

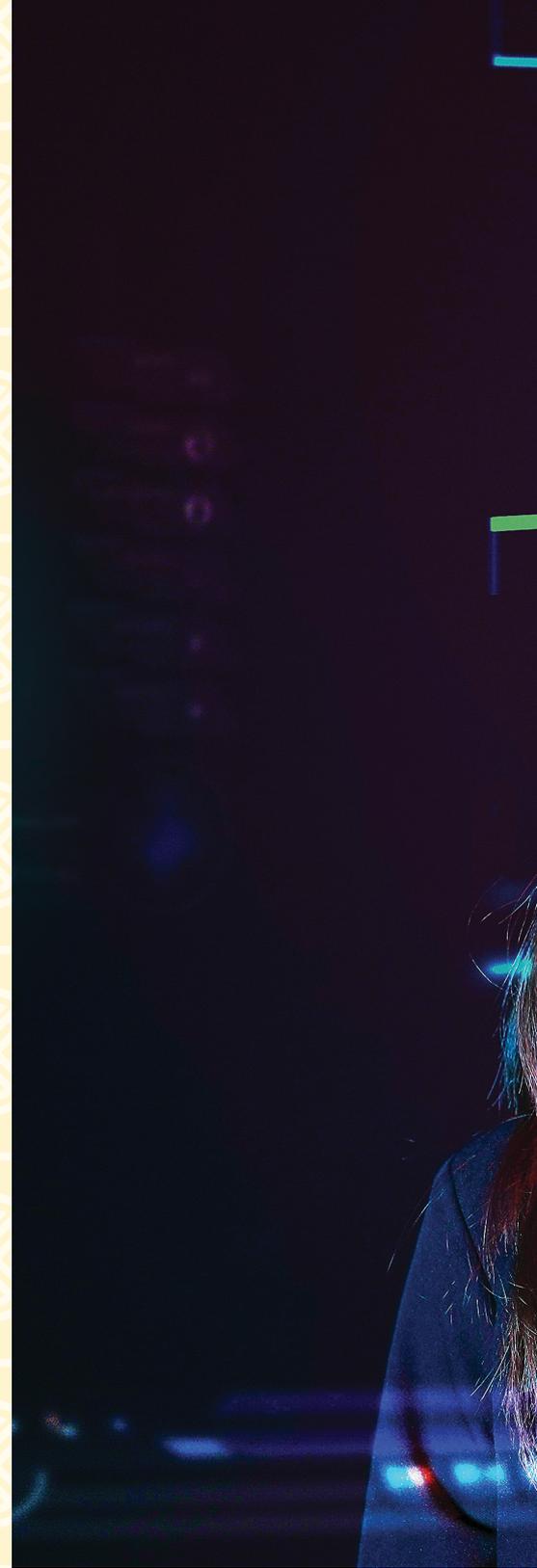
BY GUOFENG ZHANG, XIAOWEI ZHOU, FENG TIAN,
HONGBIN ZHA, YONGTIAN WANG, AND HUJUN BAO

The Present and Future of Mixed Reality in China

VIRTUAL REALITY (VR) technology can be used to generate a lifelike virtual world, allowing users to roam freely in this virtual world and interact with it. Augmented reality (AR) is a technology that superimposes and fuses virtual scenery or information with the real physical environment, and interactively presents it in front of users, so that the virtuality and reality share the same space. Mixed reality (MR) is a further development of VR and AR. Its concept was proposed by Paul Milgram and Fumio

Kishino in 1994, who described the entire continuum from reality to virtuality.¹⁴

VR/AR research in China began in the late 1990s. In 2002, the 973 project (National Basic Research Program) “Fundamental Theories, Algorithms of Virtual Reality and Its Implementation” was established, with Hujun Bao from Zhejiang University as the chief scientist. In





2009, the project received continuous support, and the focus was shifted to AR and MR. The research of these two projects lasted for 10 years. Zhejiang University, the Institute of Software and the Institute of Automation of the Chinese Academy of Sciences, Beijing Institute of Technology (BIT), Beihang University, Tsinghua University, and other institutions participated in these

two projects. The projects made many world-class research advances in VR/AR related areas such as automatic camera tracking, efficient modeling, real-time rendering, multichannel human-computer interaction, VR/AR headset display, 3D registration, as well as VR/AR engines and software platforms. The projects also trained a group of outstanding VR/AR

researchers, which significantly boosted VR/AR research in China.

With the continuous investment of national funding agencies and research institutions, the quality of China's MR research has developed rapidly. The number of publications from Chinese researchers in top conferences such as IEEE VR and the International Symposium on Mixed and Augmented Reality

(ISMAR) surged. For example, the first-author affiliations of 18.5% of accepted papers for ISMAR 2020, a top conference of AR and MR, are from China, rising from only 4% in 2010. Moreover, researchers from SenseTime and Zhejiang University received the best paper award of ISMAR 2020 for their work Mobile-3DRecon,²⁵ which was the first time authors from Chinese institutions won this award. With the increasing research impact, more and more Chinese researchers have been invited to join the organization committees of top conferences. In 2019, ISMAR was held in China for the first time. Qingping Zhao from Beihang University and Yongtian Wang from BIT served as general chairs, and Shi-Min Hu from Tsinghua University served as the science and technology chair.

Meanwhile, the nation has made further plans for the MR industry. In December 2018, the Ministry of Industry and Information Technology of China issued “Guiding Opinions on Accelerating the Development of the Virtual Reality Industry.” It proposed establishing “a relatively complete virtual reality industry chain” in China by 2020, and achieving the goal of “the overall strength of China’s virtual reality industry will be in the forefront of the world by 2025.” In March 2021, China’s 14th Five-Year Plan was officially announced, listing VR and AR as key industries in the digital economy over the next five years.

Research Contributions from Chinese Scholars in MR

To realize the immersive visual fusion and presence of virtual and real environments, key technologies such as tracking and registration, 3D modeling, realistic rendering, human-computer interaction, and natural display need to be developed. Next, we will introduce the current state of research in these key technologies in China.

3D registration and reconstruction. For a mixed reality system, 3D registration technology mainly aims to reconstruct the 3D information of the real scene and the real-time pose information of the user or the

camera. Simultaneous localization and mapping (SLAM) is a most important 3D registration technique, which realizes the inside-out tracking and localization. Although the research on SLAM in China started late, it has also made remarkable achievements. For example, the RDSLAM¹⁶ and RKSLAM¹³ proposed by Bao’s team at Zhejiang University are both well-known monocular visual SLAM systems. The former can handle dynamic environments, and the latter can run on a smartphone in real-time. The VINS-Mono¹⁵ proposed by Shaojiao Shen’s team from the Hong Kong University of Science and Technology (HKUST) has become one of the most popular open-source visual-inertial SLAM systems due to its remarkable robustness and versatility. ICE-BA,¹² which was jointly developed by Baidu and Zhejiang University, increases the speed of bundle adjustment (BA) by an order of magnitude using the incremental computation. Some Chinese companies have also successively launched SLAM-based AR/MR platforms, for example, SenseTime’s SenseAR^a and Huawei’s AR Engine.^b

With the progress of deep learning in recent years, some deep learning-based SLAM methods have also emerged. However, these methods typically require collection of large-scale training data, and are generally difficult to generalize to environments that have changed or have never been seen before. In order to solve this issue, the team of Hongbin Zha from Peking University proposed a SLAM framework based on the online learning paradigm, which enables the SLAM system to infer uncertainty and quickly adapt itself in a rapidly changing environment.^{11,22}

In the area of 3D reconstruction, both academia and industry in China made great progress. The ENFT-SfM²⁸ proposed by Zhejiang University achieves fast and robust large-scale scene reconstruction through an efficient non-consecu-

tive feature tracking method and segment-based bundle adjustment algorithm. The HSfM² proposed by the Institute of Automation of the Chinese Academy of Sciences achieves similar accuracy to the incremental reconstruction through a hybrid structure-from-motion method, and at the same time reaches an efficiency close to or even better than that of the global reconstruction. The ultra-large-scale global 3D reconstruction system proposed by HKUST realizes a distributed global motion averaging through a divide-and-conquer approach and achieves efficient 3D reconstruction from tens of thousands of images.³⁰ In industry, a series of 3D reconstruction products have emerged, such as Altizure, DJI Terra, AirlookMap, and others, which can automatically reconstruct high-precision 3D models from image data taken by drones.

Simulation and rendering.

Computer simulation has been developed for two decades in China. In the 1990s, Jiaoying Shi, Qunsheng Peng, and Enhua Wu started their pioneering research, covering considerable subjects in AR/VR. Since then, Chinese scholars have been devoting greater effort and making significant contributions. For complex elasticity, Bao’s team has proposed a series of technologies (for example, Huang et al.⁶) to greatly improve the simulation efficiency. For fluids, Enhua Wu, Guoping Wang, Shi-Min Hu, and Xiaopei Liu and their teams proposed novel methods (for example, Yan et al.²³) to efficiently simulate complex fluid phenomena including multi-phase fluids. Kun Zhou and his team focused on human bodies and solved fidelity problems in hair and face simulation.¹ The teams of Qingping Zhao, Xiaogang Jin, Min Tang, and Mingliang Xu proposed to leverage the ideas of space-time continuity to address problems in medical simulation, group animation, and cloth simulation. Wang et al.¹⁷ developed haptic devices and algorithms in VR, overcoming the problems caused by high frame rates.

In generalized real-time rendering, significant achievements have

a <http://openar.sensetime.com/>

b <https://developer.huawei.com/consumer/en/hms/huawei-arengine/>

been seen in stereo rendering, automatic shader optimization, and global illumination. Stereo and power-aware rendering is an important research direction to provide realistic content for head-mounted displays, which have limited computational resources. Bao's team developed a cutting-edge stereo shading and shader optimization architecture,²⁶ which is promising to meet the ever-increasing requirements of high-framerate and high-resolution for immersive VR. Neural rendering is a prominent direction for providing premium effects in a uniform framework. Zhejiang University, Tsinghua University, and Kujiale contributed leading solutions in this domain, ranging from real-time single-bounce indirect illumination, denoising, to path guiding.²⁰ For the rendering of specific material and effects, the teams from Zhejiang University and Nanjing University introduced state-of-the-art real-time techniques^{21,4} for rendering cloth and participating media, respectively.

Human-computer interaction.

Human-computer interaction is a key component in MR research, which involves various aspects including theories, devices, and techniques. In the area of theory, the team of Feng Tian from the Institute of Software, Chinese Academy of Sciences (ISCAS) proposed an uncertainty model based on the ternary Gaussian probability distribution.^{7,8} It can accurately predict the target acquisition error rate and has an important guiding role in the interaction designs in MR. By combining touch-screen gestures and a variety of tactile feedback mechanisms based on electrostatic force, Zhao et al.²⁹ designed a multi-channel visual- and touch-based virtual reality interactive system for 3D object interaction tasks, which improves the accuracy of 3D object operations.

In the research of interactive devices, the team of Yingqing Xu from Tsinghua University made a breakthrough in the 1:N scanning drive and push-push contact latch technology, which effectively solves the problem of large-format render-

ing with tactile dot-matrix feedback devices, and effectively improves the resolution and stability. Wang et al.¹⁹ developed wearable tactile feedback gloves that can achieve 0-4N continuous and stable force feedback.

In the research of interactive techniques, the team of Yongtian Wang from BIT proposed a calibration method based on a dynamic pinhole camera model to solve the problem of precise virtual-real fusion in close-range high-precision hand-eye collaborative interaction. ISCAS designed an interactive component (vMirror) for interactions in VR, which uses the reflection of the mirror to observe and select long-distance occluded objects, improve the selection efficiency of occluded targets, and reduce user dizziness.¹⁰

Optical display of VR/AR devices.

Head-mounted display (HMD) is an important carrier device of VR and AR. HMD for VR (VR-HMD) has been deployed in commercial applications earlier than AR-HMD, and the technical solution has become increasingly mature.

Due to disadvantages like small field of view (FOV), inferior image quality, and heavy design, the applications of early AR-HMD with off-axis mirror and relay lens are limited. The introduction of freeform surfaces not only greatly increases the design freedom, but also significantly reduces the volume and weight of AR-HMDs. Zhejiang University, Nankai University, and BIT conducted in-depth research on freeform optics. The team of Yongtian Wang from BIT proposed a closed-loop optimization design method that integrates full-FOV image quality balance and injection error pre-compensation. The freeform element they developed weighs only 8g, which greatly reduces the weight of an AR-HMD.³ In 2018, Ned+ announced the boundless AR optics module with freeform optics,^c which has a diagonal FOV of 120°.

Chinese researchers have also done excellent work to further reduce the volume and weight of AR-HMD.

c <http://www.nedplusar.com/en/index>



Terminal-cloud collaboration through 5G networks will make it possible to achieve the high-resolution, frame rate, and fidelity rendering of large-scale virtual scenes on mobile or head-mounted MR devices.



Stereo and power-aware rendering is an important research direction to provide realistic content for head-mounted displays, which have limited computational resources.

In 2015, the team of Qiang Sun from Changchun Institute of Optics and Mechanics proposed a waveguide structure with two vertically arranged half-reflective films and achieved a FOV of $20^{\circ} \times 15^{\circ}$ and two-dimensional expansion of the exit pupil.⁹ In 2020, based on a dual-layer geometrical waveguide, Wang et al.¹⁸ from BIT proposed an ultra-thin, large-FOV AR-HMD which achieved a total thickness of 3.0 mm, 62° FOV, and 10-mm exit pupil at an eye relief of 18 mm.

An AR-HMD based on diffractive optical elements has also been developed. In 2011, Yan et al.²⁴ proposed a method of dispersion-free diffraction using four holographic gratings optimization method for holographic waveguide, which achieved a circular FOV of 25° and a large pupil of about 43 mm. In 2015, Han et al.⁵ from BIT designed a waveguide display system composed of freeform elements and volume holographic gratings with a diffraction efficiency of 87.57%, which achieves a diagonal FOV of 45° . In 2017, based on space-variant volume holographic gratings, Chao et al.²⁷ proposed a quite efficient waveguide display which achieved 31.9% system efficiency, 20° FOV, and high brightness uniformity simultaneously.

The Mixed Reality Industry in China

In recent years, many VR/AR startup companies have emerged in China. Quite a few of them are focused on VR/AR all-in-one systems or core component devices. For example, Pico, QiYu VR, NOLO, and others have launched VR all-in-one systems. Shadow Creator, Nreal, and more have launched birdbath-based AR glasses. Collaborating with Ned+, BIT launched a variety of AR-HMD products based on freeform surface optics. Ned+, Lochn Optics, Lingxi, Rokid, Greater, North Ocean Photonics and others have launched waveguide AR-HMD. In recent years, Chinese start-up companies appeared at the Consumer Electronics Show and the Augmented Reality World Expo and won a series of awards.

In addition to hardware, a few tech giants in China have also launched AR/MR technology platforms. Baidu released its open platform DuMix AR^d in 2017, empowering developers with AR technology. Huawei released the AR Engine for Huawei mobile phones in 2018 and its spatial computing platform Cyberspace in 2019. SenseTime released the SenseAR developer platform in 2018 and upgraded it in 2020 to SenseMARS, a cross-hardware and cross-system MR platform providing high-precision 3D mapping and spatial computing capabilities. Compared with the international leading products, some of China's MR products (such as SenseMARS and Cyberspace) already can achieve almost comparable MR effects in large-scale scenes, and even have some differentiated advantages in some respects. For example, to the best of our knowledge, the MR platform SenseMARS jointly developed by SenseTime and Zhejiang University is the first product in the industry that is able to achieve high-precision 6DoF visual-inertial localization and AR navigation in large-scale indoor scenes on the Web and mini programs.

The development of MR in China is at a relatively early stage and mainly based on scenarios involving camera apps, short video, and live broadcasting for applications such as house viewing and navigation guides. In 2015, AR effects selfie camera app "FaceU" was released and quickly became popular and was acquired in 2018 by ByteDance for about US\$300 million. Douyin and Kuaishou have added AR special effects to their short video applications. In the real estate sector, Beike took the lead in launching the VR house viewing function in 2018. In e-commerce, Alibaba and JD.com launched the buy+ and Tiangong virtual shopping plans in 2016 and 2017 respectively to provide e-commerce with digital AR content, though they are not yet open to large-scale third-party users. In the past two years, more and more companies, including Didi,

^d <https://dumix.baidu.com/>

Huawei, SenseTime, and Baidu, have launched visual localization with AR navigation solutions or applications.

Future Perspectives

Although there is still a certain gap between China's MR products and international leaders, the gap is narrowing rapidly. In particular, with the popularization of 5G and the rapid development of cloud computing in China, high-bandwidth and low-latency networks will greatly push MR technologies towards the combination of terminal and cloud computing, and city-level MR will be realized in the near future. For example, SLAM technology can be combined with high-precision 3D maps and cloud computing to break through the bottleneck of robustness and efficiency in large-scale scenes. Also, the terminal-cloud collaboration through 5G networks will make it possible to achieve the high-resolution, frame rate, and fidelity rendering of large-scale virtual scenes on mobile or head-mounted MR devices. Zhejiang University has made outstanding contributions in the related research.

In addition, key technologies in MR, such as rendering, simulation, 3D modeling, and interaction are also being deeply integrated with AI in China. It is expected that in the next few years, China's MR products will not only work in a much larger environment but will also become more high-fidelity and in effect smarter. For instance, neural scene representation and rendering show promise to break through the limitations of traditional graphics pipelines. Tsinghua University, Microsoft Research Lab Asia, and Zhejiang University have made remarkable contributions. Moreover, current MR systems still lack sufficient intelligence to allow group participation and collaboration. Deep integration with AI is required. Some well-known companies in China—for example, Huawei and SenseTime—have devoted huge effort to MR+AI research and products.

Last but not least, the sense of

immersion, ease of use, and wearing comfort of MR equipment are essential for commercialization of MR. Optical display technology is the key. The future optical solutions of VR-HMD will mainly focus on aspheric lens VR-HMD, Fresnel lens VR-HMD, and pancake VR-HMD. AR-HMD is more challenging. Free-form AR-HMD can expand the field of view with high image quality, and commercial mass production is possible in China. Geometric waveguides and diffractive waveguides can achieve a thinner system structure. However, the former is difficult to mass produce. With China's progress in high-precision optical component processing and design, diffractive optical waveguides, in terms of optical effects, appearance, and mass production prospects, will become the mainstream of AR-HMD in China. 

References

- Cao, C., Weng, Y., Lin, S., Zhou, K. 3D shape regression for real-time facial animation. *ACM Trans. Graphics* 32, 4 (2013), 1–10.
- Cui, H., Gao, X., Shen, S., Hu, Z. HSfM: Hybrid structure-from-motion. In *Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition*, 2017, 1212–1221.
- Cheng, D., Wang, Y., Hua, H., Talha, M. Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism. *Applied Optics* 48, 14 (2009), 2655–2668.
- Guo, J., Guo, Y., Pan, J., Lu, W. BRDF analysis with directional statistics and its applications. *IEEE Trans. Visualization and Computer Graphics* 26, 3 (2020), 1476–1489.
- Han, J., Liu, J., Yao, X., Wang, Y. Portable waveguide display system with a large field of view by integrating freeform elements and volume holograms. *Optics Express* 23, 3 (2015), 3534–3549.
- Huang, J., Tong, Y., Zhou, K., Bao, H., Desbrun, M. Interactive Shape Interpolation through Controllable Dynamic Deformation. *IEEE Trans. Visualization and Computer Graphics* 17, 7 (2011), 983–992.
- Huang, J., Tian, F., Fan, X., Zhang, X., Zhai, S. Understanding the uncertainty in 1D unidirectional moving target selection. In *Proceedings of the ACM Conf. Human Factors in Computing Systems*, 2018, 1–12.
- Huang, J., Tian, F., Fan, X., Tu, H., Zhang, H., Peng, X., Wang, H. Modeling the endpoint uncertainty in crossing-based moving target selection. In *Proceedings of the 2020 CHI Conf. Human Factors in Computing Systems*, 2020, 1–12.
- Hu, X., Sun, Q., Li, J., Li, C., Liu, L., Zhang, J. Spectral dispersion modeling of virtually imaged phased array by using angular spectrum of plane waves. *Optics Express* 23, 1 (2015), 1–12.
- Li, N., Zhang, Z., Liu, C., Yang, Z., Fu, Y., Tian, F., Han, T., Fan, M. vMirror: enhancing the interaction with occluded or distant objects in VR with virtual mirrors. In *Proceedings of 2021 CHI Conf. Human Factors in Computing Systems*, 1–11.
- Li, S., Wang, X., Cao, Y., Xue, F., Yan, Z., Zha, H. Self-supervised deep visual odometry with online adaptation. In *Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition*, 2020, 6339–6348.
- Liu, H., Chen, M., Zhang, G., Bao, H., Bao, Y. ICE-BA: incremental, consistent and efficient bundle adjustment for visual-inertial SLAM. In *Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition*, 2018, 1974–1982.
- Liu, H., Zhang, G., Bao, H. Robust keyframe-based monocular SLAM for augmented reality. In *Proceedings of the IEEE Intern. Symp. Mixed and Augmented Reality*, 2016, 1–10.
- Milgram, P., Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems* E77-D 12, (1994), 1321–1329.
- Qin, T., Li, P., Shen, S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robotics* 34, 4 (2018), 1004–1020.
- Tan, W., Liu, H., Dong, Z., Zhang, G., Bao, H. Robust monocular SLAM in dynamic environments. In *Proceedings of the IEEE Intern. Symp. on Mixed and Augmented Reality*, 2013, 209–218.
- Wang, D., Zhang, Y., Hou, J., Wang, Y., Lv, P., Chen, Y., Zhao, H. iDental: A haptic-based dental simulator and its preliminary user evaluation. *IEEE Trans. Haptics* 5, 4 (2011), 332–343.
- Wang, Q., Cheng, D., Hou, Q., Gu, L., and Wang, Y. Design of an ultra-thin, wide-angle, stray-light-free near-eye display with a dual-layer geometrical waveguide. *Optics Express* 28, 23 (2020), 35376–35394.
- Wang, Z., Wang, D., Zhang, Y., Liu, J., Wen, L., Xu, W., Zhang, Y. A three-fingered force feedback glove using fiber reinforced soft bending actuators. *IEEE Trans. Industrial Electronics* 67, 9 (2019), 7681–7690.
- Xu, B., Zhang, J., Wang, R., Xu, K., Yang, Y., Li, C., Tang, R. Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation. *ACM Trans. Graphics* 38, 6 (2019), 224:1–224:12.
- Xu, C., Wang, R., Zhao, S., Bao, H. Multi-scale hybrid micro-appearance modeling and realtime rendering of thin fabrics. *IEEE Trans. Visualization and Computer Graphics* 27, 4 (2019), 2409–2420.
- Xue, F., Wang, X., Wang, J., Zha, H. Deep visual odometry with adaptive memory. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2020.
- Yan, X., Jiang, Y., Li, C., Martin, R., Hu, S. Multiphase SPH simulation for interactive fluids and solids. *ACM Trans. Graphics* 35, 4 (2016), 1–11.
- Yan, Z., Li, W., Zhou, Y., Kang, M., Zheng, Z. Virtual display design using waveguide hologram in conical mounting configuration. *Optical Engineering* 50, 9 (2011) 94001.
- Yang, X., Zhou, L., Jiang, H., Tang, Z., Wang, Y., Bao, H., Zhang, G. Mobile3DRecon: real-time monocular 3D reconstruction on a mobile phone. *IEEE Trans. Visualization and Computer Graphics* 26, 12 (2020), 3446–3456.
- Yuan, Y., Wang, R., Bao, H. Tile pair-based adaptive multi-rate stereo shading. *IEEE Trans. Visualization and Computer Graphics* 26, 6 (2020), 2303–2314.
- Yu, C., Peng, Y., Zhao, Q., Li, H., and Liu, X. Highly efficient waveguide display with space-variant volume holographic gratings. *Applied Optics* 56, 34 (2017), 9390–9397.
- Zhang, G., Liu, H., Dong, Z., Jia, J., Wong, T., Bao, H. Efficient non-consecutive feature tracking for robust structure-from-motion. *IEEE Trans. Image Processing* 25, 12 (2016), 5957–5970.
- Zhao, L., Liu, Y., Ye, D., Ma, Z., Song, W. Implementation and evaluation of touch-based interaction using electrovibration haptic feedback in virtual environments. In *Proceedings of the 2020 IEEE Conf. Virtual Reality and 3D User Interfaces*, 239–247.
- Zhu, S., Zhang, R., Zhou, L., Shen, T., Fang, T., Tan, P., Quan, L. Very large-scale global SfM by distributed motion averaging. In *Proceedings of the 2018 IEEE Conf. Computer Vision and Pattern Recognition*, 4568–4577.

Guofeng Zhang is a professor at Zhejiang University, Hangzhou, China.

Xiaowei Zhou is a research professor at Zhejiang University, Hangzhou, China.

Feng Tian is a professor at the Institute of Software, Chinese Academy of Sciences, Beijing, China.

Hongbin Zha is a professor at Peking University, Beijing, China.

Yongtian Wang is a professor at Beijing Institute of Technology, Beijing, China.

Hujun Bao is a professor at Zhejiang University, Hangzhou, China.

Copyright held by authors/owners.
Publication rights licensed to ACM.