Rogerio Schmidt Feris

July 2026

MIT-IBM Computing Research Lab
IBM Research
314 Main St
Cambridge, MA 02142, USA

Tel: 914-255-0218
rsferis@us.ibm.com
http://www.rogerioferis.org

Education

Ph.D. degree in Computer Science (2002-2006)
University of California, Santa Barbara, USA
(Advised by Prof. Matthew Turk)

M.S. degree in Computer Science (1999-2001)
University of Sao Paulo, Brazil
(Advised by Prof. Roberto Cesar)

B.S. degree in Computer Engineering (1994-1998)
University of Rio Grande (FURG), Brazil

Professional Experience

- **Principal Scientist & Senior Manager, MIT-IBM Computing Research Lab, MA (2020-present)**
  - I am currently leading a team of AI researchers, with a focus on enhancing large language models with multimodal capabilities. Notable recent achievements include:

    - **Granite Vision**: We built a lightweight large language model with vision capabilities, specifically designed to excel in enterprise use cases, particularly in visual document understanding. Our model garnered significant attention, with over **3 million downloads** as of July 2026 on Hugging Face and Ollama platforms.

    - **Generative AI Commentary**: My team successfully trained a large-scale multimodal model that generated narrations for video highlights of the 2023 Wimbledon and US Open tennis tournaments**.** This project captured international media attention, with coverage by major outlets, such as CNN, ESPN, and Fox News, showcasing the potential of AI-driven commentary in sports broadcasting.

  - I led the IBM-MIT team as a principal investigator for **DARPA Learning with Less Labels** (LwLL), together with Prof. Josh Tenenbaum. I also led the IBM-MIT-Purdue team as a principal investigator for **IARPA DIVA** (with Prof. Jeff Siskind and Prof. Aude Oliva). We have achieved quite strong results in the evaluation benchmarks for both LwLL and DIVA, **ranking 1st** for most of the tasks among all other performers and external institutions.

- **Research Manager, IBM T. J. Watson Research Center, NY (2016-2020)**
  - I led the research team that created a system called Cognitive Highlights, which was used to generate the **official highlights of the US Open/Wimbledon/Masters**

tournaments, for the first time in history. The highlights were seen by millions of fans worldwide! This work was covered by the New York Times and many other media outlets. It was also integrated as a product offering of Watson Media. We received a prestigious **Tech Emmy award**, which recognizes the impact of this work.

- I served as the coordinator of the joint program between IBM Research and Watson Visual Recognition, driving efforts towards strong research and product deliverables for the Watson Vision Services. I was also a **global leader** for the multimodal perception area at IBM, coordinating efforts across labs worldwide and leading computer vision / AI challenges.

- **Research Scientist, IBM T. J. Watson Research Center, NY (2006-2016)**
  - I was one of the key inventors of the IBM Intelligent Video Analytics (IVA), a leading-edge camera-based system with automatic event detection and search capabilities. My work on video analysis for safety and security was published in high-profile conferences and featured at ABC news and CBS 60 minutes. Among other awards, I was named IBM Master Inventor and received an IBM Outstanding Innovation Achievement Award.

- **Affiliate Associate Professor, University of Washington, WA (2008-2016)**
  - I have co-advised PhD students at University of Washington, in close collaboration with Prof. Ming-Ting Sun. I co-advised the PhD dissertation of Haowei Liu and Jun Xie (both working at Google now).

- **Adjunct Associate Professor, Columbia University, NY (2008, 2013, 2014)**
  - I co-taught advanced graduate courses on "Visual Recognition and Search" at Columbia University (Electrical Engineering and Computer Science Departments).

- **Senior Software Engineer, IBM Global Technology Services, NY (2008)**
  - In addition to working on core research, I had a one-year assignment to IBM Global Technology Services as a senior software engineer, with the goal of helping the commercialization of the IBM Intelligent Video Analytics (IVA) system. My code has been running in various components of the product.

Research Interests

- My research interests are in multimodal AI (computer vision, language, speech, and sound), including large language and multimodal models and related topics such as long-horizon memory for self-evolving agents. More recently, I've become interested in the intersection of multimodal AI and quantum computing.

Selected Awards and Distinctions

- **IEEE Fellow**, Class of 2026. This honor is given to less than 0.1% of members annually.
- IBM Corporate Technical Award – "Rogerio Feris is honored for his pioneering work of science and engineering in Generative AI expressive commentary and lively narration workflows, which can be applied to live content creation, production and distribution", 2025.

- IBM Significant External Honor, a prize that recognizes individuals who have received among the highest honors and awards, granted by an eligible external group for the impact of their technical contributions to design, engineering, and science, 2024. (in reference to the Emmy Award below)

- **Technology and Engineering Emmy Award**, National Academy of Television Arts & Sciences, for the work "AI-ML curation of Sports Highlights". The Technology & Engineering Emmy® Awards recognizes engineering technologies that either represent so extensive an improvement on existing methods or are so innovative in nature that they materially have affected television. As a key innovator, Rogerio received an Emmy statuette and participated in the National Association of Broadcasters ceremony in Las Vegas, 2023.

- Notable Area Chair, NeurIPS 2023.

- Best paper award, Honorable Mention, IEEE Winter Conference on Applications of Computer Vision (WACV), for the paper "Select, Label, and Mix: Learning Discriminative Invariant Feature Representations for Partial Domain Adaptation", 2023.

- IBM Corporate Award for AI Video Enrichment and Editing (Auto-curation of Sports Highlights), 2019. This is a prestigious company-wide award which recognizes breakthrough achievements that have led to significant market and industry success for IBM.

- 1st place, TRECVID ActEV: Activities in Extended Video, 2018.
    - Activity Detection (AD) Task – 1st Place out of 30 entries.
    - Activity and Object Detection (AOD) Task – 1st Place out of 15 entries

- Best Digital Development, Yahoo Sports Tech Awards (Wimbledon Cognitive Highlights), 2018.

- Best paper runner up, CVPR Workshop on Computer Vision in Sports, for the work "Automatic Curation of Golf Highlights using Multimodal Excitement Features", 2018.

- Best papers of ECCV 2016 - IJCV Special Issue, which feature the work "A Recurrent Encoder-Decoder Network for Sequential Face Alignment", 2016.

- Best Poster Award, Greater New York Area Workshop on Multimedia and Vision, for the paper "Designing category-level attributes for discriminative visual recognition", 2013.

- Invited Speaker at the National Academy of Engineering JAFOE meeting, which gathered **sixty of the most promising engineers under the age of 45 from Japan and the United States**, 2012.

- TRECVid Surveillance Event Detection Competition (IBM-CMU-NUS team). I was part of the team that ranked 1st in 4 out of 7 event detection tasks, 2012.

- Named IBM Master Inventor, 2011. This title is reserved for leading inventors within IBM. Each year, only 1% of the IBM Research technical population receives this award.

- IBM Outstanding Innovation Achievement Award. IBM Research award received for contributions in physical security and video analytics, 2011.

- Far-Reaching Research Award – competitive grant from IBM Research to pursue research towards building the next-generation camera networks (with Ankur Datta, Lisa Brown, and Rick Kjeldsen), 2010.

- North American Frost & Sullivan Award for Video Surveillance Software Company of the Year, 2008. Prize given to IBM in recognition of our work in video surveillance.
- IBM Emerging Leader in Multimedia, July 2005. I was among eight students selected from top universities by IBM.
- Recipient of a University of California, Santa Barbara fellowship to pursue PhD research in the US (2002-2006)
- Best Computer Science MSc Thesis in Brazil - 2nd Prize (Nationwide Competition). Award received at the Congress of Brazilian Computer Society, 2002.
- IBM Research Accomplishment Awards
    - Dynamic Neural Networks for Efficient AI, 2020 (Outstanding Innovation Award)
    - Contributions to Science in the Learning with Less Labels (LwLL) Domain – Data Limited Learning for Vision and Beyond, 2020 (Outstanding Innovation Award)
    - Efficient and Robust Object Detection, 2018
    - CoreML, Customized Learning and De-Biasing for Watson Visual Recognition, 2018
    - Research Contributions to Watson Visual Recognition Services, 2017
    - Contributions to Safety & Physical Security through Video Analytics (O-level, Outstanding), 2010

Publications

[200+ papers, h-index 77 as of July 2026]

Books

1. R. Feris, C. Lampert, and D. Parikh (Eds.). Visual Attributes. Advances in Computer Vision and Pattern Recognition, Springer, 2016.

Large Multimodal Model Releases

1. L. Karlinsky, A. Arbelle, A. Daniels, A. Nassar, A. Alfassi, B. Wu, D. Joshi, E. Schwartz, J. Kondic, N. Shabtay, P. Li, R. Herzig, S. Abedin, S. Perek, S. Harary, U. Barzelay, A. Goldfarb, A. Oliva, B. Wieles, B. Bhattacharjee, B. Huang, C. Auer, D. Gutfreund, D. Beymer, D. Wood, H. Kuehne, J. Hansen, J. Shtok, K. Wong, L. Bathen, M. Mishra, M. Lysak, M. Dolfi, M. Yurochkin, N. Livathinos, N. Harel, O. Azulai, O. Naparstek, R. Teixeira de Lima, R. Panda, S. Doveh, S. Gupta, S. Das, S. Zawad, Y. Kim, Z. He, A. Brooks, G. Goodhart, A. Govindjee, D. Leist, I. Ibrahim, A. Soffer, D. Cox, K. Soule, L. Lastras, N. Desai, S. Ofek-koifman, S. Raghavan, T. Syeda-Mahmood, P. Staar, T. Drory, and R. Feris. **Granite Vision**: a lightweight, open-source multimodal model for enterprise Intelligence. IBM Technical Report, 2025. (**more than 3 million downloads** on HuggingFace/Ollama as of July 2026).

Peer-Reviewed Conference Publications

1. R. Feris, Y. Liu, P. Li, H. Hua, and D. Kremer. Aligning Quantum Operators with Large Language Models. IEEE International Conference on Quantum Computing and Engineering (**QCE 2026**).

2. J. Kondic, P. Li, D. Joshi, I. Sanchez, B. Wiesel, S. Abedin, A. Alfassy, E. Schwartz, D. Caraballo, Y. G. Cinar, F. Scheidegger, S. I. Ross, D. K. I. Weidele, H. Hua, E. Arutyunova, R. Herzig, Z. He, Z. Wang, X. Yu, Y. Zhao, S. Jiang, M. Liu, Q. Lin, P. Staar, L. Lastras, A. Oliva, and R. Feris. ChartNet: A Million-Scale, High-Quality Multimodal Dataset for Robust Chart Understanding. Conference on Computer Vision and Pattern Recognition (**CVPR 2026**).

3. K. Li, C. Shang, L. Karlinsky, R. Feris, T. Darrell, and R. Herzig. Latent Implicit Visual Reasoning. Conference on Computer Vision and Pattern Recognition (**CVPR 2026**).

4. A. Singh, S. Marjit, W. Lin, P. Gavrikov, S. Yeung-Levy, H. Kuehne, R. Feris, S. Doveh, J. Glass, and M. Jehanzeb Mirza. TTRV: Test-Time Reinforcement Learning for Vision Language Models. Conference on Computer Vision and Pattern Recognition (**CVPR 2026**).

5. Y. Tang, D. Shimada, H. Hua, C. Huang, J. Bi, R. Feris, and C. Xu. Video-R4: Reinforcing Text-Rich Video Reasoning with Visual Rumination. Conference on Computer Vision and Pattern Recognition (**CVPR 2026**, Findings).

6. Z. Yu, L. Nguyen, J. Park, X. Gu, Z. He, S. Lee, R. Feris, and Y. Lee. DocHop: Benchmarking Out-of-domain Multi-hop Reasoning in Information-Dense Documents. International Conference on Machine Learning (**ICML 2026**).

7. B. Huang, H. Hua, Z. Yu, T. Darrell, R. Feris, and R. Herzig. DAVE: A VLM Vision Encoder for Document Understanding and Web Agents. International Conference on Learning Representations (**ICLR 2026**).

8. X. Gu, J. Mao, Z. Hong, Z. Yu, P. Li, D. Joshi, R. Feris, and Z. He. Composition-Grounded Data Synthesis for Visual Reasoning. International Conference on Learning Representations (**ICLR 2026**).

9. L. Selch, Y. Hou, M. Jehanzeb Mirza, S. Doveh, J. Glass, R. Feris, and W. Lin. PRISMM-Bench: A Benchmark of Peer-Review Grounded Multimodal Inconsistencies. International Conference on Learning Representations (**ICLR 2026**).

10. T. Chai, C. Mitra, B. Huang, G. Gare, Z. Lin, A. Arbelle, L. Karlinsky, R. Feris, T. Darrell, D. Ramanan, and R. Herzig. Activation Reward Models for Few-shot Model Alignment. Association for Computational Linguistics (**ACL 2026**, Findings).

11. A. Rouditchenko, S. Bhati, E. Araujo, S. Thomas, H. Kuehne, R. Feris, and J. Glass. Omni-R1: Do You Really Need Audio to Fine-Tune Your Audio LLM? IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU 2025**)

12. C. Mitra, B. Huang, T. Chai, Z. Lin, A. Arbelle, R. Feris, L. Karlinsky, T. Darrell, D. Ramanan, and R. Herzig. Enhancing Few-Shot Vision-Language Classification with Large Multimodal Model Features. International Conference on Computer Vision (**ICCV 2025**).

13. S. Doveh, N. Shabtay, W. Lin, E. Schwartz, H. Kuehne, R. Giryes, R. Feris, L. Karlinsky, J. Glass, A. Arbelle, S. Ullman, and M. Jehanzeb Mirza. Teaching VLMs to Localize Specific Objects from In-context Examples. International Conference on Computer Vision (**ICCV 2025**).

14. S. Kumar Maharana, B. Zhang, L. Karlinsky, R. Feris, and Y. Guo. BATCLIP: Bimodal Online Test-Time Adaptation for CLIP. International Conference on Computer Vision (**ICCV 2025**).

15. E. Araujo, A. Rouditchenko, Y. Gong, S. Bhati, S. Thomas, B. Kingsbury, L. Karlinsky, R. Feris, J. Glass, and H. Kuehne. CAV-MAE Sync: Improving Contrastive Audio-Visual Mask Autoencoders via Fine-Grained Alignment. Conference on Computer Vision and Pattern Recognition (**CVPR 2025**).

16. Y. Wang, D. Krotov, Y. Hu, Y. Gao, W. Zhou, J. McAuley, D. Gutfreund, R.Feris, and Z. He. M+: Extending MemoryLLM with Scalable Long-Term Memory. International Conference on Machine Learning (**ICML 2025**).

17. J. Kang, L. Karlinsky, H. Luo, Z. Wang, J. Hansen, J. Glass, D. Cox, R. Panda, R. Feris, and A. Ritter. Self-MoE: Towards Compositional Large Language Models with Self-Specialized Experts. International Conference on Learning Representations (**ICLR 2025**).

18. I. Huang, W. Lin, M. Mirza, J. Hansen, S. Doveh, V. Butoi, R. Herzig, A. Arbelle, H. Kuehne, T. Darrell, C. Gan, A. Oliva, R. Feris, and L. Karlinsky. ConMe: Rethinking Evaluation of Compositional Reasoning for Modern VLMs. Conference on Neural Information Processing Systems (**NeurIPS 2024**).

19. R. Wang, S. Ghosh, D. Cox, D. Antognini, A. Oliva, R. Feris, and L. Karlinsky. Trans-LoRA: towards data-free Transferable Parameter Efficient Finetuning. Conference on Neural Information Processing Systems (**NeurIPS 2024**).

20. Z. He, L. Karlinsky, D. Kim, J. McAuley, D. Krotov, and R. Feris. CAMELoT: Towards Large Language Models with Training-Free Associative Memory. ICML 2024 Workshop on Long-Context Foundation Models, 2024.

21. J. Li, S. Das, A. Oliva, D. Krotov, L. Karlinsky, and R. Feris. Long Context Understanding using Self-Generated Synthetic Data. ICML 2024 Workshop on Long-Context Foundation Models, 2024.

22. B. Chen , N. Shvetsova , A. Rouditchenko, D. Kondermann, S. Thomas, S. Chang, R. Feris, J. Glass, and H. Kuehne. What, when, and where? Self-Supervised Spatio-Temporal Grounding in Untrimmed Multi-Action Videos from Narrated Instructions. Conference on Computer Vision and Pattern Recognition (**CVPR 2024**).

23. J. Smith, L. Valkov, S. Halbe, V. Gutta, R. Feris, Z. Kira, and L. Karlinsky. Adaptive Memory Replay for Continual Learning. CVPR Workshop on Efficient Large Vision Models, 2024.

24. A. Rouditchenko, Y. Gong, S. Thomas, L. Karlinsky, H. Kuehne, R. Feris, and J. Glass. Whisper-Flamingo: Integrating Visual Features into Whisper for Audio-Visual Speech Recognition and Translation (**Interspeech 2024**).

25. J. Kang, H. Luo, Y. Zhu, J. Hansen, J. Glass, D. Cox, A. Ritter, R. Feris, L. Karlinsky. Self-Specialization: Uncovering Latent Expertise within Large Language Models. Association for Computational Linguistics (**ACL 2024**, Findings).

26. B. Pan, R. Panda, S. Jin, R. Feris, A. Oliva, P. Isola, and Y. Kim. LangNav: Language as a Perceptual Representation for Navigation. North American Chapter of the Association for Computational Linguistics (**NAACL 2024**).

27. Baughman, S. Hammer, R. Agarwal, R. Feris, G. Akay, E. Morales, L. Karlinsky, and T. Johnson. Large Scale Generative AI Text Applied to Sports and Music. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2024**).

28. S. Bhati, Y. Gong, L. Karlinsky, H. Kuehne, R. Feris, and J. Glass. DASS: Distilled Audio State Space Models Are Stronger and More Duration-Scalable Learners. IEEE Spoken Language Technology Workshop, 2024.

29. X. Sun, R. Panda, C. Chen, N. Wang, B. Pan, K. Gopalakrishnan, A. Oliva, R. Feris, and K. Saenko. Improved Techniques for Quantizing Deep Networks with Adaptive Bit-Widths. IEEE Winter Conference on Applications of Computer Vision (**WACV 2024**)

30. Z. He, G. Blackwood, R. Panda, J. McAuley, R. Feris. Synthetic Pre-Training Tasks for Neural Machine Translation. Association for Computational Linguistics (**ACL 2023** Findings).

31. H. Zhong, S. Mishra, D. Kim, S. Jin, R. Panda, H. Kuehne, L. Karlinsky, V. Saligrama, A. Oliva, and R. Feris. Learning Human Action Recognition Representations Without Real Humans. Conference on Neural Information Processing Systems (**NeurIPS 2023** Datasets Track)

32. S. Doveh, A. Arbelle, S. Harary, R. Herzig, D. Kim, P. Cascante-Bonilla, A. Alfassy, R. Panda, R. Giryes, R. Feris, S. Ullman, and L. Karlinsky. Dense and Aligned Captions (DAC) Promote Compositional Reasoning in VL Models. Conference on Neural Information Processing Systems (**NeurIPS 2023, Spotlight**).

33. M. Mirza, L. Karlinsky, W. Lin, H. Possegger, M. Kozinski, R. Feris, and H. Bischof. LaFTer: Label-Free Tuning of Zero-shot Classifier using Language and Unlabeled Image Collections. Conference on Neural Information Processing Systems (**NeurIPS 2023**).

34. R. Herzig, A. Mendelson, L. Karlinsky, A. Arbelle, R. Feris, T. Darrell, A. Globerson. Incorporating Structured Representations into Pretrained Vision & Language Models Using Scene Graphs. Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**).

35. A. Rouditchenko, S. Khurana, S. Thomas, R. Feris, L. Karlinsky , H. Huehne, D. Harwath, B. Kingsbury, and J. Glass. Comparison of Multilingual Self-Supervised and Weakly-Supervised Speech Pre-Training for Adaptation to Unseen Languages (**Interspeech 2023**).

36. K. Wang. D. Kim, R. Feris, and M. Betke. CDAC: Cross-domain Attention Consistency in Transformer for Domain Adaptive Semantic Segmentation. International Conference on Computer Vision (**ICCV 2023**).

37. P. Cascante-Bonilla, K. Shehada, J. Smith, S. Doveh, D. Kim, R. Panda, G. Varol, A. Oliva, V. Ordonez, R. Feris, and L. Karlinsky. Going Beyond Nouns With Vision & Language Models Using Synthetic Data. International Conference on Computer Vision (**ICCV 2023**).

38. W. Lin, L. Karlinsky, N. Shvetsova, H. Possegger, M. Kozinski, R. Panda, R. Feris, H. Kuehne, and H. Bischof. MAtch, eXpand and Improve: Unsupervised Finetuning for Zero-Shot Action Recognition with Language Knowledge. International Conference on Computer Vision (**ICCV 2023**).

39. J. Smith, P. Cascante-Bonilla, A. Arbelle, D. Kim, R. Panda, D. Cox, D. Yang, Z. Kira, R. Feris, and L. Karlinsky. ConStruct-VL: Data-Free Continual Structured VL Concepts Learning. Conference on Computer Vision and Pattern Recognition (**CVPR 2023**).

40. J. Smith, L. Karlinsky, V. Gutta, P. Cascante-Bonilla, D. Kim, A. Arbelle, R. Panda, R. Feris, and Zsolt Kira. CODA-Prompt: COntinual Decomposed Attention-based Prompting for

Rehearsal-Free Continual Learning. Conference on Computer Vision and Pattern Recognition (**CVPR 2023**).

41. S. Doveh, A. Arbelle, S. Harary, R. Panda, R. Herzig, E. Schwartz, R. Giryes, R. Feris, S. Ullman, and L. Karlinsky. Teaching Structured Vision & Language Concepts to Vision & Language Models. Conference on Computer Vision and Pattern Recognition (**CVPR 2023**).

42. P. Wang, R. Panda, L. Hennigen, P. Greengard, L. Karlinsky, R. Feris, D. Cox, Z. Wang, and Y. Kim. Learning to Grow Pretrained Models for Efficient Transformer Training. International Conference on Learning Representations (**ICLR 2023, notable-top-25%).**

43. Z. Wang, R. Panda, L. Karlinsky, R. Feris, H. Sun, and Y. Kim. Multitask Prompt Tuning Enables Parameter-Efficient Transfer Learning. International Conference on Learning Representations (**ICLR 2023**).

44. A. Sahoo, R. Panda, R. Feris, K., and A. Das. Select, Label, and Mix: Learning Discriminative Invariant Feature Representations for Partial Domain Adaptation. IEEE Winter Conference on Applications of Computer Vision (**WACV 2023, Best Paper Award Honorable Mention**).

45. T. Li, L. Fan, Y. Yuan, H. He, Y. Tian, R. Feris, P. Indyk, and D. Katabi. Addressing Feature Suppression in Unsupervised Visual Representations. IEEE Winter Conference on Applications of Computer Vision (**WACV 2023**).

46. A., Y. Chuang, N. Shvetsova, S. Thomas, R. Feris, B. Kingsbury, L. Karlinsky, D. Harwath, H. Kuehne, and J. Glass. C2KD: Cross-Lingual Cross-Modal Knowledge Distillation for Multilingual Text-Video Retrieval. International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2023**).

47. J. Lee, Y. Bu, P. Sattigeri, R. Panda, G. Wornell, L. Karlinsky, and R. Feris. A maximal correlation approach to imposing fairness in machine learning. International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2022**).

48. M. Baradad, R. Chen, J. Wulff, T. Wang, R. Feris, A. Torralba, and P. Isola. Procedural Image Programs for Representation Learning. Conference on Neural Information Processing Systems (**NeurIPS 2022**).

49. Y. Kim, S. Mishra, S. Jin, R. Panda, H. Kuehne, L. Karlinsky, V. Saligrama, K. Saenko, A. Oliva, and R. Feris. How Transferable are Video Representations Based on Synthetic Data? Conference on Neural Information Processing Systems (**NeurIPS 2022** Dataset Track).

50. A. Alfassy, A. Arbelle, O. Halimi, S. Harary, R. Herzig, E. Schwartz, R. Panda, M. Dolfi, C. Auer, P. Staar, K. Saenko, R. Feris, L. Karlinsky. FETA: Towards Specializing Foundational Models for Expert Task Applications. Conference on Neural Information Processing Systems (**NeurIPS 2022** Dataset Track).

51. S. Mishra, R. Panda, C. Phoo, C. Chen, L. Karlinsky, K. Saenko, V. Saligrama, and R. Feris. Task2Sim: Towards Effective Pre-training and Transfer from Synthetic Data. Conference on Computer Vision and Pattern Recognition (**CVPR 2022**).

52. S. Harary, E. Schwartz, A. Arbelle, P. Staar, S. Abu-Hussein, E. Amrani, R. Herzig, A. Alfassy, R. Giryes, H. Kuehne, D. Katabi, K. Saenko, R. Feris, and L. Karlinsky. Unsupervised Domain Generalization by Learning a Bridge Across Domains. Conference on Computer Vision and Pattern Recognition (**CVPR 2022, Oral**).

53. P. Cascante-Bonilla, H. Wu, L. Wang, R. Feris, and V. Ordonez. SimVQA: Exploring Simulated Environments for Visual Question Answering. Conference on Computer Vision and Pattern Recognition (**CVPR 2022**).

54. Y. Li, R. Panda, Y. Kim, C. Chen, R. Feris, D. Cox, and N. Vasconcelos. VALHALLA: Visual Hallucination for Machine Translation. Conference on Computer Vision and Pattern Recognition (**CVPR 2022**).

55. N. Shvetsova, B. Chen, A. Rouditchenko, S. Thomas, B. Kingsbury, R. Feris, D. Harwath, J. Glass, and H. Kuehne. Everything at Once -- Multi-modal Fusion Transformer for Video Retrieval. Conference on Computer Vision and Pattern Recognition (**CVPR 2022**).

56. T. Li, P. Cao, Y. Yuan, L. Fan, Y. Yang, R. Feris, P. Indyk, and D. Katabi. Targeted Supervised Contrastive Learning for Long-Tailed Recognition. Conference on Computer Vision and Pattern Recognition (**CVPR 2022**).

57. B. Pan, Y. Jiang, R. Panda, Z. Wang, R. Feris, and A. Oliva. IA-RED^2: Interpretability-Aware Redundancy Reduction for Vision Transformers. Conference on Neural Information Processing Systems (**NeurIPS 2021**).

58. A. Islam. C. Chen, R. Panda, L. Karlinsky, R. Feris, and R. Radke. Dynamic Distillation Network for Cross-Domain Few-Shot Recognition with Unlabeled Data. Conference on Neural Information Processing Systems (**NeurIPS 2021**).

59. R. Panda, R. Chen, Q. Fan, X. Sun, K. Saenko, A. Oliva, and R. Feris. AdaMML: Adaptive Multi-Modal Learning for Efficient Video Recognition. International Conference on Computer Vision (**ICCV 2021**).

60. A. Islam, R. Chen, R. Panda, L. Karlinsky, R. Radke, and R. Feris. A Broad Study on the Transferability of Visual Representations with Contrastive Learning. International Conference on Computer Vision (**ICCV 2021**).

61. A. Arbelle, S. Doveh, A. Alfassy, J. Shtok, G. Lev, E. Schwartz, H. Kuehne, H. Levi, P. Sattigeri, R. Panda, R. Chen, A. Bronstein, K. Saenko, S. Ullman, R. Giryes, R. Feris, and L. Karlinsky. Detector-Free Weakly Supervised Grounding by Separation. International Conference on Computer Vision (**ICCV 2021, Oral**).

62. X. Sun, R. Panda, R. Chen, A. Oliva, R. Feris, and K. Saenko. Dynamic Network Quantization for Efficient Video Inference. International Conference on Computer Vision (**ICCV 2021**).

63. B. Chen, A. Rouditchenko, K. Duarte, H. Kuehne, S. Thomas, A. Boggust, R. Panda, B. Kingsbury, R. Feris, D. Harwath, J. Glass, M. Picheny, and S. F. Chang. Multimodal Clustering Networks for Self-supervised Learning from Unlabeled Videos. International Conference on Computer Vision (**ICCV 2021**).

64. A. Rouditchenko, A. Boggust, D. Harwath, B. Chen, D. Joshi, S.Thomas, K. Audhkhasi, H. Kuehne, R. Panda, R. Feris, B. Kingsbury, M. Picheny, A.Torralba, and J. Glass. AVLnet: Learning Audio-Visual Language Representations from Instructional Videos. Interspeech 2021

65. A. Rouditchenko, A. Boggust, D. Harwath, S. Thomas, H. Kuehne, B. Chen, R. Panda, R. Feris, B. Kingsbury, M. Picheny and J. Glass. Cascaded Multilingual Audio-Visual Learning from Videos. Interspeech 2021

66. G. Bukchin, E. Schwartz, K. Saenko, O. Shahar, R. Feris, R. Giryes, and L. Karlinsky. Fine-grained Angular Contrastive Learning with Coarse Labels. Conference on Computer Vision and Pattern Recognition (**CVPR 2021, Oral**).

67. H. Wu, Y. Gao, X. Guo, Z. Al-Halah, S. Rennie, K. Grauman, and R. Feris. Fashion IQ: A New Dataset Towards Retrieving Images by Natural Language Feedback. Conference on Computer Vision and Pattern Recognition (**CVPR 2021**).

68. S. Whitehead, H. Wu, H. Ji, R. Feris, and K. Saenko. Separating Skills and Concepts for Novel Visual Question Answering. Conference on Computer Vision and Pattern Recognition (**CVPR 2021**).

69. M. Monfort, S. Jin, D. Harwath, R. Feris, J. Glass, and A. Oliva. Spoken Moments: A Large Scale Dataset of Audio Descriptions of Dynamic Events in Video. Conference on Computer Vision and Pattern Recognition (**CVPR 2021**).

70. A. Singh, O. Chakraborty, A. Varshney, R. Panda, R. Feris, K. Saenko, and A. Das. Semi-Supervised Action Recognition with Temporal Contrastive Learning. Conference on Computer Vision and Pattern Recognition (**CVPR 2021**).

71. C. Chen, R. Panda, K. Ramakrishnan, R. Feris, J. Cohn, A. Oliva, and Q. Fan. Deep Analysis of CNN-based Spatio-temporal Representations for Action Recognition. Conference on Computer Vision and Pattern Recognition (**CVPR 2021**).

72. B. Pan, R. Panda, C. Fosco, C. Lin, A. Andonian, Y. Meng, K. Saenko, A. Oliva, and R. Feris. VA-RED^2: Video Adaptive Redundancy Reduction. International Conference on Learning Representations (**ICLR 2021**).

73. Y. Meng, R. Panda, C. Lin, P. Sattigeri, L. Karlinsky, K. Saenko, A. Oliva, and R. Feris. AdaFuse: Adaptive Temporal Fusion Network for Efficient Action Recognition. International Conference on Learning Representations (**ICLR 2021**).

74. L. Karlinsky, J. Shtok, A. Alfassy, M. Lichtenstein, S. Harary, E. Schwartz, S. Doveh, P. Sattigeri, R. Feris, A. Bronstein, and R. Giryes. StarNet: towards Weakly Supervised Few-Shot Object Detection. AAAI Conference on Artificial Intelligence (**AAAI 2021**).

75. R. Panda, M. Merler, M. Jaiswal, H. Wu, K. Ramakrishnan, U. Finkler, C. Chen, M. Cho, R. Feris, D. Kung, and B. Bhattacharjee. NASTransfer: Analyzing Architecture Transferability in Large Scale Neural Architecture Search. AAAI Conference on Artificial Intelligence (**AAAI 2021**).

76. U. Finkler, M. Merler, R. Panda, M. Jaiswal, H. Wu, K. Ramakrishnan, C. Chen, M. Cho, D. Kung, R. Feris, and B. Bhattacharjee. Large Scale Neural Architecture Search with Polyharmonic Splines. AAAI Workshop on Meta-Learning for Computer Vision, 2021.

77. X. Sun, R. Panda, R. Feris, and K. Saenko. AdaShare: Learning What to Share for Efficient Deep Multi-Task Learning. Conference on Neural Information Processing Systems (**NeurIPS 2020**).

78. Y. Guo, N. Codella, L. Karlinsky, J. Codella, J. Smith, K. Saenko, T. Rosing, and R. Feris. A Broader Study of Cross-Domain Few-Shot Learning. European Conference on Computer Vision (**ECCV 2020**).

79. Y. Meng, C. Lin, R. Panda, P. Sattigeri, L. Karlinsky, K. Saenko, A. Oliva, and R. Feris. AR-Net: Adaptive Frame Resolution for Efficient Action Recognition. European Conference on Computer Vision (**ECCV 2020**).

80. Z. Tang, Y. Gao, P. Sattigeri, L. Karlinsky, R. Feris, and D. Metaxas. OnlineAugment: Online Data Augmentation with Less Domain Knowledge. European Conference on Computer Vision (**ECCV 2020**).

81. M. Lichtenstein, P. Sattigeri, R. Feris, R. Giryes, and L. Karlinsky. TAFSSL: Task-Adaptive Feature Sub-Space Learning for few-shot classification. European Conference on Computer Vision (**ECCV 2020**).

82. A. Andonian, C. Fosco, M. Monfort, A. Lee, R. Feris, C. Vondrick, and A. Oliva. We Have So Much In Common: Modeling Semantic Relational Set Abstractions in Videos. European Conference on Computer Vision (**ECCV 2020**).

83. A. Sahoo, A. Singh, R. Panda, R. Feris, and A. Das. Mitigating Dataset Imbalance via Joint Generation and Classification. ECCV Workshop on Imbalance Problems in Computer Vision, 2020.

84. C. Lin, Y. Hung, R. Feris, and L. He. Video Instance Segmentation Tracking. Conference on Computer Vision and Pattern Recognition (**CVPR 2020**).

85. Z. Wang, M. Yu, Y. Wei, R. Feris, J. Xiong, W. Hwu, T. Huang, and H. Shi. Differential Treatment for Stuff and Things: A Simple Unsupervised Domain Adaptation Method for Semantic Segmentation. Conference on Computer Vision and Pattern Recognition (**CVPR 2020**).

86. K. Ramakrishnan, R. Panda, Q. Fan, J. Henning, A. Oliva, and R. Feris. Relationship Matters: Relation Guided Knowledge Transfer for Incremental Learning of Object Detectors. CVPR Workshop on Continual Learning in Computer Vision, 2020.

87. M. Khoi-Nguyen, D. Joshi, R. Yeh, J. Xiong, R. Feris, and M. Do. Learning Motion in Feature Space: Locally- Consistent Deformable Convolution Networks for Fine Grained Action Detection. International Conference on Computer Vision (**ICCV 2019, Oral**), Seoul, Korea, 2019.

88. A. Alfassy, L. Karlinsky, A. Aides, J. Shtok, S. Harary, R. Feris, R. Giryes, and A. Bronstein. LaSO: Label-Set Operations Networks for Multi-label Few-shot Learning. Conference on Computer Vision and Pattern Recognition (**CVPR 2019, Oral**), Long Beach, California, 2019.

89. Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris. SpotTune: Transfer Learning through Adaptive Fine-tuning. Conference on Computer Vision and Pattern Recognition (**CVPR 2019**), Long Beach, California, 2019.

90. L. Karlinsky, J. Shtok, S. Harary, E. Schwartz, A. Aides, R. Feris, R. Giryes, and A. Bronstein. RepMet: Representative-based Metric Learning for Classification and One-shot Object Detection. Conference on Computer Vision and Pattern Recognition (**CVPR 2019**), Long Beach, California, 2019.

91. C. Chen, Q. Fan, N. Mallinar, T. Sercu, and R. Feris. Big-Little Net: An Efficient Multi-Scale Feature Representation for Visual and Speech Recognition. International Conference on Learning Representations (**ICLR 2019**), New Orleans, Louisiana, 2019.

92. A. Boggust, K. Audhkhasi, D. Joshi, D. Harwath, S. Thomas, R. Feris, D. Gutfreund, Y. Zhang, A. Torralba, M. Picheny, and James Glass. Grounding Spoken Words in Unlabeled Video. CVPR Workshop on Sight and Sound, 2019.

93. K. Ramakrishnan, M. Monfort, B. McNamara, A. Lascelles, D. Gutfreund, R. Feris, and A. Oliva. Identifying Interpretable Action Concepts in Deep Networks. CVPR Workshop on Explainable AI, 2019.

94. M. Jaiswal et al. Video-Text Compliance: Activity Verification Based on Natural Language Instructions. ICCV Workshop on Large Scale Holistic Video Understanding, 2019.

95. Z. Shen, H. Shi, J. Yu, H. Phan, R. Feris, L. Cao, D. Liu, X. Wang, T. Huang, M. Savvides. Learning Object Detection from Scratch via Gated Feature Reuse. British Machine Vision Conference (BMVC), 2019.

96. E. Schwartz*, L. Karlinsky*, J. Shtok, S. Harary, M. Marder, A. Kumar, R. Feris, R. Giryes, and A. Bronstein. Delta-Encoder: an Effective Sample Synthesis Method for Few-shot Object Recognition. Neural Information Processing Systems (**NeurIPS 2018, Spotlight**), Montreal, Canada, 2018. (* equal contribution)

97. X. Guo*, H. Wu*, Y. Cheng, S. Rennie, G. Tesauro, and R. Feris. Dialog-based Interactive Image Retrieval. Neural Information Processing Systems (**NeurIPS 2018**), Montreal, Canada, 2018. (*equal contribution)

98. A. Kumar, P. Sattigeri, K. Wadhawan, L. Karlinsky, R. Feris, W. T. Freeman, and G. Wornell. Co-regularized Alignment for Unsupervised Domain Adaptation. Neural Information Processing Systems (**NeurIPS 2018**), Montreal, Canada, 2018.

99. R. Gao, R. Feris and K. Grauman. Learning to Separate Object Sounds by Watching Unlabeled Video. European Conference on Computer Vision (**ECCV 2018, Oral**), Munich, Germany, 2018.

100. B. Cheng*, Y. Wei*, H. She, R. Feris, J. Xiong, and T. Huang. Revisiting RCNN: On Awakening the Classification Power of Faster RCNN. European Conference on Computer Vision (**ECCV 2018**), Munich, Germany, 2018. (* equal contribution)

101. Z. Wu*, T. Nagarajan*, A. Kumar, S. Rennie, L. Davis, K. Grauman, and R. Feris. BlockDrop: Dynamic Inference Paths in Residual Networks. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2018, Spotlight**), Salt Lake City, Utah, June 2018. (* equal contribution)

102. X. Peng, Z. Tang, F. Yang, R. Feris, and D. Metaxas. Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2018**), Salt Lake City, Utah, June 2018.

103. M. Merler, D. Joshi, K. Mac, Q. Nguyen, S. Hammer, J. Kent, J. Xiong, M. Do, J. Smith, and R. Feris. The Excitement of Sports: Automatic Highlights using Audio-Visual Cues. CVPR Workshop on Sight and Sound, 2018.

104. N. Codella, C. Lin, A. Halpern, M. Hind, R. Feris, and J. Smith. Collaborative Human-AI (CHAI): Evidence-Based Interpretable Melanoma Classification in Dermoscopic Images. MICCAI Workshop on Interpretability of Machine Intelligence in Medical Image Computing, 2018.

105. N. Codella, D. Anderson, T. Philips, A. Porto, K. Massey, J. Snowdon, R. Feris, and J. Smith. Segmentation of both Diseased and Healthy Skin from Clinical Photographs in a Primary Care Setting. International Engineering in Medicine and Biology Conference, 2018.

106. M. Beigi, L. M Brown, Q. Fan, J. Henning, C. Lin, H. Shi, C. Shu, and R. Feris. Object-Centric Spatio-Temporal Activity Detection and Recognition. NIST TRECVID Workshop, 2018.

107. Y. Lu, A. Kumar, S. Zhai, Y. Cheng, T. Javidi, and R. Feris. Fully Adaptive Feature Sharing in Multi-Task Networks with Applications in Person Attribute Classification. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2017, Spotlight**), Honolulu, Hawaii, July 2017.

108. S. Zhai, H. Wu, A. Kumar, Y. Cheng, Y. Lu, Z. Zhang,and R. Feris. S3Pool: Pooling with Stochastic Spatial Sampling. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2017**), Honolulu, Hawaii, July 2017.

109. M. Merler, D. Joshi, Q. Nguyen, S. Hammer, J. Kent, J. R Smith, and R. Feris. Automatic Curation of Golf Highlights using Multimodal Excitement Features. Workshop of Computer Vision in Sports (in conjunction with CVPR), 2017.

110. Z. Cai, Q. Fan, R. Feris, and N. Vasconcelos. A Unified Multi Scale Deep Convolutional Neural Network for Fast Object Detection. European Conference on Computer Vision (**ECCV 2016**), Amsterdam, Netherlands, 2016.

111. X. Peng, R. Feris, X. Wang, and D. Metaxas. A Recurrent Encoder-Decoder Network for Sequential Face Alignment. European Conference on Computer Vision (**ECCV 2016, Oral**), Amsterdam, Netherlands, 2016.

112. J. Wang, Y. Cheng, and R. Feris. Walk and Learn: Facial Attribute Representation Learning from Egocentric Video and Contextual Data. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2016, Oral**), Las Vegas, Nevada, June 2016.

113. Y. Cheng, F. Yu, R. Feris, S. Kumar, A. Choudhary, and S. F. Chang. An Exploration of Parameter Redundancy in Deep Networks with Circulant Projections. IEEE International Conference on Computer Vision (**ICCV 2015**), Santiago, Chile, December 2015.

114. J. Huang, R. Feris, Q. Chen, and S. Yan. Cross-domain Image Retrieval with a Dual Attribute-aware Ranking Network. IEEE International Conference on Computer Vision (**ICCV 2015**), Santiago, Chile, December 2015.

115. Q. Chen, J. Huang, R. Feris, L. Brown, J. Dong, and S. Yan. Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2015**), Boston, Massachusetts, June 2015.

116. J. Bowler, R. Feris, L. Cao, J. Wang, and M. Zhou. Automated Axon Segmentation from Highly Noisy Microscopic Videos. Winter Conference on Applications of Computer Vision (WACV 2015), Kona, Hawaii, 2015.

117. R. Feris, R. Bobbitt, and S. Pankanti. Efficient 24/7 Object Detection in Surveillance Videos. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2015), Germany, August 2015.

118. Y. Cheng, L. Brown, Q. Fan, R. Feris, S. Pankanti, and T. Zhang. RiskWheel: Interactive Visual Analytics for Surveillance Event Detection. IEEE International Conference on Multimedia and Expo (ICME 2014, Oral), Chengdu, China, 2014.

119. J. Xie, L. Chou, R. Feris, and M.T. Sun. Single Depth Image Super resolution and Denoising via Coupled Dictionary Learning with Local Constraints and Shock Filtering. IEEE International Conference on Multimedia and Expo (ICME 2014, Oral), 2014.

120. J. Xie, R. Feris, and M.T. Sun. Edge Guided Single Depth Image Super Resolution. IEEE International Conference on Image Processing (ICIP 2014), 2014.

121. Y. Cui, Y. Xiang, K. Rong, L. Cao, and R. Feris. A Spatial-Color Layout Feature for Representing Galaxy Images. IEEE Winter Conference on Applications of Computer Vision (WACV 2014), Steamboat Springs, Colorado, 2014.

122. L. Brown, R. Feris, and S. Pankanti. Temporal Non-Maximum Suppression for Pedestrian Detection Using Scene Context. International Conference on Pattern Recognition (ICPR 2014), Stockholm, Sweden, 2014.

123. R. Feris, L. Brown, S. Pankanti, and M.T. Sun. Appearance-based Object Detection under Varying Environmental Conditions. International Conference on Pattern Recognition (ICPR 2014, Oral), Stockholm, Sweden, 2014.

124. A. Mattos and R. Feris. Fusing Well-Crafted Feature Descriptors for Efficient Fine-Grained Classification. IEEE International Conference on Image Processing (ICIP 2014), Paris, France 2014.

125. A. Mattos, R. Herrmann, K. Shigeno, and R. Feris. A Mission-Oriented Citizen Science Platform for Efficient Flower Classification Based on Combination of Feature Descriptors. ICMR Workshop on Environmental Multimedia Retrieval, Glasgow, UK, 2014.

126. A. Mattos, R. Herrmann, K. Shigeno, and R. Feris. Flower Classification for a Citizen Science Mobile App. International Conference on Multimedia Retrieval (ICMR 2014), Glasgow, UK, 2014.

127. R. Feris, R. Bobbit, L. Brown, and S. Pankanti. Attribute-based People Search: Lessons Learnt from a Practical Surveillance System. ACM International Conference on Multimedia Retrieval (ICMR 2014), Oral Presentation, Glasgow, UK, 2014.

128. K. Scherbaum, R. Feris, J. Petterson, V. Blanz, and H. Seidel. Fast Face Detector Training Using Tailored Views. IEEE International Conference on Computer Vision (**ICCV 2013**), Sydney, Australia, December 2013.

129. Q. Chen, Z. Song, R. Feris, A. Datta, L. Cao, Z. Huang, and S. Yan. Efficient Maximum Appearance Search for Large-Scale Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2013**), Portland, Oregon, June 2013.

130. F. Yu, L. Cao, R. Feris, J. Smith, and S. Chang. Designing Category-Level Attributes for Discriminative Visual Recognition. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2013**), Portland, Oregon, June 2013.

131. Q. Chen et al. Spatio-Temporal Fisher Vector Coding for Surveillance Event Detection. ACM International Conference on Multimedia (ACM MM 2013), Barcelona, Spain, 2013.

132. J. Xie, J. Hsu, R. Feris, and M. Sun. Fine Registration of 3D Point Clouds with ICP Using an RGB-D Camera. IEEE International Symposium on Circuits and Systems (ISCAS 2013), Beijing, China, 2013.

133. X. Wang, L. Cao, R. Feris, A. Datta, and T. Xan. Hierarchical Feature Pooling with Structure Learning: A new method for Pedestrian Detection. CVPR Workshop on Structured Prediction: Tractability, Learning, and Inference, Portland, Oregon, 2013.

134. J. Leandro, R. Cesar, and R. Feris. Shape Analysis using the Spectral Graph Wavelet Transform. IEEE eScience Conference, Beijing, China 2013.

135. R. Feris, A. Datta, M. T. Sun, and S. Pankanti. Boosting Object Detection Performance in Crowded Surveillance Videos. Winter Conference on Applications of Computer Vision (WACV 2013), Florida, USA, 2013.

136. F. Rashed, B. Siddiquie, R. Feris, and L. Davis. Domain Adaptive Object Detection. Winter Conference on Applications of Computer Vision (WACV 2013), Florida, USA, 2013.

137. Y. Cai et al. CMU-IBM-NUS@TRECVID 2012: Surveillance Event Detection. NIST Technical Report, 2012. First place in the retrospective surveillance event detection task.

138. R. Feris, B. Siddiquie, and S. Pankanti. Learning Detectors from Large Datasets for Object Retrieval in Video Surveillance. International Conference on Multimedia and Expo (ICME 2012), Melbourne, Australia, 2012.

139. B. Siddiquie, R. Feris, A. Datta, and L. Davis. Unsupervised Model Selection for View-Invariant Object Detection in Surveillance Environments. International Conference on Pattern Recognition (ICPR 2012, Oral), Tsukuba City, Japan, 2012.

140. A. Datta, L. Brown, R. Feris, and S. Pankanti. Appearance Modeling for Person Re-Identification using Weighted Brightness Transfer Functions. International Conference on Pattern Recognition (ICPR 2012), Tsukuba City, Japan, 2012.

141. B. Siddiquie, R. Feris, and L. Davis. Image Ranking and Retrieval Based on Multi-Attribute Queries. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2011, Oral Presentation**), Colorado Springs, USA, 2011.

142. R. Feris, B. Siddiquie, Y. Zhai, J. Petterson, L. Brown, and S. Pankanti. Attribute-based Vehicle Search in Crowded Surveillance Videos. ACM International Conference on Multimedia Retrieval (ICMR 2011, Oral Presentation), Trento, Italy, 2011.

143. A. Datta, R. Feris, and D. Vaquero. Hierarchical Ranking of Facial Attributes. IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, California, 2011.

144. R. Feris, J. Petterson, B. Siddiquie, L. Brown, and S. Pankanti. Large-Scale Vehicle Detection in Challenging Urban Surveillance Environments. Winter Conference on Applications of Computer Vision (WACV 2011), Kona, Hawaii, 2011.

145. H. Liu, R. Feris, V. Krueger, and M.T. Sun. Unsupervised Action Classification Using Space-Time Link Analysis. IEEE International Symposium on Circuits and Systems (ISCAS 2010), Paris, France, 2010.

146. L. Chen, J. McAuley, R. Feris, T. Caetano, and M. Turk. Shape Classification Through Structured Learning of Matching Measures. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2009**), Miami, Florida, June 2009.

147. D. Vaquero, R. Raskar, R. Feris, and M. Turk. A Projector-Camera Setup for Geometry-Invariant Frequency Demultiplexing. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2009**), Miami, Florida, June 2009.

148. D. Vaquero, R. Feris, L. Brown, and A. Hampapur. Attribute-based people search in surveillance environments. Winter Conference on Applications of Computer Vision (WACV 2009, Oral Presentation), Snowbird, Utah, December 2009.

149. A. Hampapur, R. Bobbitt, L. Brown, M. Desimone, R. Feris, R. Kjeldsen, M. Lu, C. Mercier, C. Milite, S. Russo, C. Shu, Y. Zhai. Video Analytics in Urban Environments. IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2009.

150. R. Feris, Y. Tian, Y. Zhai, and A. Hampapur. Facial Image Analysis Using Local Feature Adaptation Prior to Learning. IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, Netherlands, September 2008.

151. Y. Tian, R. Feris, and A. Hampapur. Real-Time Detection of Abandoned and Removed Objects in Complex Environments. IEEE International Workshop on Visual Surveillance (in conjunction with ECCV 2008), Marseille, France, 2008.

152. D. Vaquero, R. Feris, M. Turk, and R. Raskar. Characterizing the Shadow Space of Camera-Light Pairs. IEEE Conference on Computer Vision and Pattern Recognition (**CVPR 2008**), Anchorage, Alaska, June 2008.

153. L. Chen, R. Feris, Y. Zhai, L. Brown, and A. Hampapur. An Integrated System for Moving Object Classification in Surveillance Videos. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2008), Santa Fe, New Mexico, September 2008.

154. L. Chen, R. Feris, and M. Turk. Efficient Partial Shape Matching Using the Smith-Waterman Algorithm. CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, Anchorage, Alaska, June 2008.

155. A. Hampapur, L. Brown, R. Feris, A. Senior, C. Shu, Y. Tian, Y. Zhai, and M. Lu. Searching Surveillance Video. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2007), London, UK, September 2007.

156. A. Senior, L. Brown, A. Hampapur, C. Shu, Y. Zhai, R. Feris, Y. Tian, S. Borger, and C. Carlson. Video Analytics for Retail: the IBM Smart Surveillance Retail Solution. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2007), London, UK, September 2007.

157. R. Feris, Y. Tian, and A. Hampapur. Capturing People in Surveillance Video. IEEE International Workshop on Visual Surveillance, 2007.

158. R. Feris, R. Raskar, and M. Turk. Dealing with Multi-scale Depth Changes and Motion in Depth Edge Detection. Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2006), Manaus, Brazil, October 2006. Awarded "One of the Best Image Processing and Computer Vision Papers".

159. H. Guan, J. Chang, L. Chen, R. Feris, and M. Turk. Multi-view Appearance-based 3D Hand Pose Estimation. IEEE Workshop on Vision for Human Computer Interaction (in conjunction with CVPR 2006), New York, NY, June 2006.

160. H. Guan, R. Feris, and M. Turk. The Isometric Self-Organizing Map for Hand Pose Estimation. International Conference on Face and Gesture Recognition, Southampton, UK, 2006.

161. R. Feris, R. Raskar, L. Chen, K. Tan, and M. Turk. Discontinuity Preserving Stereo with Small Baseline Multi-Flash Illumination. International Conference on Computer Vision (**ICCV 2005, Oral Presentation)**, Beijing, China, 2005.

162. Y. Zana, R. Cesar, R. Feris, and M. Turk. Face Verification in Polar Frequency Domain: a Biologically Motivated Approach. International Symposium on Visual Computing, Lake Tahoe, NV, 2005.

163. R. Feris, R. Raskar, K. Tan, and M. Turk. Specular Reflection Reduction with Multi-Flash Imaging. Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2004), Curitiba, Brazil, October 2004. Also accepted as a poster in SIGGRAPH 2004.

164. R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi. Exploiting Depth Discontinuities for Vision-based Fingerspelling Recognition. IEEE Workshop on Real-Time Vision for Human-Computer Interaction (in conjunction with CVPR 2004), Washington DC, USA, June 2004.

165. K. Tan, J. Kobler, R. Feris, P. Dietz, and R. Raskar. Shape Enhanced Surgical Visualizations and Medical Illustrations with Multi-flash Imaging. International Conference on Medical Imaging Computing and Computer Assisted Intervention (MICCAI 2004), Rennes, France 2004.

166. C. Hu, Y. Chang, R. Feris, and M. Turk. Manifold Based Analysis of Facial Expression. IEEE Workshop on Face Processing in Video (in conjunction with CVPR 2004), Washington DC, USA, June 2004.

167. C. Hu, R. Feris, and M. Turk. Real-time View-Based Face Alignment Using Active Wavelet Networks. Workshop on Analysis and Modeling of Faces and Gestures (in conjunction with ICCV 2003), Nice, France, October 2003.

168. C. Hu, R. Feris, and M. Turk. Active Wavelet Networks for Face Alignment. British Machine Vision Conference (BMVC 2003), Norwich, 2003.

169. M. Turk, C. Hu, R. Feris, F. Lashkari, and A. Beall. TLA Based Face Tracking. International Conference on Vision Interfaces, Calgary, Canada, 2002.

170. R. Feris, J. Gemmell, K. Toyama, and V. Krueger. Hierarchical Wavelet Networks for Facial Feature Localization. International Conference on Automatic Face and Gesture Recognition, Washington D.C., USA, May 20-21, 2002.

171. R. Feris, V. Krueger, and R. M. Cesar Jr. Efficient Real-Time Face Tracking in Wavelet Subspace. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems (in conjunction with ICCV 2001), Vancouver, BC, 2001.

172. R. Feris and R. M. Cesar Jr. Locating and Tracking Facial Landmarks Using Gabor Wavelet Networks. Lecture Notes in Computer Science, vol. 2013, pp. 311-320. International Conference on Advances in Pattern Recognition, Rio de Janeiro, Brazil, May 2001.

173. V. Krueger and R. Feris. Wavelet Subspace Method for Real-time Face Tracking. Proc. Pattern Recognition. DAGM Symposium, Munich, Germany 2001.

174. R. Feris, T. E. Campos, and R. M. Cesar Jr. A Project for Face Recognition from Video Sequences Using GWN and Eigenfeature Selection. Workshop in Artificial Intelligence and Computer Vision, Atibaia, Brazil, November 2000.

175. R. Feris and R. M. Cesar Jr. Tracking Facial Features Using Gabor Wavelet Networks. Symposium on Computer Graphics and Image Processing (SIBGRAPI 2000), Gramado, Brazil, October 2000.

176. T. Campos, R. Feris, and R. M. Cesar Jr. Improved Face versus Non-Face Discrimination Using Fourier Descriptors through Feature Selection. Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2000), Gramado, Brazil, October 2000.

177. R. Feris, T. Campos, and R. M. Cesar Jr. Detection and Tracking of Facial Features in Video Sequences. Lecture Notes on Artificial Intelligence, vol. 1793, pp. 127-135. Proceedings of MICAI-2000, Acapulco, Mexico, April 2000.

178. T. Campos, R. Feris, and R. M. Cesar Jr. Eigenfaces versus Eigeneyes: First Steps Towards Performance Assessment of Representations for Face Recognition. Lecture Notes on Artificial Intelligence, vol. 1793, pp. 193-201. Proceedings of MICAI-2000, Acapulco, Mexico, April 2000.

179. R. Feris and W. Lages. Stereo Image Matching Using Correlation and Relaxation Labeling. IV Congress for Scientific Initiation and Postgraduation at Aeronautics Institute of Technology, pp. 193-199, Sao Jose dos Campos-SP, Brazil, October 1998 (in portuguese).

Journal Papers

1. A. Rouditchenko, S. Thomas, H. Kuehne, R. Feris, and J. Glass. mWhisper-Flamingo for Multilingual Audio-Visual Noise-Robust Speech Recognition. IEEE Signal Processing Letters, 2025.

2. M. Monfort, K. Ramakrishnan, A. Andonian, B. McNamara, A. Lascelles, B. Pan, Q. Fan, D. Gutfreund, R. Feris, and A. Oliva. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022.

3. J. Lee, Y. Bu, P. Sattigeri, R. Panda, G. Wornell, L. Karlinsky, and R. Feris. A Maximal Correlation Framework for Fair Machine Learning. Entropy 24(4), 461, 2022.

4. E. Schwartz, L. Karlinsky, R. Feris, R. Giryes, and A. Bronstein. Baby steps towards few-shot learning with multiple semantics. Pattern Recognition Letters, 2022.

5. S. Doveh, E. Schwartz, C. Xue, R. Feris, A. Bronstein, R. Giryes, and L. Karlinsky. MetAdapt: Meta-Learned Task-Adaptive Architecture for Few-Shot Classification. Pattern Recognition Letters, 2021.

6. X. Peng , R. Feris, X. Wang, and D. Metaxas. Red-net: A recurrent encoder-decoder network for video-based face alignment. International Journal of Computer Vision (**IJCV**), 2018.

7. M. Merler, D. Joshi, Q. Nguyen, S. Hammer, J. Kent, J. Xiong, M. Do, J. Smith, and R. Feris. Automatic Curation of Sports Highlights using Multimodal Excitement Features. IEEE Transactions on MultiMedia (TMM) 2018.

8. J. Xie, R. Feris, and M.T. Sun. Edge-Guided Single Depth Image Super Resolution. IEEE Transactions on Image Processing (TIP), vol. 25, no. 1, pp. 428-438, 2016.

9. J. Xie, Y. Hsu, R. Feris, and M.T. Sun. Fine registration of 3D point clouds fusing structural and photometric information using an RGB-D camera. Journal of Visual Communication and Image Representation (JVCI), vol. 32, pp. 194-204, 2015.

10. J. Xie, R. Feris, S. Yu, and M.T. Sun. Joint Super Resolution and De-noising from a Single Depth Image. IEEE Transactions on Multimedia (TMM), vol.17, no.9, pp.1525-1537, 2015.

11. R. Feris, B. Siddiquie, J. Petterson, Y. Zhai, A. Datta, L. Brown, and S. Pankanti. Large-Scale Vehicle Detection, Indexing, and Search in Urban Surveillance Videos. IEEE Transactions on Multimedia, 2012.

12. S. Pankanti, L. Brown, J. Connell, A. Datta, Q. Fan, R. Feris, N. Haas, Y. Li, N. Ratha, and H. Thinh. Practical Computer Vision: Example Techniques and Challenges. IBM Journal of Research and Development, 201.

13. Y. Tian, R. Feris, H. Liu, A. Hampapur, and M. Sun. Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos. IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews, 2010.

14. H. Liu, R. Feris, V. Krueger, and M.T. Sun. Unsupervised Action Classification Using Space-Time Link Analysis. Eurasip Journal on Advances in Signal Processing, 2010.

15. A. Hampapur et al. Analytics Driven Asset Management. IBM Smart Cities Journal, 2010.

16. R. Feris, R. Raskar, L. Chen, K. Tan, and M. Turk. Multi-Flash Stereopsis: Depth Edge Preserving Stereo with Small Baseline Illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), vol. 30, no. 1, pp. 147-159, 2008.

17. Y. Zana, R. Cesar, R. Feris and M. Turk. Local Approach for Face Verification in Polar Frequency Domain. Image and Vision Computing Journal, vol. 24, no. 8, pp. 904-913, 2006.

18. R. Feris, M. Turk, R. Raskar, and K. Tan. Specular Highlights Detection and Reduction with Multi-Flash Photography. International Journal of the Brazilian Computer Society, vol. 1, no. 12, pp. 35-42, 2006.

19. Y. Chang, C. Hu, R. Feris, and M. Turk. Manifold-Based Analysis of Facial Expressions. Image and Vision Computing Journal, vol.24, no. 6, pp. 605-614, 2006.

20. R. Raskar, K. Tan, R. Feris, J. Kobler, J. Yu, and M. Turk. Harnessing Real-World Depth Edges with Multi-Flash Imaging. IEEE Computer Graphics and Applications (IEEE CG&A), vol. 25, no. 1, pp. 32-38, January 2005.

21. R. Feris, V. Krueger, and R. Cesar. A Wavelet Subspace Method for Real-time Face Tracking. Journal of Real-time Imaging, vol. 10, pp. 339-350, 2004.

22. R. Raskar, K. Tan, R. Feris, J. Yu, and M. Turk. Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging. ACM Transactions on

Graphics (**SIGGRAPH 2004**), Vol. 23, Issue 3, August 2004. Also accepted in SIGGRAPH Emergent Technologies, 2004.

Book Chapters

1. R. Feris, C. Lampert and D. Parikh. Introduction to Visual Attributes. In Visual Attributes (eds: R. S. Feris, C. Lampert, and D. Parikh), Springer, 2017.

2. H. Liu, R. Feris, and M.T. Sun. Benchmarking Datasets for Human Activity Recognition. Visual Analysis of Humans – Looking at People, Springer 2011.

3. Y. Zhai, R. Feris, A. Hampapur, S. Russo, and S. Pankanti. Parsing Object Events in Heavy Urban Traffic. Object Tracking, Intech, 2011.

4. H. Liu, M.T. Sun, and R. Feris. Video Activity Recognition. Multimedia Analysis, Processing and Communication, Z. Li, J. Kacprzyk, D. Tao, E. Izquierdo, W. Lin, and H. Wang ed., Springer, 2010.

5. R. Feris, A. Hampapur, Y. Zhai, R. Bobbitt, L. Brown, D. Vaquero, Y-L. Tian, H. Liu, and M-T. Sun. Case Study: IBM Smart Surveillance System. Intelligent Video Surveillance: Systems and Technology, by Taylor & Francis Group, LLC, 2009.

6. Y. Zhai, R. Feris, L. Brown, R. Bobbitt, A. Hampapur, S. Pankanti, Q. Fan, A. Yanagawa, Y. Tian, and S. Velipasalar. Composite Event Detection in Multi-Camera and Multi-Sensor Surveillance Networks. Multi-Camera Networks: Concepts and Applications, by Elsevier, 2009.

7. Y. Tian, R. Feris, L. Brown, D. Vaquero, Y. Zhai, and A.Hampapur. Multi-Scale People Detection and Motion Analysis for Video Surveillance. Machine Learning for Human Motion Analysis: Theory and Practice, IGI Global, 2009.

8. D. Vaquero, R. Feris, L. Brown, A. Hampapur, and Matthew Turk. Attribute-based People Search. Intelligent Video Surveillance: Systems and Technology, by Taylor & Francis Group, LLC, 2009.

9. Y. Tian, A. Hampapur, L. Brown, R. Feris, M. Lu, A. Senior, C. Shu, and Y. Zhai. Event Detection, Query, and Retrieval for Video Surveillance. Artificial Intelligence for Maximizing Content Based Image Retrieval, 2008.

10. R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi. Recognition of Isolated Fingerspelling Gestures Using Depth Edges. B. Kisacanin, V. Pavlovic and T. Huang (eds.), Real-time Vision for Human-Computer Interaction, Springer-Verlag, 2005.

Theses and Reports

1. M. Stallone et al. Scaling Granite Code Models to 128K Context. IBM Technical Report, 2024.

2. A. Arbelle, G. Blackwood, L. Karlinsky, A. Sahoo, J. Schtok, and R. Feris. LATERAL: Learning Automatic, Transfer-Enhanced, and Relation-Aware Labels. Technical Report, DARPA Learning with Less Labels, 2023.

3. M. Merler, N. Ratha, R. Feris, and J. Smith. Diversity in Faces. arXiv 2019.

4. B. Cheng, Y. Wei, R. Feris, J. Xiong, W. Hwu, T. Huang, and H. Shi. Decoupled Classification Refinement: Hard False Positive Suppression for Object Detection. arXiv 2018.

5. R. Feris. Detection and Analysis of Depth Discontinuities with Lighting and Viewpoint Variation. PhD thesis, University of California, Santa Barbara, 2006.

6. R. Feris, J. Gemmell, K. Toyama, and V. Krueger. Facial Feature Detection using a Hierarchical Wavelet Face Database. Microsoft Research Technical Report MSR-TR-2002-05, January 2002.

7. R. Feris. Efficient Real-time Face Tracking in Wavelet Subspace. MSc. thesis, University of Sao Paulo, May 2001. Awarded as the Second Best Computer Science MSc. Thesis in Brazil (Nationwide Competition, 2002).

Demos and Posters

1. D. Joshi, M. Merler , Q. Nguyen, S. Hammer, J. Kent, J. Smith, and R. Feris. IBM High-Five: Highlights from Intelligent Video Engine. ACM Multimedia, 2017.

2. F. Yu, L. Cao, R. Feris, J. Smith, and S. F. Chang. Designing Category-level Attributes for Discriminative Visual Recognition. Greater New York Area Workshop on Multimedia and Vision, 2013. (Best Poster Award)

3. IBM Centennial Demo – Traffic Analysis Using Video Analytics. Lincoln Center, Manhattan, New York 2011.

4. R. Feris, L. Brown, Q. Fan, A. Datta, S. Pankanti, A. Hampapur, Y. Zhai, R. Bobbitt, R. Kjeldsen, M. Lu, C. Shu, S. Russo, and C. Milite. IBM Smart Surveillance System: Searchable Video Analytics for Urban Environments and Retail Stores. CVPR DEMO, 2010.

5. A. Agrawal, V. Branzoi, R. Chellappa, R. Feris, R. Raskar, K. Tan, and M. Turk. Depth Edges in Real-Time Using a Multi-Flash Camera. CVPR DEMO, 2005.

6. R. Raskar, K. Tan, R. Feris, J. Yu, and M. Turk. Non-photorealistic Camera: Automatic Stylization with Multi-Flash Imaging. SIGGRAPH Emergent Technologies, 2004.

7. R. Feris, R. Raskar, K. Tan, and M. Turk. Specular Reflection Reduction Using a Multi-Flash Camera. SIGGRAPH poster, 2004.

Issued Patents

1. 12,548,333 - Dynamic network quantization for efficient video inference
2. 12,499,332 - Translating text using generated visual representations and artificial intelligence
3. 12,154,307 - Interpretability-aware redundancy reduction for vision transformers
4. 12,067,082 - Temporal contrastive learning for semi-supervised video action recognition
5. 11,954,910 - Dynamic multi-resolution processing for video classification
6. 11,263,488 - System and method for augmenting few-shot object classification with semantic information from multiple sources
7. 11,176,587 - Method, a system, and a computer readable storage medium for auto recommendations for an online shopping cart

8. 11,100,145 - Dialog-based image retrieval with contextual information
9. 11,079,924 - Cognitive graphical control element
10. 10,977,303 - Image retrieval using interactive natural language dialog
11. 10,679,047 - System and method for pose-aware feature learning
12. 10,607,089 - Re-identifying an object in a test image
13. 10,595,101 - Auto-curation and personalization of sports highlights
14. 10,424,342 - Facilitating people search in video surveillance
15. 10,390,986 - Control device for controlling a rigidity of an orthosis and method of controlling a rigidity of an orthosis
16. 10,275,608 - Object-centric video redaction
17. 10,169,661 - Filtering methods for visual object detection
18. 10,163,355 - Dynamic management system, method, and recording medium for cognitive drone-swarms
19. 10,163,042 - Finding missing persons by learning features for person attribute classification based on deep learning
20. 10,089,551 - Self-optimized object detection using online detector selection
21. 10,040,551 - Drone delivery of coffee based on a cognitive state of an individual
22. 9,495,599 - Determination of train presence and motion state in railway environments
23. 9,477,890 - Object detection using limited learned attribute ranges
24. 9,471,852 - User-configurable settings for content obfuscation
25. 9,460,361 - Foreground analysis based on tracking information
26. 9,460,349 - Background understanding in video data
27. 9,443,148 - Visual monitoring of queues using auxiliary devices
28. 9,430,874 - Estimation of object properties in 3D world
29. 9,424,659 - Real time processing of video frames
30. 9,396,548 - Multi-cue object detection and analysis
31. 9,342,594 - Indexing and searching according to attributes of a person
32. 9,330,312 - Multispectral detection of personal attributes for video surveillance
33. 9,330,111 - Hierarchical ranking of facial attributes
34. 9,322,647 - Determining camera height using distr of object heights and object image heights
35. 9,299,162 - Multi-mode video event indexing
36. 9,280,833 - Topology determination for non-overlapping camera network
37. 9,262,445 - Image ranking based on attribute correlation
38. 9,251,425 - Object retrieval in video data using complementary detectors
39. 9,245,186 - Semantic parsing of objects in video
40. 9,224,049 - Detection of static object on thoroughfare crossings
41. 9,224,046 - Multi-view object detection using appearance model transfer from similar scenes
42. 9,165,375 - Automatically determining field of view overlap among multiple cameras
43. 9,134,399 - Attribute-based person tracking across multiple cameras
44. 9,104,919 - Multi-cue object association
45. 9,082,201 - Surface contamination determination system
46. 9,069,104 - Pathway management using model analysis and forecasting
47. 9,058,669 - Incorporating video meta-data in 3D models
48. 8,948,454 - Boosting object detection performance in videos
49. 8,934,670 - Real time processing of video frames for triggering an alert
50. 8,837,776 - Rule-based combination of a hierarchy of classifiers for occlusion detection
51. 8,824,791 - Color correction for static cameras
52. 8,811,663 - Object detection in crowded scenes
53. 8,774,532 - Calibration of video object classification
54. 8,675,917 - Abandoned object recognition using pedestrian detection
55. 8,620,026 - Video-based detection of multiple object types under varying poses
56. 8,488,881 - Object segmentation at a self-checkout
57. 8,483,481 - Foreground analysis based on tracking information

58. 8,249,301 - Video object classification
59. 8,170,276 - Object detection system based on a pool of adaptive features
60. 8,107,678 - Detection of abandoned and removed objects in a video stream
61. 7,738,725 - Stylized rendering using a multi-flash camera

Invited Talks

- "Building Lightweight Multimodal Models for Enterprise AI"
  - Keynote speaker, ICCV Workshop on Efficient Computing under Limited Resources, 2025
  - Northeastern University, 2026

- Invited Speaker, Google CSR Workshop / LatinX in Computer Vision Research Workshop, 2024. "Augmenting Large Language Models with Long-term Memory and Multimodal Capabilities"

- Keynote speaker, Conference on Graphics, Patterns and Images (Sibgrapi 2023), Brazil, 2023. "Learning with Real and Unreal Data in the Era of Foundation Models"

- Invited Speaker, IJCAI Workshop on Generalizing from Limited Resources in the Open World, 2023. "Learning Trusted Models with Less Data"

- Google Research India, 2022. "Representation Learning based on Synthetic Data"

- "Computational Visual Pathways for Multi-Task Learning and Simulation"
  - Flatiron Institute, 2022
  - Binghamton University, 2022
  - University of British Columbia, 2021

- Boston University AIR Distinguished Speaker Series, 2021. "Dynamic Neural Networks for Efficient Multimodal Video Understanding"

- Invited speaker, CVPR Workshop on Learning from Limited and Imperfect Data, 2021. "How Transferable are Contrastive Representations?"

- Keynote speaker, CVPR Workshop on Multimodal Learning and Applications, 2021. "Adaptive Multimodal Learning for Efficient Video Understanding"

- Keynote speaker, CVPR LatinX in CV Workshop. "Adaptive Multimodal Learning for Efficient Video Understanding"

- Keynote speaker, ICML LatinX in AI workshop, 2020. "Dynamic Neural Networks for Efficient Image and Video Classification"

- Keynote speaker, CVPR Workshop on Diagram Image Retrieval and Analysis, 2020. "Visual Learning Beyond Natural Images"

- Invited speaker, What's Next in AI – AI We can Scale, MIT-IBM Watson AI Lab event. "Learning to Learn"

- Invited speaker, NeurIPS Workshop on Energy Efficient Machine Learning and Cognitive Computing, 2019. "Adaptive (Dynamic) Neural Networks for Efficient Inference"

- Invited speaker, CVPR Workshop on Learning from Imperfect Data, 2019. "Learning More from Less: Weak Supervision and Beyond"

- Keynote speaker, CVPR Workshop on Understanding Subjective Attributes of Data, 2019. "Is It All Relative? Interactive Fashion Search based on Relative Natural Language Feedback"

- Invited speaker, CVPR Workshop on Energy Efficient Machine Learning and Cognitive Computing, 2019. "Speeding up Deep Neural Networks with Adaptive Computation and Efficient Multi-Scale Architectures"

- Keynote speaker, Conference on Graphics, Patterns and Images (SIBGRAPI 2016), Sao Jose dos Campos, Brazil, 2016. "Representation Learning Beyond Human Labels: Practical Applications in Visual Analysis of People"

- Keynote speaker, International Joint Conference on Artificial Intelligence (IJCAI) BeyondLabeler Workshop, New York City, USA, 2016. "Representation Learning Beyond Human Labels: Practical Applications in Visual Analysis of People"

- Invited Speaker, NYC Computer Vision Meetup, 2015. "Deep Domain Adaptation for Describing People Based on Fine-grained Attributes"

- Invited Speaker, Workshop on Recent Trends in Computer Vision, Boston, USA, 2015. "Deep Domain Adaptation for Describing People Based on Fine-grained Attributes".

- Invited Speaker, International Conference on Multimedia Retrieval (ICMR) Practitioner's day, Glasgow, Scotland, 2014. "Multimedia Analytics at IBM Research"

- NYU Center for Urban Science and Progress, New York City, 2014. "Smart Surveillance: Towards Visual Search based on Fine-Grained Semantic Attributes". Host: Claudio Silva

- Invited Speaker, CVPR Workshop on Fine-Grained Visual Categorization, Portland, Oregon, 2013. "Fine-Grained Attribute Classification and Search: Two Case Studies and Lessons Learnt"

- IBM Research, Brazil, 2013. "Learning Visual Semantics". Host: Andrea Mattos

- Invited Speaker, National Academy of Engineering JAFOE meeting, Irvine, California, 2012. "Intelligent Video Surveillance"

- University of Delaware, 2012. "Attribute-based Object Retrieval". Host: Jingyi Yu

- Invited Speaker, ICMR 2011 Practitioner's day, Trento, Italy 2011. "Attribute-based Object Retrieval"

- City College of New York, May 2010. "Searching for People and Vehicles in Urban Environments". Host: Yingli Tian

- Cognitec, Dresden, Germany, September 2008. "The IBM Smart Surveillance System". Host: Roger Kelesoglu

- Invited speaker, IBM Conference on Maturing and Leveraging Biometrics Analytics, 2008. "Person Identification through Search Based on Physical Attributes"

- University of Washington, 2007. "The IBM Smart Surveillance System". Host: Linda Shapiro
- Princeton University, 2007. "Computer Vision for Digital Photography and Smart Surveillance". Host: Fei-Fei Li
- Kodak Research, Intelligent Systems Lab, Rochester, NY, 2007. "Multi-Flash Photography and Applications". Host: Jiebo Luo
- HP Laboratories, Computer Vision and Graphics group, Palo Alto, CA, 2007. "Detection and Modeling of Depth Discontinuities with Lighting and Viewpoint Variation". Host: Bruce Culbertson
- Adobe Advanced Technology Labs, Seattle, 2006. "Discontinuity Detection and Modeling with Lighting and Viewpoint Variation" (also presented at University of Washington). Host: David Salesin
- Jet Propulsion Laboratory, Pasadena, CA, 2006. "Detection and Analysis of Depth Discontinuities in Computer Vision". Host: Larry Matthies
- IBM T.J. Watson Research, Hawthorne, NY, 2005. "What Can We Do with a Multi-Flash Camera?" Host: Deepak Turaga (as part of Emerging Leaders in Multimedia Seminar)
- Microsoft Research, Redmond, 2001. "A Wavelet Subspace Method for Real-time Face Tracking". Host: Kentaro Toyama

Academic Experience

- **Adjunct Associate Professor, Columbia University**
  - Instructor: Visual Recognition and Search, Spring 2014 (Graduate course)
    http://rogerioferis.com/VisualRecognitionAndSearch2014/home.html
    Evaluation: 5.00/5.00
  - Instructor: Visual Recognition and Search, Spring 2013 (Graduate course)
    http://rogerioferis.com/VisualRecognitionAndSearch2013/home.html
    Evaluation: 4.56/5.00
  - Instructor: Automatic Video Surveillance, Spring 2008 (Graduate Course)
    http://www.andrewsenior.com/technical/surveillanceclass/
- **Affiliate Associate Professor, University of Washington**
  - Co-advised the dissertations of UW PhD students and was involved in several other academic activities (such as selection of PhD students)
- **Short Courses / Tutorials**
  - ICCV 2019 Tutorial on Visual Learning with Limited Labeled Data
    https://sites.google.com/view/learning-with-limited-data/home
    with Kate Saenko, Leonid Karlinsky, Ross Girshick, Judy Hoffman, Kevin Swersky, and Lorenzo Torresani
  - ICCV 2015 Tutorial on Tools for Efficient Object Detection
    http://mp7.watson.ibm.com/ICCV2015/ObjectDetectionICCV2015.html

with Piotr Dollar, Xiaoyu Wang, Kaiming He, Ross Girshick, Rodrigo Benenson, and Jan Hosang

- CVPR 2014 Tutorial on Learning Visual Semantics: Models, Massive Computation, and Innovative Applications
  http://mp7.watson.ibm.com/LearningVisualSemantics/index.htm
  with Liangliang Cao, John Smith, and Shih-Fu Chang

- CVPR 2012 Tutorial on Looking at People: Benchmarking Human Activity Recognition with Haowei Liu and Ming-Ting Sun

- **PhD Thesis Committees**

  - Jovana Kondic, MIT, 2026
  - Zhuoran Yu, UW-Madison, 2026
  - Bowen Pan, MIT, 2024
  - Navaneet KL Murthy, University of California, Davis, 2024
  - Karim Ahmed, Dartmouth College, 2019
  - Brendan Jou, Columbia University, 2016
  - Xiaodong Yang, City University of New York, 2015
  - Shizhi Chen, City University of New York, 2013
  - Eanes Pereira, Universidade Federal de Campina Grande, 2012
  - Haowei Liu, University of Washington, 2011

- **PhD and MEng thesis co-advisor**

  - Maura Kelleher, MIT. "Real-Time Prosodic Tennis Commentary via Streaming Vision-Language Models", MEng, 2026.
  - Jerry Li, MIT. "Improving LLM Long Context Understanding via Synthetic Data and Adaptive Compression", MEng, 2024.
  - Irene Huang, MIT. "Rethinking the Evaluation of Compositional Reasoning for Modern VLMs", MEng, 2024.
  - Sonia Purohit, MIT. "AI Commentator: Narrating Sports Games through Multimodal Perception and Large Language Models", MEng, 2023.
  - Howard Zhong, MIT. "Learning Privacy-Preserving Transferable Video Representations", MEng, 2023.
  - Yo-whan Kim, MIT. "How Transferable are Video Representations Based on Synthetic Data?", MEng, 2022.
  - Jun Xie, University of Washington. "Utilizing Depth Information for Emerging 3D Applications", PhD, 2016.
  - Haowei Liu, University of Washington. "Human Activity Recognition from Video", PhD, 2011.

Professional Activities

- **Program Chair**

  - IEEE Winter Conference on Applications of Computer Vision (WACV) 2017
  - 1st GNY Area Vision and Multimedia Meeting, 2011.

- **Associate Editor**, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018-2023

- **Senior Area Chair**
    - Neural Information Processing Systems (NeurIPS) – 2024, 2025, 2026

- **Area Chair**
    - Conference on Computer Vision and Pattern Recognition (CVPR) - 2015, 2016, 2017, 2020, 2021, 2023, 2025, 2026
    - Neural Information Processing Systems (NeurIPS) - 2018, 2019, 2020, 2021, 2022, 2023
    - International Conference on Learning Representations (ICLR) – 2021, 2022, 2023, 2024
    - European Conference on Computer Vision (ECCV) – 2020, 2024
    - International Conference on Machine Learning (ICML) - 2021
    - International Conference on Computer Vision (ICCV) – 2015, 2023
    - ACM Multimedia - 2017
    - International Symposium on Visual Computing (ISVC) - 2015

- **Organizing Committee**
    - ICLR Workshop on New Frontiers in Associative Memories, 2026
    - CVPR Workshop on What is Next in Multimodal Foundation Models?, 2026
    - CVPR Workshop on Emerging Directions in Data for Multimodal Foundation Models, 2026
    - ICCV Workshop on Memory and Vision, 2025
    - ICCV Workshop on What is Next in Multimodal Foundation Models?, 2025
    - ICML Workshop on Long-Context Foundation Models, 2025
    - CVPR Workshop on What is Next in Multimodal Foundation Models?, 2025
    - CVPR Workshop on What is Next in Multimodal Foundation Models?, 2024
    - ICCV Workshop on What is Next in Multimodal Foundation Models?, 2023
    - CVPR Workshop on Dynamic Neural Networks Meets Computer Vision, 2021
    - CVPR Workshop on Neural Architecture Search, 2021
    - CVPR Workshop on Visual Learning with Limited Labels: Zero-Shot, Few-Shot, Any-Shot, and Cross-Domain Few-Shot Learning, 2020
    - CVPR Workshop on Computer Vision for Fashion, Art, and Design, 2020
    - CVPR Workshop on Neural Architecture Search and Beyond for Representation Learning, 2020
    - CVPR Workshop on Diagram Image Retrieval and Analysis, 2020
    - CVPR Workshop on Fair, Data-Efficient and Trusted Computer Vision, 2020
    - CVPR Workshop on Skin Image Analysis, 2020 (Steering Committee)
    - ECCV Workshop on Multimodal Video Analysis, 2020
    - CVPR Workshop on Bias Estimation in Face Analytics, 2019
    - CVPR Workshop on Skin Image Analysis, 2019 (Steering Committee)
    - ICCV Workshop on Linguistics Meets Image and Video Retrieval, 2019
    - ICCV Workshop on Moving Cameras: from Body Cameras to Drones, 2019
    - ICCV Workshop on Computer Vision for Fashion, Art, and Design, 2019 (Steering Committee)
    - ICCV Workshop on Multimodal Video Analysis and Moments in Time Challenge, 2019
    - ICML Workshop on Towards learning with limited labels: Equivariance, Invariance, and Beyond, 2018
    - ECCV Workshop on Bias Estimation in Face Analytics, 2018
    - ECCV Workshop on Computer Vision for Fashion, Art, and Design, 2018
    - CVPR Workshop on Moving Cameras: from Body Cameras to Drones, 2016
    - ECCV Workshop on Parts and Attributes, 2014
    - ECCV Workshop on Parts and Attributes, 2012
    - GNY Area Vision and Multimedia Meeting, 2012

- ECCV Workshop on Parts and Attributes, 2010
- IBM Watson Emerging Leaders in Multimedia Workshop, 2006 and 2007
- I Workshop on Artificial Intelligence and Computer Vision, 1997

- **Chair**
  - Industry Chair, ACM Multimedia 2022
  - Corporate Relations Chair, CVPR 2018
  - Industry and Practitioner Chair, ICMR 2016
  - Doctoral Consortium Chair, FG 2011
  - Corporate Relations Chair, FG 2011

- **LDV Vision Summit:** judge for the Entrepreneurial Computer Vision Challenges (http://www.ldv.co/visionsummit/2015/speakers), 2015

- **Award Committee Member**
  - City University of Hong Kong (CUHK) International Selection Committee for PhD thesis Award, 2015
  - Member of the committee for selection of the best paper award – International Conference in Image Processing (ICIP 2008)

- Watson Chair of the Multimedia Professional Interest Community (2013 – 2017)
- Doctoral Consortium Co-Chair, IEEE Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, USA, 2011
- Grant reviewer: Expert proposal evaluator for FIT-IT, Vienna, Austria, 2011.
- Department Intern Coordinator, IBM Research, 2010
- Seminar PIC Coordinator, IBM Research, 2010-2011
- Program Committee Member: ACL (2026), CVPR (2007-2014), ICCV (2007, 2009, 2011), ECCV (2008,2010,2012,2026), AVSS (2008), WACV (2011), ICIP (2009), ICPR (2010,2014), ACCV (2007,2009), Visual Surveillance Workshop (2009,2010) and others
- Reviewer for major conferences and journals in computer vision/graphics, including ACM SIGGRAPH, IEEE Transactions on PAMI, Pattern Recognition Letters, IVC Journal, CVIU Journal, CVPR, ICCV, Face and Gesture Recognition, Eurographics, ISMAR, etc.

Media Press Coverage

- IBM Research Team Models Unitary Operator Mapping for Unified Quantum-Linguistic Inputs. Quantum Zeitgeist, 2026.
- MIT researchers teach AI models to interpret charts. MIT News, 2026.
- Method teaches generative AI models to locate personalized objects. MIT News, 2025.
- AI learns how vision and sound are connected, without human intervention. MIT News, 2025.
- Researchers use large language models to help robots navigate. MIT News, 2024.
- How memory augmentation can improve large language model efficiency and flexibility. IBM Research blog, 2024.
- Helping computer vision and language models understand what they see. MIT News, 2023.

- A simpler path to better computer vision. MIT News, 2022
- In machine learning, synthetic data can offer real performance improvements. MIT News, 2022.
- A safer, lower-cost alternative to real data for pretraining computer vision models. IBM Research blog, 2022.
- Hallucinating to Better Text Translation. Communications of the ACM / MIT News, 2022.
- IBM's StarNet brings explainable AI to image classification. VentureBeat, 2020.
- Shrinking deep learning's carbon footprint. MIT News, 2020.
- Toward a machine learning model that can reason about everyday actions. MIT News, 2020.
- IBM researchers develop a pair of low-power, high-performance computer vision systems. VentureBeat, 2018.
- Coffee delivery drone patented by IBM. BBC News, 2018.
- Enjoy Those U.S. Open Highlights. A Computer Picked Them for You. New York Times, 2017.
- How an IBM Computer Picks U.S. Open Highlights. Fortune, 2017.
- IBM's Watson Serves Up This Year's U.S. Open Highlights. NBC News, 2017.
- How Artificial Intelligence May be Tennis's Ace for Serving Grand Slam Fans. Newsweek, 2017.
- IBM's Watson is creating US Open tennis highlight videos. Engadget, 2017.
- IBM uses AI to serve up Wimbledon highlights. CNet, 2017.
- IBM To Provide Wimbledon Highlights Using Artificial Intelligence. SportTechie, 2017.
- IBM Watson is creating highlight reels at the Masters. ZDNet, 2017.
- Fighting terrorism in New York City. CBS 60 Minutes, 2011. (starting at minute 7:20)
- ABC7 puts video analytics to the test. ABC News, 2010.
- Cameras help confirm Scott suicide ruling. ABC News, 2009.
- Intelligent Iris. ABC News, 2008.
- The Nonphotorealistic camera. SlashDot, 2004.
- Master's dissertation / Brazilian Media: My work on face tracking was featured at Jornal do SBT and TV USP.
- High-School Project / Brazilian Media: In 1993, during my technical high-school, I developed a software for constructing facial sketches from witnesses' descriptions (coding in assembly for accessing the video memory card, dealing with bit planes, etc.) My work was covered by several newspapers, magazines, and television programs, including Ciencia Viva and RBS TV.