

Chenliang Xu

Prepared on February 19, 2018

CONTACT	3005 Wegmans Hall 250 Hutchison Road Department of Computer Science University of Rochester Rochester, NY 14627	PHONE: (585) 275-5426 EMAIL: chenliang.xu@rochester.edu WEB: www.cs.rochester.edu/~cxu22 DATE OF BIRTH: 12 March 1988 CITIZENSHIP: China
CURRENT APPOINTMENT	Assistant Professor Department of Computer Science University of Rochester	ROCHESTER, NY 9/2016 - PRESENT
RESEARCH INTERESTS	Computer vision and its relations to natural language, robotics, and data science with a focus on problems in video understanding such as video segmentation, activity recognition, and multi-modal vision-and-x modeling.	
RESEARCH AREA	Computer Vision Deep Learning Multimodal Modeling	
EDUCATION	University of Michigan Ph.D. in Computer Science and Engineering ADVISOR: Jason J. Corso DISSERTATION TITLE: "Scale-Adaptive Video Understanding"	ANN ARBOR, MI 8/2016
	State University of New York at Buffalo M.S. in Computer Science and Engineering	BUFFALO, NY 2/2012
	Nanjing University of Aeronautics and Astronautics B.S. in Information and Computing Science	NANJING, CHINA 6/2010
DISTINCTIONS	National Science Foundation BIGDATA Award <i>Audio-Visual Scene Understanding</i>	2017
	Best Paper Award at 14th Sound and Music Computing Conference	2017
	University of Rochester AR/VR Pilot Grant Award <i>Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech</i>	2017
	Rackham Conference Travel Grant – University of Michigan	2015, 2016
	Travel Award of IEEE ICCV 2013	2013
	Best Open Source Code Third Prize at IEEE CVPR 2012	2012
	Best Demo Prize at the 2nd Greater New York Multimedia and Vision Meeting	2012
	First Tier Jiangsu Province Award – Mathematics Modeling Competition in China	2007
	Outstanding Student Scholarship – Nanjing Univ. of Aero. and Astro.	2007, 2008, and 2009
PUBLICATIONS	JOURNAL ARTICLES ¹ <i>Qualifiers added where known: Impact Factor (IF), h5-Index (h5) provided by Google Scholar.</i>	
	J22. C. Xu and J. J. Corso. Libsvx: A supervoxel library and benchmark for early video processing. <i>International Journal of Computer Vision</i> , 119(3):272–290, 2016. IF: 4.27; h5: 63 (IJCV)	
	J21. T. Han, H. Yao, C. Xu, X. Sun, Y. Zhang, and J. J. Corso. Dancelets mining for video recommendation based on dance styles. <i>IEEE Transactions on Multimedia</i> , 19(4):712–724, 2017	IF: 2.54; h5: 51 (TMM)

¹IJCV and TPAMI are among the journals with highest impact factors across all computer science categories.

- J20. C. Xu, R. F. Doell, S. J. Hanson, C. Hanson, and J. J. Corso. A study of actor and action semantic retention in video supervoxel segmentation. *International Journal of Semantic Computing*, 7(4):353–375, 2013. *Selected as a Best Paper from ICSC.*
h5: 12 (IJSC)

CONFERENCE PROCEEDINGS (PEER-REVIEWED)²

Qualifiers added where known: Acceptance Rate (AR), h5-Index (h5) provided by Google Scholar.

- C19. L. Ding and C. Xu. Weakly-supervised action segmentation with iterative soft boundary assignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. (CVPR)
- C18. T. Chen, C. Xu, and J. Luo. Improving text-based person search by spatial matching and adaptive threshold. In *IEEE Winter Conference on Applications of Computer Vision*, 2018. AR: 37%; h5:31 (WACV)
- C17. L. Zhou, C. Xu, and J. J. Corso. Towards automatic learning of procedures from web instructional videos. In *AAAI Conference on Artificial Intelligence*, 2018. AR: 24%; h5:56 (AAAI)
- C16. L. Chen, Y. Wu, A. M. DSouza, A. Z. Abidin, A. Wismüller, and C. Xu. Mri tumor segmentation with densely connected 3d cnn. In *SPIE Conference on Medical Imaging*, 2018. Oral Presentation (SPIE)
- C15. B. Li, C. Xu, and Z. Duan. Audio-visual source association for string ensembles through multi-modal vibrato analysis. In *Sound and Music Computing*, 2017. Best Paper Award (SMC)
- C14. Y. Yan, C. Xu, D. Cai, and J. J. Corso. Weakly supervised actor-action segmentation via robust multi-task ranking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. AR: 28%; h5:140 (CVPR)
- C13. C. Xu and J. J. Corso. Actor-action semantic segmentation with grouping process models. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016. Spotlight Presentation AR: 9.7%; h5: 140 (CVPR)
- C12. C. Xu, S.-H. Hsieh, C. Xiong, and J. J. Corso. Can humans fly? action understanding with multiple classes of actors. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2015. AR: 28%; h5: 140 (CVPR)
- C11. C. Xu, S. Whitt, and J. J. Corso. Flattening supervoxel hierarchies by the uniform entropy slice. In *Proceedings of the IEEE International Conference on Computer Vision*, 2013. AR: 28%; h5: 92 (ICCV)
- C10. C. Xu, R. F. Doell, S. J. Hanson, C. Hanson, and J. J. Corso. Are actor and action semantics retained in video supervoxel segmentation? In *Proceedings of IEEE International Conference on Semantic Computing*, 2013. AR: 30%; h5: 15 (ICSC)
- C9. C. Xu, P. Das, R. F. Doell, and J. J. Corso. A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013. AR: 26%; h5: 140 (CVPR)
- C8. C. Xu, C. Xiong, and J. J. Corso. Streaming hierarchical video segmentation. In *Proceedings of European Conference on Computer Vision*, 2012. Oral Presentation AR: 2.8%; h5: 76 (ECCV)
- C7. C. Xu and J. J. Corso. Evaluation of super-voxel methods for early video processing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

²CVPR, ICCV and ECCV are premier conferences in computer vision. For each, typical number of submissions is around 2000 and the overall acceptance rate is around 27%. CVPR is the highest rated publication venue for computer vision and eighth-highest across all engineering and computer science, according to Google Scholar metrics.

WORKSHOP PAPERS

- W6. A. Newswanger and C. Xu. One-shot video object segmentation with iterative online fine-tuning. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017 (CVPRW)
- W5. L. Chen, S. Srivastava, Z. Duan, and C. Xu. Deep cross-modal audio-visual generation. In *ACM International Conference on Multimedia Thematic Workshops*, 2017. (ACMMMWW)
- W4. L. Zhou, C. Xu, P. Koch, and J. J. Corso. Watch what you just said: Image captioning with text-conditional attention. In *ACM International Conference on Multimedia Thematic Workshops*, 2017. (ACMMMWW)
- W3. S. Oh, A. Perera, I. Kim, M. Pandey, K. Cannons, H. Hajimirsadeghi, A. Vahdat, G. Mori, B. Miller, S. McCloskey, Y.-C. Cheng, Z. Huang, C.-H. Lee, C. Xu, R. Kumar, W. Chen, J. J. Corso, L. Fei-Fei, D. Koller, V. Ramanathan, K. Tang, A. Joulin, and A. Alahi. Trecvid 2013 genie: Multimedia event detection and recounting. In *NIST TRECVID Workshop*, 2013.
- W2. A. Perera, S. Oh, M. Pandey, T. Ma, A. Hoogs, A. Vahdat, K. Cannons, H. Hajimirsadeghi, G. Mori, S. McCloskey, B. Miller, S. Venkatesha, P. Davalos, P. Das, C. Xu, J. J. Corso, R. Srihari, I. Kim, Y.-C. Cheng, Z. Huang, C.-H. Lee, K. Tang, L. Fei-Fei, and D. Koller. Trecvid 2012 genie: Multimedia event detection and recounting. In *NIST TRECVID Workshop*, 2012.

BOOKS AND THESES

- T1. C. Xu. *Scale-Adaptive Video Understanding*. PhD thesis, University of Michigan, 2016.

FUNDING &
INDUSTRIAL GIFTS

TOTAL FUNDING: \$731,625 (\$681,625 as PI)
Funding is sorted by start date (recent first).

- F4. **PI:** *Audio-Visual Scene Understanding* \$650,000
SOURCE: NSF BIGDATA Program 9/2017-8/2021
COLLABORATORS: Duan (ECE)
OBJECTIVE: The main objective is to develop algorithms to achieve a human-like audio-visual bimodal scene understanding that overcomes the limitations in single-modality analysis through big data analysis of Internet Videos.
- F3. **Co-PI:** *Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech* \$50,000
SOURCE: University of Rochester AR/VR Pilot Program 8/2017-7/2018
COLLABORATORS: Maddox (PI, BME), Duan (ECE)
OBJECTIVE: Develop a real-time talking face by analyzing the correlations between auditory and visual signals to assist hearing-impaired people.
- F2. **PI:** *Video Understanding Beyond Human Activities* GPU Donation
SOURCE: NVIDIA 8/2017
OBJECTIVE: The main objective is to achieve a collaborative recognition of actors, objects being the agent of action, and their actions in videos that overcomes the limitations of separated or cascaded approaches.
- F1. **PI:** *Real-Time Streaming Video Object Detection* \$31,625
SOURCE: Markable, Inc. 8/2017-1/2018
OBJECTIVE: Develop an online real-time video object detection system that handles various video challenges, such as motion blur, video defocus, object occlusion and rare pose.

PROFESSIONAL
SERVICES

ORGANIZING COMMITTEE

- * IEEE CVPR Workshop *Brave New Ideas For Motion and Spatio-Temporal Representations* 2017

PROGRAM COMMITTEE

- * IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, 2017, 2018
* The AAAI Conference on Artificial Intelligence (AAAI) 2017, 2018
* IEEE International Conference on Computer Vision (ICCV) 2015, 2017
* British Machine Vision Conference (BMVC) 2015, 2016, 2017

- * International Conference on Semantic Computing (ICSC) 2017
- * Neural Information Processing Systems (NIPS) 2016
- * Asian Conference on Computer Vision (ACCV) 2016
- * International Symposium CompIMAGE'16 2016
- * IEEE/ISPRS 3rd Joint Workshop on Multi-Sensor Fusion for Dynamic Scene Understanding 2015
- * ISPRS Geospatial Week – Image Sequence Analysis 2015
- * XXIIIrd ISPRS Congress 2016
- * Indian Conference on Computer Vision, Graphics and Image Processing 2014

JOURNAL REVIEWER

- * IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)
- * IEEE Transactions on Image Processing (IEEE TIP)
- * IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT)
- * IEEE Transactions on Multimedia (IEEE TMM)
- * International Journal of Computer Vision (IJCV)
- * Pattern Recognition (PR)
- * Image and Vision Computing (IVC)
- * Computer Vision and Image Understanding (CVIU)
- * Signal Processing: Image Communication
- * Machine Vision and Applications
- * IET Computer Vision
- * OSA Biomedical Optics Express
- * Electronic Letters on Computer Vision and Image Analysis
- * IPSJ Transactions on Computer Vision and Applications

UNIVERSITY SERVICES	Computer Science PhD Admission Committee	2017, 2018
	Computer Science Website Committee	2018
	Data Science MS Admission Committee	2018

INVITED TALKS	When Computer Vision Meets Audition: <i>From Cross-Modal Generation to Audio-Visual Scene Understanding</i> - Rochester Institute of Technology	11/2017
	Towards the What, Who and Where of Video Understanding - The DAVIS Challenge on Video Object Segmentation, IEEE CVPR Workshop	7/2017
	Actor-Action Semantic Segmentation with Grouping Process Models - Midwest Vision Workshop	4/2016
	Scale-Adaptive Video Understanding - School of Computing, University of Utah	3/2016
	- Department of Computer Science, Stevens Institute of Technology	3/2016
	- Department of Computer Science, University of Rochester	3/2016
	- Department of Computer Science & Engineering, Michigan State University	2/2016
	Pulling Information from Scales - SPEECS Seminar, University of Michigan	1/2016
	Action Understanding with Multiple Classes of Actors - AI Lab Mini-Symposium, University of Michigan	5/2015
	Hierarchical Video Segmentation: Methods, Perception and Application - SPEECS Seminar, University of Michigan	9/2014
	LIBSVX and Video Segmentation Evaluation - IEEE CVPR 2014 Tutorial on Video Segmentation	6/2014
	Are Actor and Action Semantics Retained in Video Supervoxel Segmentation? - IEEE International Conference on Semantic Computing	9/2013

TEACHING	CSC 249/449 Machine Vision <i>University of Rochester, Spring 2018</i> Instructor for the cross-listed undergraduate- and graduate-level course that covers both basics and advances in modern computer vision.
	CSC 298/578 Deep Learning and Graphical Models

University of Rochester, Spring 2017

Instructor for the cross-listed undergraduate- and graduate-level course that covers both essentials in deep learning and probabilistic graphical models. It is a new CS course at UR.

CSC 577 Advanced Topics in Computer Vision

University of Rochester, Fall 2016, Fall 2017

Instructor for the graduate-level seminar course that covers special topics in computer vision.

EECS 598 Foundations of Computer Vision

University of Michigan, Fall 2015

Graduate student instructor for this course. Lecture tutorials and reviews, hold office hours, and grade student presentations.

Beyond the M.S. Mentoring Program

University of Michigan, Fall 2015

Encourage and provide guidance for master's students to continue on for a Ph.D.

STUDENT ADVISING	PHD ADVISEES	
	P3. Yapeng Tian	FALL 2017 - PRESENT
	P2. Jing Shi	FALL 2017 - PRESENT
	P1. Jie Chen (co-supervised with Jiebo Luo)	FALL 2017 - PRESENT
	PHD COMMITTEE	
	Chris Bates, Jianbo Yuan	
	MS STUDENTS	
	M5. Lele Chen	SPRING 2017 - PRESENT
	PROJECT: <i>Visual-Audio Generative Adversarial Networks</i>	
	M4. Li Ding	SUMMER 2017
	PROJECT: <i>Video Action Segmentation with LSTM and Attention Models</i>	
	M3. Sudhanshu Srivastava	SPRING, SUMMER 2017
	PROJECT: <i>Pose-Guided Human Activity Video Generation</i>	
	M2. Mingyang Zhou	SPRING 2016
	PROJECT: <i>Reconstruct Human Actions in Videos using Sparse Coding</i>	
	M1. Yao Li	SPRING 2013
	PROJECT: <i>Parallel Graph-Based Hierarchical Video Segmentation</i>	
	UNDERGRADUATES	
	U11. Amos Newswanger	SUMMER 2017, FALL 2017
	U10. Max Torop	SUMMER 2017
	U9. Zhiheng Li (Visiting Student from Wuhan University)	SUMMER 2017
	U8. Wei Zhao (Visiting Student from USTC)	SUMMER 2017
	U7. Wes Smith (REU Student from University of Edinburgh)	SUMMER 2017
	U6. Moses Bug (REU Student from Brandeis University)	SUMMER 2017
	U5. Shengqi Suizhu	FALL 2016
	U4. Austin Schaffer	SPRING 2014
	U3. Libing Wu	SPRING 2014
	U2. Spencer Whitt	SUMMER 2013
	U1. Tyler Ganter	SUMMER 2013

SOFTWARE &
DATA SETS

A2D 2015
A dataset and benchmark for action recognition and segmentation with multiple classes of actors. It considers seven actor classes (adult, baby, dog, etc.) and eight action classes (climb, crawl, eat, etc.). It contains 3782 videos with at least 99 instances per valid actor-action tuple.
<http://www.cs.rochester.edu/~cxu22/a2d/>

Video2Text.net 2013
A website and web-service for automatic conversion of videos to natural language sentences based on the video content. This website showcases our work in the vision+language domain.
<http://www.video2text.net>

YouCook 2013
Dataset of third-person cooking videos categorized into six styles of cooking and selected from open-source web videos of different kitchens and complexity levels. It contains object and action bounding boxes as well as multiple natural language descriptions of each video.
<http://www.cs.rochester.edu/~cxu22/youcook/>

LIBSVX 2012, 2013
Supervoxel library: a set of methods for early video processing by computing supervoxel segmentations as well as a quantitative benchmark for fair comparisons of those segmentations.
<http://www.supervoxels.com>
– *Winner Best Demo Prize at 2nd Greater New York Multimedia and Vision Meeting.* 6/2012
– *Winner Best Open Source Code 3rd Prize at IEEE CVPR 2012.* 6/2012

PAST POSITIONS

Research Assistant ANN ARBOR, MI
Electrical Engineering and Computer Science 9/2014 - 8/2016
University of Michigan
ADVISOR: Jason J. Corso
PRIMARY FOCUS: Investigating unified models of video representation integrating probabilistic methods, machine learning, and dynamic adaptive graphs to advance our ability to solve the video understanding problem.

Research Intern CUPERTINO, CA
NEC Lab American, INC. 5/2014 - 8/2014
MENTOR: Manmohan Chandraker
PROJECT: Video labeling and 3D localization for monocular road scene understanding.

Research Assistant BUFFALO, NY
Computer Science and Engineering 8/2011 - 5/2014
State University of New York at Buffalo
ADVISOR: Jason J. Corso
PRIMARY FOCUS: Improving the representation and indexing of objects and events in large-scale video analysis by efficiently encoding the low-level perceptual entities in the video and grounding them with rich high-level semantics.

REFERENCE LIST

Jason J. Corso (Ph.D. Advisor), Associate Professor, Department of Electrical Engineering and Computer Science, University of Michigan, Phone: (734) 647-8833, Email: jjcorso@eecs.umich.edu.
Irfan Essa, Professor, School of Interactive Computing, Georgia Institute of Technology, Phone: (404) 894-6856, Email: irfan@gatech.edu.
Benjamin Kuipers, Professor, Department of Electrical Engineering and Computer Science, University of Michigan, Phone: (734) 647-6887, Email: kuipers@umich.edu.
Jia Deng, Assistant Professor, Department of Electrical Engineering and Computer Science, University of Michigan, Phone: (734) 764-8517, Email: jiadeng@umich.edu.