Chenliang Xu

Prepared on August 13, 2018

Contact

3005 Wegmans Hall
250 Hutchison Road
University of Rochester
Rochester, NY 14627

Phone: (585) 275-5426
Email: chenliang.xu@rochester.edu
Web: www.cs.rochester.edu/~cxu22

Current Appointment

Assistant Professor — Rochester, NY
Department of Computer Science — 9/2016 - Present
University of Rochester

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 — Ann Arbor, MI
Ph.D. in Computer Science and Engineering — 8/2016
Advisor: Jason J. Corso
Dissertation Title: "*Scale-Adaptive Video Understanding*"

State University of New York at Buffalo — Buffalo, NY
M.S. in Computer Science and Engineering — 2/2012

Nanjing University of Aeronautics and Astronautics — Nanjing, China
B.S. in Information and Computing Science — 6/2010

Distinctions

National Science Foundation BIGDATA Award — 2017
Audio-Visual Scene Understanding

Best Paper Award at 14th Sound and Music Computing Conference — 2017

University of Rochester AR/VR Pilot Grant Award — 2017
Real-Time Synthesis of a Virtual Talking Face from Acoustic Speech

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

Peer-Reviewed Journal Articles[1], Conference and Workshop Proceedings[2]
Qualifiers added where known: Acceptance Rate (AR), Impact Factor (IF), h5-Index (h5) provided by Google Scholar.

P22. Y. Tian, J. Shi, B. Li, Z. Duan, and **C. Xu**. Audio-visual event localization in unconstrained videos. In *European Conference on Computer Vision*, 2018. *h5: 98* **(ECCV)**

P21. L. Chen, Z. Li, R. K. Maddox, Z. Duan, and **C. Xu**. Lip movements generation at a glance. In *European Conference on Computer Vision*, 2018. *h5: 98* **(ECCV)**

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

[2] CVPR, 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.

P20. L. Ding and **C. Xu**. Weakly-supervised action segmentation with iterative soft boundary assignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
AR: 28%; h5:158 **(CVPR)**

P19. S. E. Eskimez, R. K. Maddox, **C. Xu**, and Z. Duan. Generating talking face landmarks from speech. In *International Conference on Latent Variable Analysis and Signal Separation*, 2018.
(LVA-ICA)

P18. 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)**

P17. 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)**

P16. 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)**

P15. 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)**

P14. L. Chen, S. Srivastava, Z. Duan, and **C. Xu**. Deep cross-modal audio-visual generation. In *ACM International Conference on Multimedia Thematic Workshops*, 2017. **(ACMMMW)**

P13. 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. **(ACMMMW)**

P12. 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)

P11. 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)**

P10. T. Han, H. Yao, **C. Xu**, X. Sun, Y. Zhang, and J. J. Corso. Dancelets mining for video recommendation based on dance styles. *IEEE Transactions on Multimedia*, 19(4):712–724, 2017.

IF: 2.54; h5: 51 **(TMM)**

P9. **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)**

P8. **C. Xu** and J. J. Corso. Libsvx: A supervoxel library and benchmark for early video processing. *International Journal of Computer Vision*, 119(3):272–290, 2016. *IF: 4.27; h5: 63* **(IJCV)**

P7. **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)**

P6. **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)**

P5. **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)**

P4. **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)**

P3. P. Das, **C. Xu**, R. F. Doell, and J. J. Corso. A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013. *AR: 26%; h5: 140* **(CVPR)**

P2. **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)**

P1. **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.
AR: 24%; h5: 140 **(CVPR)**

Thesis and Technical Reports

T6. M. Chakraborty, **C. Xu**, and Andrew D White. Encoding and selecting coarse-grain mapping operators with hierarchical graphs. Technical report, arXiv:1804.04997, 2018.
(arXiv Report)

T5. L. Ding and **C. Xu**. Tricornet: A hybrid temporal convolutional and recurrent network for video action segmentation. Technical report, arXiv:1705.07818, 2017. **(arXiv Report)**

T4. **C. Xu**, C. Xiong, and J. J. Corso. Action understanding with multiple classes of actors. Technical report, arXiv:1704.08723, 2017. **(arXiv Report)**

T3. **C. Xu**. *Scale-Adaptive Video Understanding*. PhD thesis, University of Michigan, 2016.
(Ph.D. Thesis)

T2. 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.
(NIST Report)

T1. 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.
(NIST Report)

Funding

Total Funding: $1.7 million ($1.2 million as PI)
Funding is sorted by start date (recent first).

F7. **PI**: *RI: Small: Learning Dynamics and Evolution towards Cognitive Understanding of Videos* — $449,990, 9/2018-8/2021
Source: Robust Intelligence, NSF
Collaborators: Jiebo Luo (Co-PI, CS)
Objective: The main objective is to build computational models to study higher-order inference in understanding web instructional videos.

F6. **Co-PI**: *CDS&E: D3SC: Applying Video Segmentation to Coarse-grain Mapping Operators in Molecular Simulations* — $488,605, 8/2018-7/2021
Source: Division of Chemistry, NSF
Collaborators: Andrew White (PI, CE)
Objective: The main objective is to apply advances in computer vision, e.g., video segmentation and action recognition, to improve models of multiscale systems in chemistry.

F5. **PI**: *Deep Reinforcement Learning for Instructional Video Grounding* $60,000
Source: Tencent AI Lab Rhino-Bird Program 6/2018-5/2019
Objective: The main objective is to develop a novel deep reinforcement learning framework to perform higher-order inference in understanding web instructional videos.

F4. **PI**: *BIGDATA: F: Audio-Visual Scene Understanding* $650,000
Source: BIGDATA Program, NSF 9/2017-8/2021
Collaborators: Zhiyao Duan (Co-PI, 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: Ross K. Maddox (PI, BOE), Zhiyao Duan (Co-PI, 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 Actvities* GPU Donation
Source: NVIDIA GPU Grant Program 8/2017
Objective: The main objective is to achieve a joint recognition of actors and their actions in videos that overcomes the limitations of separated or cascaded approaches.

F1. **PI**: *Real-Time Streaming Video Object Detection* $30,000
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, 2019
* IEEE International Conference on Computer Vision (ICCV) 2015, 2017
* British Machine Vision Conference (BMVC) 2015, 2016, 2017, 2018
* International Conference on Semantic Computing (ICSC) 2017
* Neural Information Processing Systems (NIPS) 2016
* Asian Conference on Computer Vision (ACCV) 2016, 2018
* 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 Transcations on Pattern Analysis and Machine Intelligence (IEEE TPAMI)
* IEEE Transactions on Image Processing (IEEE TIP)
* IEEE Transcations 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

PhD Admission Committee, Department of Computer Science 2017, 2018, 2019

Lab Committee, Department of Computer Science 2019

Website Committee, Department of Computer Science 2018

MS Admission Committee, Goergen Institute for Data Science 2018

Invited Talks

When Computer Vision Meets Audition:
From Cross-Modal Generation to Audio-Visual Scene Understanding
- Data Science Summer Colloquium Series, University of Rochester 6/2018
- 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

NSF REU Site: Computational Methods for Understanding Music, Media, and Minds
University of Rochester, Summer 2017, Summer 2018
Teach a multi-day deep learning workshop to NSF REU site students.

Pre-College Data Science Program
Hajim Engineering, University of Rochester, Summer 2018
Teach a deep learning tutorial to a group of 17 high school students (including from South Korea, China, Spain and Haiti).

CSC 249/449 Machine Vision
University of Rochester, Spring 2018
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, Fall 2018
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

D5. Zhiheng Li — Start in Fall 2018

D4. Lele Chen — Start in Fall 2018

D3. Jie Chen (co-supervised with Jiebo Luo) — Fall 2017 - Present
D2. Jing Shi — Fall 2017 - Present
D1. Yapeng Tian — Fall 2017 - Present

PhD Committee
Chris Bates, Jianbo Yuan, Feng Yang, Shupeng Gui, Haofu Liao, Haichuan Yang, Taylan Sen

MS Students
M10. Jing Bi — Summer 2018
M9. Hao Huang — Summer 2018
M8. Wei Zhang — Summer 2018
M7. Wentian Zhao — Summer 2018
M6. Shaojie Wang — Summer 2018
M5. Lele Chen — Spring 2017 - Spring 2018
M4. Li Ding — Summer 2017
M3. Sudhanshu Srivastava — Spring, Summer 2017
M2. Mingyang Zhou — Spring 2016
M1. Yao Li — Spring 2013

Undergraduates
U15. Justin Goodman (REU Student from University of Maryland) — Summer 2018
U14. Marc Moore (REU Student from Mississippi State University) — Summer 2018
U13. Chenxiao Guan (Xerox Fellow Student) — Summer 2018
U12. Tianyou Xiao — Summer 2018
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

AVE 2018
Audio-Visual Event (AVE) dataset is a large video dataset that consists of 4143 10-second videos with both audio and video tracks for 28 audio-visual events and their temporal boundary annotations. It is the largest dataset for sound event detection.
`http://www.cs.rochester.edu/~cxu22/d/ave/`

YouCook2 2018
YouCook2 is the largest task-oriented, instructional video dataset in the vision community. It contains 2000 long untrimmed videos from 89 cooking recipes; on average, each distinct recipe has 22 videos. The procedure steps for each video are annotated with temporal boundaries and described by imperative English sentences.
`http://youcook2.eecs.umich.edu/`

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.