” *European Conference on Computer Vision (ECCV)*, Munich, Germany, September 2018. [*acceptance rate: 31.8%*]
- [42] C. Fan, J. Lee, and M. S. Ryoo, “Forecasting Hands and Objects in Future Frames”, *European Conference on Computer Vision Workshops (ECCVW)*, Munich, Germany, September 2018.
- [43] A. Piergiovanni and M. S. Ryoo, “Learning Latent Super-Events to Detect Multiple Activities in Videos,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, June 2018. [*acceptance rate: 29.6%*]
- [44] A. Piergiovanni and M. S. Ryoo, “Fine-grained Activity Recognition in Baseball Videos,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, June 2018.
- [45] M. S. Ryoo, K. Kim, and H. J. Yang, “Extreme Low Resolution Activity Recognition with Multi-Siamese Embedding Learning,” *AAAI Conference on Artificial Intelligence (AAAI)*, New Orleans, LA, February 2018. [*acceptance rate: 24.6%*]
- [46] J. Lee and M. S. Ryoo, “Learning Robot Activities from First-Person Human Videos Using Convolutional Future Regression,” *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, September 2017.
- [47] I. Gori, J. K. Aggarwal, L. Matthies, and M. S. Ryoo, “Multi-Type Activity Recognition from a Robot's Viewpoint,” *the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, Melbourne, Australia, August 2017 (invited). [*acceptance rate: 26.0%*]

- [48] C. Fan, J. Lee, M. Xu, K. K. Singh, Y. J. Lee, D. J. Crandall, and M. S. Ryoo, “Identifying First-person Camera Wearers in Third-person Videos,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 2017. [*acceptance rate: 29.2%*]
- [49] T. Shu, X. Gao, M. S. Ryoo, and S.-C. Zhu, “Learning Social Affordance Grammar from Videos: Transferring Human Interactions to Human-Robot Interactions,” *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, May 2017.
- [50] M. S. Ryoo, B. Rothrock, C. Fleming, and H. J. Yang, “Privacy-Preserving Human Activity Recognition from Extreme Low Resolution,” *AAAI Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, February 2017. [*acceptance rate: 24.6%*]
- [51] A. Piergiovanni¹, C. Fan¹, and M. S. Ryoo, “Learning Latent Sub-events in Activity Videos Using Temporal Attention Filters,” *AAAI Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, February 2017 (¹equal contribution). [*acceptance rate: 24.6%*]
- [52] T. Shu, M. S. Ryoo, and S.-C. Zhu, “Learning Social Affordance for Human-Robot Interaction,” *the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, New York City, NY, July 2016. [*acceptance rate: 24%*]
- [53] M. S. Ryoo, B. Rothrock, and L. Matthies, “Pooled Motion Features for First-Person Videos,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, June 2015. [*acceptance rate: 28%*]
- [54] M. S. Ryoo, T. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies, “Robot-Centric Activity Prediction from First-Person Videos: What Will They Do to Me?,” *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Portland, OR, March 2015. [*acceptance rate: 25.4%*]
Best Paper Award Nominee
- [55] L. Xia, I. Gori, J. K. Aggarwal, and M. S. Ryoo, “Robot-Centric Activity Recognition from First-Person RGB-D Videos,” *IEEE Winter Conference on Applications of Computer Vision (WACV)*, HI, January 2015. [*1st-round acceptance rate: 30.0%*]
- [56] Y. Iwashita, A. Takamine, R. Kurazume, and M. S. Ryoo, “First-Person Animal Activity Recognition from Egocentric Videos,” *International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, August 2014.
- [57] Y. Iwashita¹, M. S. Ryoo¹, T. J. Fuchs, and C. Padgett, “Recognizing Humans in Motion: Trajectory-based Aerial Video Analysis,” *British Machine Vision Conference (BMVC)*, Bristol, U.K., September 2013 (¹equal contribution). [*acceptance rate: 29.8%*]
- [58] M. S. Ryoo and L. Matthies, “First-Person Activity Recognition: What Are They Doing to Me?,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, June 2013. [*acceptance rate: 26.2%*]
- [59] J. H. Joung, M. S. Ryoo, S. Choi, and S. R. Kim, “Reliable Object Detection and Segmentation Using Inpainting,” *IEEE/RSJ International Intelligent Robots and Systems (IROS)*, Algarve, Portugal, October 2012.
- [60] M. S. Ryoo, “Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos,” *International Conference on Computer Vision (ICCV)*, Barcelona, Spain, November 2011. [*acceptance rate: 23.7%*]
- [61] M. S. Ryoo, “Interactive Learning of Human Activities Using Active Video Composition,” *International Workshop on Stochastic Image Grammars (SIG)*, in *Proceedings of International Conference on Computer Vision Workshops (ICCVW)*, Barcelona, Spain, November 2011.
- [62] J. H. Joung, M. S. Ryoo, S. Choi, W. Yu, and H. Chae, “Background-aware Pedestrian/Vehicle Detection System for Driving Environments,” *IEEE Conference on Intelligent Transportation Systems (ITSC)*, Washington, D.C., October 2011.
- [63] M. S. Ryoo and W. Yu, “One Video is Sufficient? Human Activity Recognition Using Active Video Composition,” *IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, Hawaii, January 2011.

- [64] M. S. Ryoo, J. Lee, J. H. Joung, S. Choi, and W. Yu, "Personal Driving Diary: Constructing a Video Archive of Everyday Driving Events," *IEEE Workshop on Applications of Computer Vision (WACV)*, Kona, Hawaii, January 2011.
- [65] M. S. Ryoo, J. H. Joung, S. Choi, and W. Yu, "Incremental Learning of Novel Activity Categories from Videos," *the 16th International Conference on Virtual Systems and Multimedia (VSMM)*, Seoul, Korea, October 2010 (invited).
- [66] M. S. Ryoo, C.-C. Chen, J. K. Aggarwal, and A. Roy-Chowdhury, "An Overview of Contest on Semantic Description of Human Activities (SDHA) 2010," *International Conference on Pattern Recognition (ICPR) Contests*, Istanbul, Turkey, August 2010. [*acceptance rate: 38%*]
- [67] M. S. Ryoo¹, J. T. Lee¹, and J. K. Aggarwal, "Video Scene Analysis of Interactions between Humans and Vehicles Using Event Context," *ACM International Conference on Image and Video Retrieval (CIVR)*, Xian, China, July 2010 (invited, ¹equal contribution). [*oral acceptance rate: 10.5%*]
- [68] J. T. Lee, M. S. Ryoo, and J. K. Aggarwal, "View Independent Recognition of Human-Vehicle Interactions Using 3-D Models," *IEEE Workshop on Motion and Video Computing (WACV/WMVC)*, Snowbird, UT, December 2009.
- [69] M. S. Ryoo and J. K. Aggarwal, "Spatio-Temporal Relationship Match: Video Structure Comparison for Recognition of Complex Human Activities," *International Conference on Computer Vision (ICCV)*, Kyoto, Japan, October 2009. [*acceptance rate: 23.2%*]
- [70] M. S. Ryoo and J. K. Aggarwal, "Human Activities: Handling Uncertainties Using Fuzzy Time Intervals," *International Conference on Pattern Recognition (ICPR)*, Tampa, FL, December 2008.
- [71] M. S. Ryoo and J. K. Aggarwal, "Observe-and-Explain: A New Approach for Multiple Hypotheses Tracking of Humans and Objects," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, June 2008. [*acceptance rate: 31.6%*]
- [72] M. S. Ryoo and J. K. Aggarwal, "Recognition of High-level Group Activities Based on Activities of Individual Members," *IEEE Workshop on Motion and Video Computing (WACV/WMVC)*, Copper Mountain, CO, January 2008. [*oral acceptance rate: 33.3%*]
- [73] J. T. Lee, M. S. Ryoo, M. Riley, and J. K. Aggarwal, "Real-time Detection of Illegally Parked Vehicles using 1-D Transformation," *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, London, UK, September 2007.
- [74] M. Bhargava, C.-C. Chen, M. S. Ryoo, and J. K. Aggarwal, "Detection of Abandoned Objects in Crowded Environments," *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, London, UK, September 2007.
- [75] M. S. Ryoo and J. K. Aggarwal, "Hierarchical Recognition of Human Activities Interacting with Objects," *International Workshop on Semantic Learning Applications in Multimedia (SLAM)*, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN, June 2007.
- [76] M. S. Ryoo and J. K. Aggarwal, "Robust Human-Computer Interaction System Guiding a User by Providing Feedback," *the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, January 2007. [*acceptance rate: 34.7%*]
- [77] M. S. Ryoo and J. K. Aggarwal, "Semantic Understanding of Continued and Recursive Human Activities," *International Conference on Pattern Recognition (ICPR)*, Vol. 1, pp. 379~382, Hong Kong, August 2006.
- [78] M. S. Ryoo and J. K. Aggarwal, "Recognition of Composite Human Activities through Context-Free Grammar based Representation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 1709-1719, New York City, NY, June 2006. [*acceptance rate: 28.1%*]
- [79] H. S. Yang, Y. Seo, M. S. Ryoo, and H. Jung, "Affective Communication System with Emotional Memories for Multimodal Interaction with Humanoids," *the 11th International Conference on Virtual Systems and Multimedia (VSMM)*, October 2005.
- [80] D. Pardoe, M. Ryoo, and R. Miikkulainen, "Evolving Neural Network Ensembles for Control Problems," *Genetic and Evolutionary Computation Conference (GECCO)*, Washington, D.C., June 2005.

- [81] M. S. Ryoo, Y. Seo, H. Jung, and H. S. Yang, “Affective Dialogue Communication System with Emotional Memories for Humanoid Robots,” *International Conference on Affective Computing and Intelligent Interaction (ACII)*, LNCS 3784, pp. 819-827, October 2005.
- [82] H. Jung, Y. Seo, M. S. Ryoo, and H. S. Yang, “Affective Communication System with Multimodality for Humanoid Robot AMI,” *IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids)*, Los Angeles, CA, November 2004.

Journal publications

- [83] J. Shang, X. Li, K. Kahatapitiya, Y.-C. Lee*, and M. S. Ryoo*, “StARformer: Transformer with State-Action-Reward Representations for Robot Learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, early access, September 2022.
- [84] A. Piergiovanni, A. Angelova*, and M. S. Ryoo, “Tiny Video Networks,” *Applied AI Letters*, October 2021.
- [85] A. Wu*, A. Piergiovanni, and M. S. Ryoo, “Model-Based Robot Imitation with Future Image Similarity,” *International Journal of Computer Vision (IJCV)*, 2019.
- [86] R. Hadidi*, J. Cao, M. Woodward, M. S. Ryoo, and H. Kim, “Distributed Perception by Collaborative Robots,” *IEEE Robotics and Automation Letters (RA-L)*, 2018. [[IROS 2018 presentation](#)]
- [87] M. S. Ryoo* and L. Matthies, “First-Person Activity Recognition: Feature, Temporal Structure, and Prediction,” *International Journal of Computer Vision (IJCV)*, 119(3):307–328, 2016.
- [88] I. Gori, J. K. Aggarwal, L. Matthies, and M. S. Ryoo*, “Multi-Type Activity Recognition in Robot-Centric Scenarios,” *IEEE Robotics and Automation Letters (RA-L)*, 1(1):593-600, 2016. [[ICRA 2016 presentation](#)]
ICRA 2016 Best Paper Award in Robot Vision
- [89] M. S. Ryoo*, S. Choi¹, J. H. Joung¹, J.-Y. Lee¹, and W. Yu, “Personal Driving Diary: Automated Recognition of Driving Events from First-Person Videos,” *Computer Vision and Image Understanding (CVIU)*, 117(10): 1299-1312, October 2013 (¹equal contribution).
- [90] J. K. Aggarwal* and M. S. Ryoo, “Toward a Unified Framework of Motion Understanding,” *Image and Vision Computing (ImaVis)*, 30(8):465-466, August 2012.
- [91] M. S. Ryoo* and J. K. Aggarwal, “Stochastic Representation and Recognition of High-level Group Activities,” *International Journal of Computer Vision (IJCV)*, 93(2):183-200, June 2011.
- [92] J. K. Aggarwal and M. S. Ryoo*, “Human Activity Analysis: A Review,” *ACM Computing Surveys (CSUR)*, 43(3), April 2011.
- [93] M. S. Ryoo*, K. Grauman, and J. K. Aggarwal, “A Task-Driven Intelligent Workspace System to Provide Guidance Feedback,” *Computer Vision and Image Understanding (CVIU)*, 114(5):520-534, May 2010.
- [94] J. T. Lee*, M. S. Ryoo, M. Riley, and J. K. Aggarwal, “Real-time Illegal Parking Detection in Outdoor Environments Using 1-D Transformation,” *IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)*, 19(7):1014-1024, July 2009.
- [95] M. Bhargava, C.-C. Chen*, M. S. Ryoo, and J. K. Aggarwal, “Detection of Object Abandonment Using Temporal Logic,” *Machine Vision and Applications (MVA)*, 20(5):271-281, June 2009.
- [96] M. S. Ryoo* and J. K. Aggarwal, “Semantic Representation and Recognition of Continued and Recursive Human Activities,” *International Journal of Computer Vision (IJCV)*, 82(1), 1-24, April 2009.

Theses

- “Semantic Representation and Recognition of Human Activities,” Ph.D. Dissertation, the University of Texas at Austin, August 2008.
Outstanding Dissertation Award Nominee

* Corresponding author

- “Semantic Understanding of Continued and Recursive Activities using Context-Free Grammar,” M.S. Thesis, the University of Texas at Austin, August 2006.
Outstanding Thesis Award Nominee
- “Affective Dialogue Communication System with Emotional Memories for Humanoid Robots,” B.S. Thesis, Korea Advanced Institute of Science and Technology (KAIST), August 2004.

Other Awards and Honors

- **Best Paper Award**
CVPR Workshop on Deep Learning for Robot Vision (DLRV), 2017 (sponsored by Google/Facebook/ACRV).
Lee & Ryoo, “Learning Robot Activities from First-Person Human Videos Using Convolutional Future Regression”
- **Best Poster Award**
The 10th Joint Workshop on Machine Perception and Robotics (MPR), Beijing, Oct. 2014.
Iwashita, Takamine, Kurazume & Ryoo, “First-Person Animal Activity Recognition from Egocentric Videos”
- **Best Video Award** (sponsored by *IEEE RO-MAN*)
The 6th Korea Robotics Society Annual Conference (KRoC), 2011.
Ryoo et al., “Personal Driving Diary: Constructing a Video Archive of Everyday Driving Events”
- **Outstanding Dissertation/Thesis Award Nominee, 2007 and 2009**
The only candidate nominated by the Department of ECE, the University of Texas at Austin.
- **UT Engineering Doctoral Fellowship, 2006.9 ~ 2008.8**
Full tuition and \$10,000 annual supplemental stipend to support research (3 years granted).
Supported by the College of Engineering, the University of Texas at Austin.
- **David Bruton Jr. Graduate School Fellowship, 2006**
- **Korea Foundation for Advanced Studies Fellowship** (supported by *SK*), 2004.8 ~ 2008.8
Full scholarship awarded for tuition and living expenses, \$50,000 annually, for Ph.D. study (5 years granted).
One of the five recipients selected from nationwide (South Korea) in Computer Science.
- **Professional Development Award, the University of Texas at Austin, 2006 and 2007**
- **KAIST Undergraduate Scholarship, 2000.3 ~ 2004.1**
Full scholarship awarded for tuition and living expenses.

Professional Activities

Organizer/Chair

- Local organizing chair, ACM/IEEE International Conference on Human-Robot Interaction (HRI) 2018
- Lead organizer, the 4th Workshop on Egocentric (First-Person) Vision, with CVPR 2016
- Organizer, the 3rd Workshop on Egocentric (First-Person) Vision, with CVPR 2014
- Lead organizer, ICPR Contest on Semantic Description of Human Activities (SDHA), with ICPR 2010

Program committee member

- Area chair, International Conference on Computer Vision (ICCV), 2023
- Area chair, International Conference on Machine Learning (ICML), 2023
- Area chair, the Thirty-Sixth Annual Conference on Neural Information Processing (NeurIPS), 2022
- Area chair, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022
- Area chair, International Conference on Learning Representations (ICLR), 2022
- Area chair, International Conference on Computer Vision (ICCV), 2021
- Area chair, International Conference on Learning Representations (ICLR), 2021

- Area chair, International Conference on Learning Representations (ICLR), 2020
- Area chair, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019
- Area chair, IAPR International Conference on Machine Vision Applications (MVA), 2017
- Area chair, IEEE Winter Conference on Applications of Computer Vision (WACV), 2016

Journal reviewer

- IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), International Journal of Computer Vision (IJCV), Computer Vision and Image Understanding (CVIU), IEEE Transactions on Image Processing (T-IP), IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT), Image and Vision Computing Journal (ImaVis), IEEE Transactions on Systems, Man and Cybernetics (SMC), etc.

Advising

Past students

- Alan Wu, Department of ISE, Indiana University (Ph.D., 2023), joined MIT Lincoln Lab
- AJ Piergiovanni, Department of CS, Indiana University (Ph.D., 2020), joined Google Brain
- Alex Seewald, Department of CS, Indiana University (M.S., 2018), joined Cerner Corporation
- Maria Soledad Elli, Department of CS, Indiana University (M.S., 2017), joined Intel

Current students

- Cristina Mata, Department of CS, Stony Brook University (Ph.D. student)
- Kumara Kahatapitiya, Department of CS, Stony Brook University (Ph.D. student)
- Jinghuan Shang, Department of CS, Stony Brook University (Ph.D. student)
- Xiang Li, Department of CS, Stony Brook University (Ph.D. student)
- Jongwoo Park, Department of CS, Stony Brook University (Ph.D. student)
- Ryan Burgert, Department of CS, Stony Brook University (Ph.D. student)
- Kanchana Ranasinghe, Department of CS, Stony Brook University (Ph.D. student)

Ph.D. defense/proposal committee member of

- Xiaofang Wang, CMU (Ph.D. 2022)
- Srijan Das, INRIA (Ph.D. 2020)
- Chenyou Fan, Department of CS, Indiana University (Ph.D. 2018)
- Sven Bambach, Department of CS, Indiana University (Ph.D. 2016)
- Stefan Lee, Department of CS, Indiana University (Ph.D. 2016)
- Josh Harguess, Department of ECE, the University of Texas at Austin (Ph.D. 2011)
- Birgi Tamersoy, Department of ECE, the University of Texas at Austin (M.S. 2009)

Teaching Experience

- **CSE 525 Robotics, Stony Brook University:** Spring 2020/2021/2022/2023,
Instructor: M. S. Ryoo
Scope: an introductory Robot Learning course for graduate students focusing on deep reinforcement learning.
- **CSE 527 Computer Vision, Stony Brook University:** Fall 2021,
Instructor: M. S. Ryoo
Scope: an introductory Computer Vision course for graduate students.
- **CS/INFO B490/I400 Intro to Computer Vision, Indiana University Bloomington:** Spring 2016/2017/2018,
Instructor: M. S. Ryoo
Scope: an introductory Computer Vision course for undergraduate students.

- ***CS/INFO B659/I590 Vision for Intelligent Robotics, Indiana University Bloomington:*** Fall 2015/2016/2017,
Instructor: M. S. Ryoo
Scope: a graduate seminar course on state-of-the-art Computer Vision algorithms and their applications to Robotics.
- ***ME/CS 132a Introduction to Vision-based Robot Navigation, California Institute of Technology:*** Winter 2015,
Instructors: L. Matthies, R. Brockers, B. Rothrock, T. Fuchs, S. Weiss, and M. S. Ryoo
Scope: current topics in robotics research in the areas of autonomous navigation and vision, including perception.