Dr. Yan Zhang

+41766389833
yz-cnsdqz.github.io/
5VpkLO8AAAAJ
yan-zhang-3119b11a
yan.zhang@inf.ethz.ch
yz-cnsdqz

Professional Summary

My current research is about graphics, computer vision, and machine learning, in particular on the digital human technologies. My recent works focus on full body motion modeling, global motion generation, vision-based performance capture in 3D human-scene interaction scenarios, as well as their applications in mixed reality, game production, and other fields. I am extending my research by incorporating physical simulation, large language models, gaze modeling, neural rending, etc., towards building intelligent digital avatars that have lifelike behaviors.

Academic Positions

Postdoc Researcher — *ETH Zurich*
Aug. 2020 - Present — *Computer Vision and Learning Group, led by Prof. Dr. Siyu Tang*

Research Assistant — *ETH Zurich*
Jan.2020 - Jul. 2020 — *Computer Vision and Learning Group, led by Prof. Dr. Siyu Tang*

Research Assistant — *Max-Planck Institute for Intelligent Systems*
Oct. 2018 - Jan. 2020 — *Perceiving Systems Department, led by Dr. Michael J. Black*

Education

Ph.D (Dr.rer.nat.) in Computer Science — *Dec 2015 – July 2020*
University of Ulm, Germany

Supervisor: Prof. Dr. Heiko Neumann
Dissertation: Human Action Parsing in Untrimmed Videos and its Applications for Elderly People Healthcare
Grade: Sehr gut (magna cum laude).

Computer Science Graduate School — *Apl 2011 – Dec. 2015*
Saarland University, Germany

M.Sc in Electronic Engineering — *Sep. 2009 – Nov. 2010*
University of Manchester, UK

Advanced Control and System Engineering

B.Eng in Mechanical Engineering — *Sep. 2005 – Jul. 2009*
Southwest Jiaotong University, China

Mechanical Design, Manufacturing, and Automation

Awards

Best paper award, International Conference on 3D Vision (3DV), 2020

Teaching

Co-supervised Master theses and semester projects

- Siwei Zhang: Proximity learning of articulation and contact in 3D environments, 2020
- Dexin Yang: 4D Human Body Capture from Egocentric Video via 3D Scene Grounding, 2020
- Dexin Yang: Physically Plausible Human Motion Recovery with Motion Prior, 2021
- Yan Wu & Jiahao Wang: Stochastic Whole-Body Grasping with Contact, 2021
- Jonathan Lehner: Virtual human navigation in large-scale CAD architecture models, 2022
- Kaifeng Zhao: Semantically Controllable Human-Scene Interaction Synthesis, 2022
- Lukas Bösiger: Navigating virtual humans in mixed reality, 2022.
- Yelan Tao: Virtual Humans meet Real Drones: Drone flight simulation in populated scenes, 2023
- Haoliang Shang: Retargeting SMPL-X bodies to Metahumans, 2023
- Chuqiao Li: 3D human pose estimation from egocentric images, 2023

Lectures

- Human motion modeling I and II, Virtual Humans at ETH Zurich, 2022
- The wanderings of Odysseus in 3D scenes, Mixed Reality at ETH Zurich, 2022

Service and Activity

Review and Area Chair

- Area Chair, 3DV, 2024
- CVPR (2022, 2021), ICCV (2021), ECCV (2022), WACV (2022, 2021), 3DV (2022, 2021), Siggraph Asia (2023), TPAMI (2019), EuroGraphics (2023)

Organizing Committee

- HBHA workshop, ECCV, 2022

Other Affiliations

- Guest scientist, Max-Planck Institute for Intelligent Systems, 2020-2022
- Affiliated Postdoc, Max-Planck ETH center for learning systems (CLS), 2020-Present

Invited Talks

- The Wanderings of Odysseus in 3D scenes, Rank Symposium on Neural Rendering in Computer Vision, UK, 2022
- Data-driven human motion generation, ETH AI Center faculty workshop on Robotics and Autonomous Systems, 2022.
- Human motion modeling, Huawei Tengman Technical Symposium, 2022
- Human Motion Capture and Synthesis, Google Research Zurich, 2021.

Publications (reverse chronological order)

[1] Synthesizing Diverse Human Motions in 3D Indoor Scenes.
K. Zhao, **Y. Zhang**, S. Wang, T. Beeler, S. Tang
International Conference on Computer Vision (ICCV), 2023.

[2] Probabilistic Human Mesh Recovery in 3D Scenes from Egocentric Views.
S. Zhang, Q. Ma, **Y. Zhang**, S. Aliakbarian, D. Cosker, S. Tang
International Conference on Computer Vision (ICCV), 2023. **Oral Presentation.**

[3] Egobody: Human body shape and motion of interacting people from head-mounted devices.
S. Zhang, Q. Ma, **Y. Zhang**, Z. Qian, T. Kwon, M. Pollefeys, F. Bogo, S. Tang
European Conference on Computer Vision (ECCV), 2022.

[4] Compositional human-scene interaction synthesis with semantic control.
K. Zhao, S. Wang, **Y. Zhang**, T. Beeler, S. Tang
European Conference on Computer Vision (ECCV), 2022.

[5] Saga: Stochastic whole-body grasping with contact.
Y. Wu, J. Wang, **Y. Zhang**, S. Zhang, O. Hilliges, F. Yu, S. Tang
European Conference on Computer Vision (ECCV), 2022.

[6] The wanderings of odysseus in 3d scenes.
Y. Zhang, S. Tang
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[7] 4d human body capture from egocentric video via 3d scene grounding.
M. Liu, D. Yang, **Y. Zhang**, Z. Cui, J.M. Rehg, S. Tang
International conference on 3D vision (3DV), 2021.

[8] Learning motion priors for 4d human body capture in 3d scenes.
S. Zhang, **Y. Zhang**, F. Bogo, M. Pollefeys, S. Tang
International Conference on Computer Vision (ICCV), 2021. **Oral Presentation.**

[9] LEAP: Learning articulated occupancy of people.
M. Mihajlovic, **Y. Zhang**, M.J. Black, S. Tang
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[10] We are more than our joints: Predicting how 3d bodies move.
Y. Zhang, M.J. Black, S. Tang
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[11] PLACE: Proximity learning of articulation and contact in 3D environments.
S. Zhang, **Y. Zhang**, Q. Ma, M.J. Black, S. Tang
International Conference on 3D Vision (3DV), 2020.

[12] Grasping field: Learning implicit representations for human grasps.
K. Karunratanakul, J. Yang, **Y. Zhang**, M.J. Black, K. Muandet, S. Tang
International Conference on 3D Vision (3DV), 2020. **Best Paper Award**

[13] Generating 3d people in scenes without people.
Y. Zhang, M. Hassan, H. Neumann, M.J. Black, S. Tang
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. **Oral Presentation**

[14] Local temporal bilinear pooling for fine-grained action parsing.
Y. Zhang, S. Tang, K. Muandet, C. Jarvers, H. Neumann
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[15] Multi-Modal Pain Intensity Recognition Based on the SenseEmotion Database.
P. Thiam, V. Kessler, M. Amirian, P. Bellmann, G. Layher, **Y. Zhang**, M. Velana, S. Gruss, S. Walter, H.C. Traue, D. Schork, J. Kim, E. Andé, H. Neumann, F. Schwenker
IEEE Transactions on Affective Computing, 2019.

[16] Human Motion Parsing by Hierarchical Dynamic Clustering.
Y. Zhang, S. Tang, H. Sun, H. Neumann
British Machine Vision Conference (BMVC), 2018.

[17] An empirical study towards understanding how deep convolutional nets recognize falls.
Y. Zhang, H. Neumann
European Conference on Computer Vision Workshop(ECCV), 2018.

[18] Visual confusion recognition in movement patterns from walking path and motion energy.
Y. Zhang, G. Layher, S. Walter, V. Kessler, H. Neumann
International Conference on Smart Homes and Health Telematics (ICOST), 2017.

[19] Continuous activity understanding based on accumulative pose-context visual patterns.
Y. Zhang, G. Layher, H. Neumann
International Conference on Image Processing Theory, Tools and Applications (IPTA), 2017.

[20] Tissue classification for laparoscopic image understanding based on multispectral texture analysis.
Y. Zhang, S. J Wirkert, J. Iszatt, H. Kenngott, M. Wagner, B. Mayer, C. Stock, N.T. Clancy, D. S Elson, L. Maier-Hein
SPIE Journal of Medical Imaging, 2017.

[21] Improving two-thumb text entry on touchscreen devices.
A. Oulasvirta, A. Reichel, W. Li, **Y. Zhang**, M. Bachynskyi, K. Vertanen, P.O. Kristensson
Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), 2013.

Projects (selective)

Inhabiting the virtual, Flight Assembled Architecture Revisited — *ETH Zurich*

2021-now

- **Has been featured at the home page of ETH Zurich.**
- Collaboration with Gramazio Kohler Research, architecture department of ETH Zurich.
- A large digital city is populated by diverse digital humans wandering autonomously, powered by generative motion models and RL-based control.
- Developed an system based on Nvidia Omniverse, which synthesizes human motions online.
- On-site exhibition in Guggenheim Museum Bilbao (2022) and Autostadt Wolfsburg (2023).
- As a byproduct, a Hololens 2-based software is developed to place virtual humans in motion into the ETH main building.
- **MY ROLE:** project leader at the computer science department side.

Markerless Interaction Capture in Immersive Design Lab — *ETH Zurich*

2022-now

- Collaboration with Immersive Design Lab (IDL), architecture department of ETH Zurich.
- Capturing interactive behaviors of people in IDL, based on multiview RGB cameras.
- Extending to volumetric capture system for novel view synthesis.
- Extending to multi-modal capture system, including audio, point cloud, etc.
- **MY ROLE:** project leader at the computer science department side.

Interaction Capture for Mixed Reality — *ETH Zurich*

2021-now

- Capturing human-human interactions and human-scene interactions based on multiview RGBD sensors and Microsoft Hololens
- Funded by Microsoft Swiss Joint Research Center
- An egocentric interaction capture dataset EgoBody has been created.
- **MY ROLE:** advisor and collaborator

SenseEmotion — *Ulm University*

2015-2018

- Pain recognition, behavior understanding, face analysis, etc. for elderly people healthcare.

- Multimodal capture system including cameras, IMU, audio, and physiological sensors was developed, and pilot study was performed at the clinic.
- **MY ROLE:** responsible for developing multi-modal body capture system and developing novel action understanding methods

Technical skills

Coding and Software	PyTorch, Python, C++/CUDA, Blender, Unity, Nvidia Omniverse