Editorial

Special Issue on "Augmented Reality, Virtual Reality & Semantic 3D Reconstruction"

Zhihan Lv [1,*], Jing-Yan Wang [2], Neeraj Kumar [3,4] and Jaime Lloret [5]

1. School of Data Science and Software Engineering, Qingdao University, Qingdao 266000, China
2. PEGASUS FZ LLC, Abu Dhabi 51133, United Arab Emirates; jywang.ieee@gmail.com
3. Computer Science and Engineering, Thapar Institute of Engineering and Technology, Deemed University, Patiala 147004, Punjab, India; neeraj.kumar@thapar.edu
4. School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248001, Uttarakhand, India
5. Department of Communications, Polytechnic University of Valencia, 46022 Valencia, Spain; jlloret@dcom.upv.es
* Correspondence: lvzhihan@gmail.com

1. Introduction

Augmented Reality is a key technology that will facilitate a major paradigm shift in the way users interact with data and has only just recently been recognized as a viable solution for solving many critical needs. Enter augmented reality (AR) technology, which can be used to visualize data from hundreds of sensors simultaneously, overlaying relevant and actionable information over your environment through a headset. Semantic 3D reconstruction makes AR technology much more promising, with much more semantic information. Although, there are several methods currently available as post-processing approaches to extract semantic information from the reconstructed 3D models, the obtained results are uncertainty, and are evenly incorrect. Thus, it is necessary to explore or develop a novel 3D reconstruction approach to automatic recover 3D geometry models and obtained semantic information in simultaneous.

The rapid advent of deep learning brought new opportunities to the field of semantic 3D reconstruction from photo collections. Deep learning-based methods are not only able to extract semantic information but can also be used to enhance some fundamental techniques in semantic 3D reconstruction: those fundamental techniques include feature matching or tracking, stereo matching, camera pose estimation, and multiview stereo. Moreover, deep learning techniques can be used to extract priors from photo collections, the obtained information in turn can improve the quality of 3D reconstruction.

The aim of this Special Issue is to provide a platform for researchers to share innovative work in the field of semantic 3D reconstruction, virtual reality, and augmented reality, including deep learning-based approaches to 3D reconstruction, and software platforms of deep learning for virtual reality and augmented reality.

2. Augmented Reality, Virtual Reality and Semantic 3D Reconstruction

As highly immersive virtual reality (VR) content, 360° video allows users to observe all viewpoints within the desired direction from the position where the video is recorded. In 360° video content, virtual objects are inserted into recorded real scenes to provide a higher sense of immersion. Lee et al. [1] propose a new method for previsualization and 3D composition that overcomes the limitations of existing methods. This system achieves real-time position tracking of the attached camera using a ZED camera and a stereovision sensor, and real-time stabilization using a Kalman filter. The proposed system shows high time efficiency and accurate 3D composition.

Dynamic hand gesture recognition based on one-shot learning requires full assimilation of the motion features from a few annotated data. However, how to effectively extract

the spatio-temporal features of the hand gestures remains a challenging issue. Ma et al. [2] propose a skeleton-based dynamic hand gesture recognition using an enhanced network (GREN) based on one-shot learning by improving the memory-augmented neural network, which can rapidly assimilate the motion features of dynamic hand gestures. Besides, the network effectively combines and stores the shared features between dissimilar classes, which lowers the prediction error caused by unnecessary hyperparameters updating, and improves the recognition accuracy with the increase of categories. The experimental results demonstrate that the GREN network is feasible for skeleton-based dynamic hand gesture recognition based on one-shot learning.

Human cognitive processes in wayfinding may differ depending on the time taken to accept visual information in environments. Kim [3] investigated users' wayfinding processes using eye-tracking experiments, simulating a complex cultural space to analyze human visual movements in perception and the cognitive processes through visual perception responses. The results show that the methods for analyzing the gaze data may vary in terms of processing, analysis, and scope of the data depending on the purpose of the virtual reality experiments. Further, they demonstrate the importance of what purpose statements are given to the subject during the experiment and the possibility of a technical approach being used for the interpretation of spatial information.

Ref. [4] report concerns a study of the impact of a semi-immersive VR system in a group of 25 children in a kindergarten context. The children were involved in several different games and activity types. Their reactions and behaviors were recorded through observation grids addressing task comprehension, participation and enjoyment, interaction and cooperation, conflict, strategic behaviors, and adult-directed questions concerning the activity, the device or general help requests. The grids were compiled at the initial, intermediate and final timepoint during each session. The results show that the activities are easy to understand, enjoyable, and stimulate strategic behaviors, interaction and cooperation, while they do not elicit the need for many explanations. These results are discussed within a neuroconstructivist educational framework and the suitability of semi-immersive, virtual-reality-based activities for cognitive empowerment and rehabilitation purposes is discussed.

As a classical method widely used in 3D reconstruction tasks, the multisource Photometric Stereo can obtain more accurate 3D reconstruction results compared with the basic Photometric Stereo, but its complex calibration and solution process reduces the efficiency of this algorithm. Wang et al. [5] propose a multisource Photometric Stereo 3D reconstruction method based on the fully convolutional network (FCN). The experimental results show that their method has a good effect on solving the main problems faced by the classical method.

The Diagnosis of Attention Deficit/Hyperactivity Disorder (ADHD) requires an exhaustive and objective assessment in order to design an intervention that is adapted to the peculiarities of the patients. The authors of [6] aimed to determine if the most commonly used ADHD observation scale—the Evaluation of Attention Deficit and Hyperactivity (EDAH) scale—is able to predict performance in a Continuous Performance Test based on Virtual Reality (VR-CPT). The findings may partially explain why the impulsive–hyperactive and the combined presentations of ADHD might be considered as unique and qualitatively different subcategories of ADHD. These results also highlighted the importance of measuring not only the observable behaviors of ADHD individuals, but also the scores in performance tests that are attained by the patients themselves.

Image matching techniques offer valuable opportunities for the construction industry. Sabzevar et al. [7] developed and evaluated an orientation and positioning approach that decreased the variation in camera viewpoints and image transformation on construction sites. The results show that images captured while using this approach had less image transformation in contrast to images not captured using this approach.

Super-resolution reconstruction is an increasingly important area in computer vision. To alleviate the problems that super-resolution reconstruction models based on generative

adversarial networks are difficult to train and contain artifacts in reconstruction results, Jiang and Li [8] presented a TSRGAN model which was based on generative adversarial networks. The author redefined the generator network and discriminator network. The experimental results show that the method made the average Peak Signal to Noise Ratio of reconstructed images reach 27.99 dB and the average Structural Similarity Index reach 0.778 without losing too much speed, which was superior to other comparison algorithms in objective evaluation index. What is more, TSRGAN significantly improved subjective visual evaluations. Experimental results prove the effectiveness and superiority of TSRGAN algorithm.

As virtual reality (VR) and the corresponding 3D documentation and modelling technologies evolve into increasingly powerful and established tools for numerous applications in architecture, monument preservation, conservation/restoration and the presentation of cultural heritage, new methods for creating information-rich interactive 3D environments are increasingly in demand. In [9], the authors describe the development of an immersive virtual reality application for the Imperial Cathedral in Königslutter. A specialized technical workflow was developed to build the virtual environment in Unreal Engine 4 (UE4) and integrate the panorama photographs. A simple mechanic was developed using the native UE4 node-based programming language to switch between these two modes of visualization.

Semantic modeling is a challenging task that has received widespread attention in recent years. With the help of mini Unmanned Aerial Vehicles (UAVs), multiview high-resolution aerial images of large-scale scenes can be conveniently collected. In [10], Wei et al. propose a semantic Multi-View Stereo (MVS) method to reconstruct 3D semantic models from 2D images. The graph-based semantic fusion procedure and refinement based on local and global information can suppress and reduce the reprojection error. In the work by Zha et al. [11] a group of images captured from an eye-in-hand vision system carried on a robotic manipulator are segmented by deep learning and geometric features and create a semantic 3D reconstruction using a map stitching method. The results demonstrate that the quality of segmented images and the precision of semantic 3D reconstruction are effectively improved by their method.

Consumer depth cameras bring about cheap and fast acquisition of 3D models. However, the precision and resolution of these consumer depth cameras cannot satisfy the requirements of some 3D face applications. Zhang et al. [12] present a super-resolution method for reconstructing a high resolution 3D face model from a low resolution 3D face model acquired from a consumer depth camera. They evaluated the method both qualitatively and quantitatively, and the experimental results validate their method.

Personalized production is moving the progress of industrial automation forward, and demanding new tools for improving the decision-making of the operators. In [13], the author presents a new, projection-based augmented reality system for assisting operators during electronic component assembly processes. The paper describes both the hardware and software solutions, and depicts the results obtained during a usability test with the new system.

Lip reading recognition is a new technology in the field of human–computer interaction. It is particularly important in a noisy environment and within the hearing-impaired population. This information is a visual language that benefits from Augmented Reality (AR). Wen and Lu [14] implemented the mobile end lip-reading recognition system based on Raspberry Pi for the first time, and the recognition application has reached the latest level of their research. Proved by experimental results, their model has fewer parameters and lower complexity. The accuracy of the model in the test dataset is 86.5%.

Augmented reality (AR) has evolved hand in hand with advances in technology, and today is considered as an emerging technique in its own right. The aim of the study in [15] was to analyze students' perceptions of how useful AR is in the school environment. During the study, a teaching proposal using AR related to the content of some curricular areas was put forward in the framework of the 3P learning model. The participants' perceptions of

this technique were analyzed according to each variable, both overall and by gender, via a questionnaire. The initial results indicate that this technique is, according to the students, useful for teaching the curriculum. The conclusion is that AR can increase students' motivation and enthusiasm while enhancing teaching and learning at the same time.

Recently, associations between the release of scents to the visual content of the scenario has been studied. Alraddadi [16] proposed an approach that combines audio and visual contents to automatically trigger scents through an olfactory device using deep learning techniques. The proposed approach can be applied to different virtual environments as long as scents can be associated with visual and auditory content.

Pham et al. [17] develop a construction hazard investigation system leveraging object anatomization on an Interactive Augmented Photoreality platform (iAPR). A prototype is developed and evaluated objectively through interactive system trials with educators, construction professionals, and learners. The findings demonstrate that the iAPR platform has significant pedagogic methods to improve learners' construction hazard investigation knowledge and skills, which improve safety performance.

Feature tracking in image collections significantly affects the efficiency and accuracy of Structure from Motion (SFM). Insufficient correspondences may cause errors. In [18], the author presents a Superpixel-based feature tracking method for structure from motion. The experimental results show that the proposed method achieves better performance with respect to the state of the art methods.

The present study in [19] focuses on determining the performance and scientific production of augmented reality in higher education (ARHE). A total of 552 scientific publications on the Web of Science (WoS) have been analyzed. The results show that scientific productions on ARHE are not abundant; the main limitation of the study is that the results only reveal the status of this issue in the WoS database.

Author Contributions: Z.L., J.-Y.W., N.K. and J.L. all participated in the review of the papers published in this special issue to the same extent, and jointly supervised the quality of the papers. Z.L. was responsible for writing the relevant editorial content. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J.; Ha, S.; Gentet, P.; Hwang, L.; Kwon, S.; Lee, S. A Novel Real-Time Virtual 3D Object Composition Method for 360° Video. *Appl. Sci.* **2020**, *10*, 8679. [CrossRef]
2. Ma, C.; Zhang, S.; Wang, A.; Qi, Y.; Chen, G. Skeleton-Based Dynamic Hand Gesture Recognition Using an Enhanced Network with One-Shot Learning. *Appl. Sci.* **2020**, *10*, 3680. [CrossRef]
3. Kim, J.Y.; Kim, M.J. Exploring Visual Perceptions of Spatial Information for Wayfinding in Virtual Reality Environments. *Appl. Sci.* **2020**, *10*, 3461. [CrossRef]
4. Lorusso, M.L.; Travellini, S.; Giorgetti, M.; Negrini, P.; Reni, G.; Biffi, E. Semi-Immersive Virtual Reality as a Tool to Improve Cognitive and Social Abilities in Preschool Children. *Appl. Sci.* **2020**, *10*, 2948. [CrossRef]
5. Wang, R.; Wang, X.; He, D.; Wang, L.; Xu, K. FCN-Based 3D Reconstruction with Multi-Source Photometric Stereo. *Appl. Sci.* **2020**, *10*, 2914. [CrossRef]
6. Areces, D.; Rodríguez, C.; García, T.; Cueli, M. Is an ADHD Observation-Scale Based on DSM Criteria Able to Predict Performance in a Virtual Reality Continuous Performance Test? *Appl. Sci.* **2020**, *10*, 2409. [CrossRef]
7. Foroughi Sabzevar, M.; Gheisari, M.; Lo, J. Development and Assessment of a Sensor-Based Orientation and Positioning Approach for Decreasing Variation in Camera Viewpoints and Image Transformations at Construction Sites. *Appl. Sci.* **2020**, *10*, 2305. [CrossRef]
8. Jiang, Y.; Li, J. Generative Adversarial Network for Image Super-Resolution Combining Texture Loss. *Appl. Sci.* **2020**, *10*, 1729. [CrossRef]
9. Walmsley, A.P.; Kersten, T.P. The Imperial Cathedral in Königslutter (Germany) as an Immersive Experience in Virtual Reality with Integrated 360° Panoramic Photography. *Appl. Sci.* **2020**, *10*, 1517. [CrossRef]

10. Wei, Z.; Wang, Y.; Yi, H.; Chen, Y.; Wang, G. Semantic 3D Reconstruction with Learning MVS and 2D Segmentation of Aerial Images. *Appl. Sci.* **2020**, *10*, 1275. [CrossRef]
11. Zha, F.; Fu, Y.; Wang, P.; Guo, W.; Li, M.; Wang, X.; Cai, H. Semantic 3D Reconstruction for Robotic Manipulators with an Eye-In-Hand Vision System. *Appl. Sci.* **2020**, *10*, 1183. [CrossRef]
12. Zhang, F.; Zhao, J.; Wang, L.; Duan, F. 3D Face Model Super-Resolution Based on Radial Curve Estimation. *Appl. Sci.* **2020**, *10*, 1047. [CrossRef]
13. Ojer, M.; Alvarez, H.; Serrano, I.; Saiz, F.A.; Barandiaran, I.; Aguinaga, D.; Querejeta, L.; Alejandro, D. Projection-Based Augmented Reality Assistance for Manual Electronic Component Assembly Processes. *Appl. Sci.* **2020**, *10*, 796. [CrossRef]
14. Wen, J.; Lu, Y. Automatic Lip Reading System Based on a Fusion Lightweight Neural Network with Raspberry Pi. *Appl. Sci.* **2019**, *9*, 5432. [CrossRef]
15. López-García, A.; Miralles-Martínez, P.; Maquilón, J. Design, Application and Effectiveness of an Innovative Augmented Reality Teaching Proposal through 3P Model. *Appl. Sci.* **2019**, *9*, 5426. [CrossRef]
16. Alraddadi, S.; Alqurashi, F.; Tsaramirsis, G.; Al Luhaybi, A.; Buhari, M.S. Aroma Release of Olfactory Displays Based on Audio-Visual Content. *Appl. Sci.* **2019**, *9*, 4866. [CrossRef]
17. Pham, H.C.; Dao, N.-N.; Cho, S.; Nguyen, P.T.; Pham-Hang, A.-T. Construction Hazard Investigation Leveraging Object Anatomization on an Augmented Photoreality Platform. *Appl. Sci.* **2019**, *9*, 4477. [CrossRef]
18. Cao, M.; Jia, W.; Lv, Z.; Zheng, L.; Liu, X. Superpixel-Based Feature Tracking for Structure from Motion. *Appl. Sci.* **2019**, *9*, 2961. [CrossRef]
19. López Belmonte, J.; Moreno-Guerrero, A.-J.; López Núñez, J.A.; Pozo Sánchez, S. Analysis of the Productive, Structural, and Dynamic Development of Augmented Reality in Higher Education Research on the Web of Science. *Appl. Sci.* **2019**, *9*, 5306. [CrossRef]