







Article

A Novel Real-Time Virtual 3D Object Composition Method for 360° Video

Jaehyun Lee ¹, Sungjae Ha ², Philippe Gentet ², Leehwan Hwang ¹, Soonchul Kwon ³
and Seunghyun Lee ^{4,*}

¹ Department of Plasma Bio-Display, Kwangwoon University, Seoul 01897, Korea; noa6142@kw.ac.kr (J.L.); hlh2143@kw.ac.kr (L.H.)

² Spatial Computing Convergence Center, Kwangwoon University, Seoul 01897, Korea; sungjae@kw.ac.kr (S.H.); pgentet@kw.ac.kr (P.G.)

³ Graduate School of Smart Convergence, Kwangwoon University, Seoul 01897, Korea; ksc0226@kw.ac.kr

⁴ Ingenium College, Kwangwoon University, Seoul 01897, Korea

* Correspondence: shlee@kw.ac.kr; Tel.: +82-940-5290

Received: 26 October 2020; Accepted: 30 November 2020; Published: 4 December 2020



Abstract: 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. These techniques are called 3D composition. For a realistic 3D composition in a 360° video, it is important to obtain the internal (focal length) and external (position and rotation) parameters from a 360° camera. Traditional methods estimate the trajectory of a camera by extracting the feature point from the recorded video. However, incorrect results may occur owing to stitching errors from a 360° camera attached to several high-resolution cameras for the stitching process, and a large amount of time is spent on feature tracking owing to the high-resolution of the video. We propose a new method for pre-visualization 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 stereo-vision sensor, and real-time stabilization using a Kalman filter. The proposed system shows high time efficiency and accurate 3D composition.

Keywords: virtual reality; 3D composition; pre-visualization; stereo vision; 360° video

1. Introduction

Three-hundred-and-sixty-degree video is receiving attention as highly immersive virtual reality (VR) content, where users can observe all viewpoints in their desired direction from the fixed position where the video is recorded, through the intentions of the videographer (who dictates environment position and height). Such video has been used to create highly realistic virtual environments not only in the media industry, including the capture of live performances, movies, and broadcasting, but also in education and games. It can provide a higher sense of immersion to users through the insertion of a computer-graphics-based virtual object, and subsequent user interaction with this inserted virtual object. These techniques have become essential elements for VR content. Typical examples include synthesizing virtual characters or objects in VR movies or displaying information markers in a 3D virtual space. This technique of inserting virtual objects into 360° video is called 3D composition.

In general, 360° video is viewed by wearing a head-mounted display (HMD). Many people experience physical discomfort and symptoms such as headaches, disorientation, and nausea when they wear an HMD [1]. This is VR motion sickness. One of the reasons this occurs is due to the user receiving insufficient updates regarding sensory information from the vestibular system [2].

When 360° video content includes fast camera movement, visual information keeps changing but the user's actual body position is fixed, which causes motion sickness. For this reason, most 360° video clips are taken from a fixed position. Synthesizing a virtual object into a fixed 360° video clip does not require a long processing time. The clip can be inserted at the desired position from the center of the camera. There have recently been various types of VR content used in film, education, and tourism which include stable movements filmed using special drones or cars. In the case of a 360° video clip including camera motion, a process of synchronizing the motion of the Red-Green-Blue (RGB) camera (actual camera) and a virtual camera is applied for the 3D composition. This process works by extracting internal (focal length) and external (position and rotation) parameters from the RGB camera used to capture a real scene [3,4]. From these parameters, we can retrieve the motion of the RGB camera, and this is called camera tracking [5]. The traditional 3D composition method estimates the trajectory of the camera by analyzing the feature points of each frame from the captured images. This method has a disadvantage in that the video resolution and camera-tracking processing times are proportional, and the composition results can only be confirmed after several processes (e.g., recording and camera tracking).

In this paper, we propose a novel method using stereo vision that can extract a depth map in real-time for 3D composition, rather than the traditional method using captured images.

2. Background Theory and Related Studies

2.1. 3D Composition

For a realistic 3D composition, it is mandatory that the RGB camera in the real space and the virtual camera in the virtual space have the same viewpoint. In the traditional method, the internal and external parameters can be estimated by searching the feature points from bright spots and dark spots and analyzing the feature point correspondence between each frame. Typical examples of this include simultaneous localization and mapping (SLAM) [6–8] and structure-from-motion (SfM) [9,10]. The external parameters extracted by these algorithms can be linked with virtual cameras in various 3D programs such as 3D Max and Maya, as applied in video production, and the Unity 3D and Unreal engines for game production. Figure 1 shows a traditional 3D composition method.

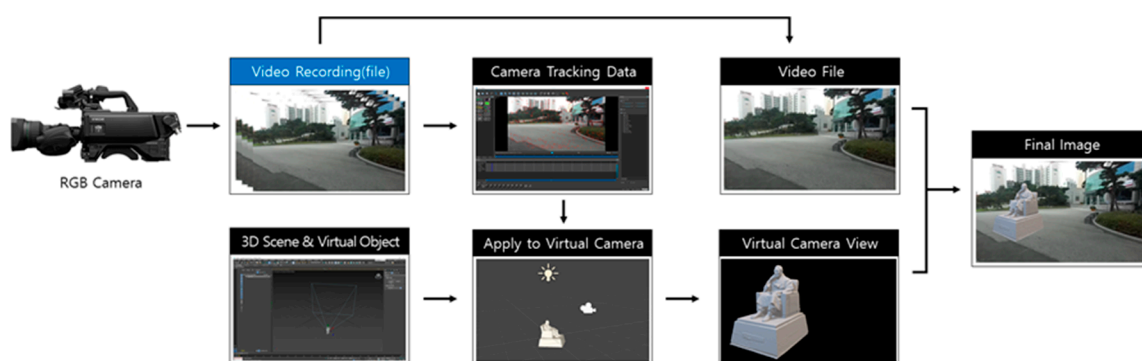


Figure 1. The traditional process using camera-tracking software (Boujou, After Effects) for creating a 3D composition by extracting feature points and estimating camera trajectory from video frames. The blue box shows the recording step (production) and the black boxes show the post-recording steps (post-production).

In general, the 3D composition method tends to depend on the camera-tracking result. Therefore, if the camera-tracking process fails the video must be reshot, which wastes time and money. In previous studies, we reported that various factors may lead to the failure of camera tracking, including an occlusion by a person or object, and motion blur caused by fast camera movement [11]. However, this is more likely to occur in a 2D video shot with relatively numerous camera movements. For a 360° video clip there is a low possibility of camera tracking failures from such factors because stable camera

movements are applied to prevent user motion sickness when wearing an HMD. Nevertheless, there is a factor that has not yet been mentioned, caused by a difference in the production processes between 2D and 360° video. In 360° video more than two cameras are used for capturing each different camera view, and after recording in real-time a 360° panoramic view is created through a matching process called “stitching”, which overlaps parts from each video clip [12]. During this stitching process errors can occur as a result of inaccurate matching due to lens distortion. These errors interfere with the tracking of the feature points in a 360° video clip containing camera movement. As a result, accurate 3D composition is hindered, and human resources are wasted. Figure 2 shows such stitching errors.



Figure 2. Errors of stitching in a 360° video.

There have been various studies undertaken with the aim of solving this problem. Most of them use a method of applying camera tracking to perspective views of a 360° video clip before the stitching process. One such method proposed by Michiels et al. uses a perspective view from one of the 360° camera rigs to obtain an undistorted image for eliminating the stitching errors [13]. In addition, Huang et al. proposed a method for obtaining stable tracking results, which uses an image correction by overlapping the point where the distortion occurs with the position difference between frames [14]. Furthermore, tracking algorithms for spherical images such as spherical scale invariant feature transform (SSIFT) [15] and spherical oriented fast and rotated brief (SPHORB) have been developed [16]. These methods can reduce the stitching errors caused by a misplaced feature point, but basically, it is progressed from the recorded video. In addition, most 360° video clips have a high resolution of more than 4K, which means a significant amount of time is consumed in camera tracking.

2.2. Stereo Vision

Representative algorithms for estimating the location of a device in real space and generating a map of the surrounding environment are simultaneous localization and mapping (SLAM) [4–6] and visual inertial odometry (VIO) [17,18]. SLAM and VIO can be applied to different types of sensors such as stereo vision, time-of-flight (ToF), and lidar, depending on the environment. Among them, stereo vision uses two cameras to extract the depth map and calculate the three-dimensional position of the feature point to calculate the relative motion. It has the advantage of being relatively inexpensive when compared with lidar and it can measure a wider distance than ToF [19].

In this paper, we used a ZED, which was developed by Stereo Lab [20]. A ZED is a stereo-vision device which uses the SLAM algorithm to provide various software tools, a software development kit (SDK) to generate 3D environment mapping and point clouds from real scenes for estimating position tracking in real-time. Various studies have been conducted on the accuracy of ZED. Ibragimov et al. investigated various Robot Operating System (ROS)-based visual SLAM methods and analyzed their feasibility for a mobile robot application in a homogeneous indoor environment. It was verified that the odometry errors of the ZED are as low as those of lidar [21]. In addition, Alapetite et al. compared the ZED with OptiTrack to analyze its accuracy [22].

In this study, we used the real-time positional tracking value of the ZED as the external parameter value of a mounted 360° camera. In addition, we converted the extracted data into a script suitable for a 3D program (e.g., 3D Max, Maya, Unity) to create a virtual camera.

2.3. Related Studies

There have been various studies relating to the 3D composition of virtual objects in 360° video clips. These studies are based on VR, augmented reality, and mixed reality (MR). Focusing on research on 360° video, Rhee et al. implemented a real-time lighting and material expression of virtual objects, according to the positional change reconstructing the camera trajectory from the captured 360° video [23]. Furthermore, the proposed MR360 is used to synthesize virtual objects with real background images. However, it is based on a fixed 360° video, and thus it differs from our proposed method, which contains camera movement [24].

Similarly, Tarko et al. implemented real-time 3D composition using the Unity game engine through a stabilization process after camera tracking [25]. However, camera tracking was based on the captured image. Here, real-time indicates a real-time composition in a 3D program after the tracking process, not during the recording step. Our proposed method is a real-time composition performed at the same time as the video recording.

We recently proposed a novel system that uses Microsoft HoloLens to track positions precisely for match-moving techniques [11] and studied a virtual camera for making motion-graphics using transformed data from the ZED [26]. In this paper, we propose a stabilized 3D composition system and a pre-visualization system using the ZED based on these previous studies.

3. Proposed System and Experiment

In this paper, we propose a novel system that uses ZED stereo vision to track the trajectory precisely for 3D composition in a 360° video. The proposed system also includes a pre-visualization system that can be confirmed to result from a 3D composition while recording the 360° video. Figure 3 shows the complete workflow of the proposed system.

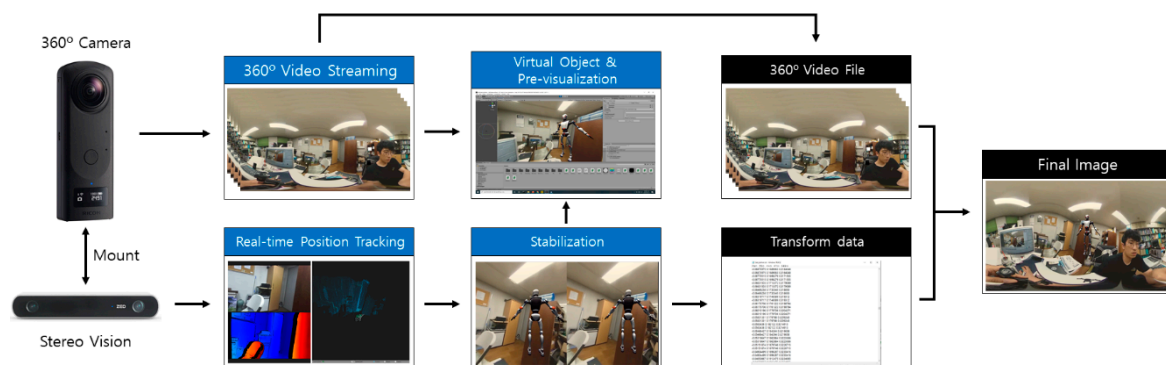


Figure 3. The proposed system workflow using stereo vision for extracting the external parameters of the 360° cameras, which were mounted together. The blue boxes show the recording step (production) and the black boxes show the post-recording steps (post-production).

3.1. Real-Time 3D Composition Using Stereo Vision

In our proposed system, we use a 360° camera “Z1”, developed by Ricoh-theta [27]. Z1 can record in 4K (3840 × 2160). It can also use real-time video streaming with stitching to 3D programs such as Unity and Unreal. This 360° camera system and the ZED mounted on a rig are connected to a PC through a USB 3.0 port. In addition, the ZED is configured such that it faces the same direction as the front of the 360° camera. The 360° camera is used to record the images of the real scene, and at the same time, the ZED extracts the external parameter by generating a depth map in real-time. The ZED generates the initial value of position data (0,0,0) when the program starts, so the difference in the

physical distance between the ZED and Z1 is not considered. Figure 4 shows the rig-mounted 360° camera and the ZED.

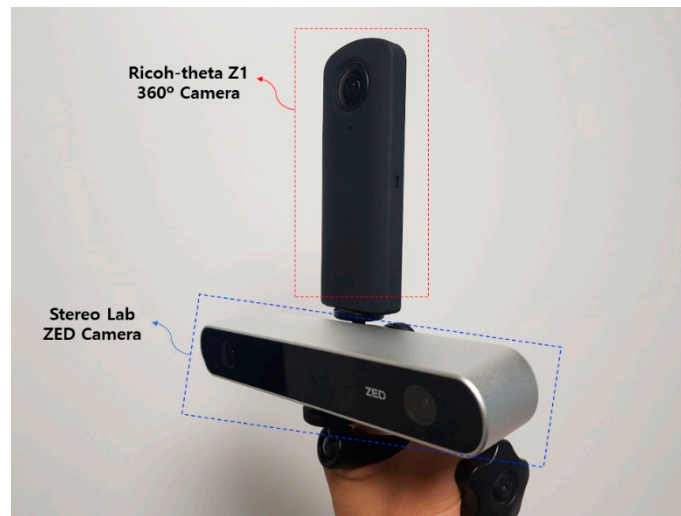


Figure 4. 360° camera and stereo-vision ZED camera.

The extraction and saving of external stereo-vision parameters are applied within Unity 3D, which is used for simultaneous processing with a pre-visualization system to confirm the composition result. For our method, we propose a stabilization process for external parameters in order to obtain better performance from the noise that generally contains stereo vision. The external parameters extracted from the ZED are saved as new data through a linear Kalman filter in real-time.

The Kalman filter is an algorithm that was developed by Kalman during the early 1960s [28,29]. It is used to track the optimal value by removing the noise included in the measured value using the prior and prediction states. It consists of a prediction step and an update step. In the prediction step, an expected value is calculated when the input value is received, according to the prior estimated value. In the update step, an accurate value is calculated based on the prior predicted value and the actual measured value. In other words, a correct value is derived by repeatedly applying the prediction and update steps. It is suitable for real-time processing because it makes predictions based on the immediately preceding data, rather than all previous data [30–32].

The trajectory data stabilized through the Kalman filter can be saved in various formats for application to 3D programs during post-production. In this paper, we saved the data using the 3ds Max file scripting language (.ms) to create a virtual camera in 3ds Max. Figure 5a shows the 3ds Max script file and Figure 5b shows the 3D composition in the 3ds Max program.

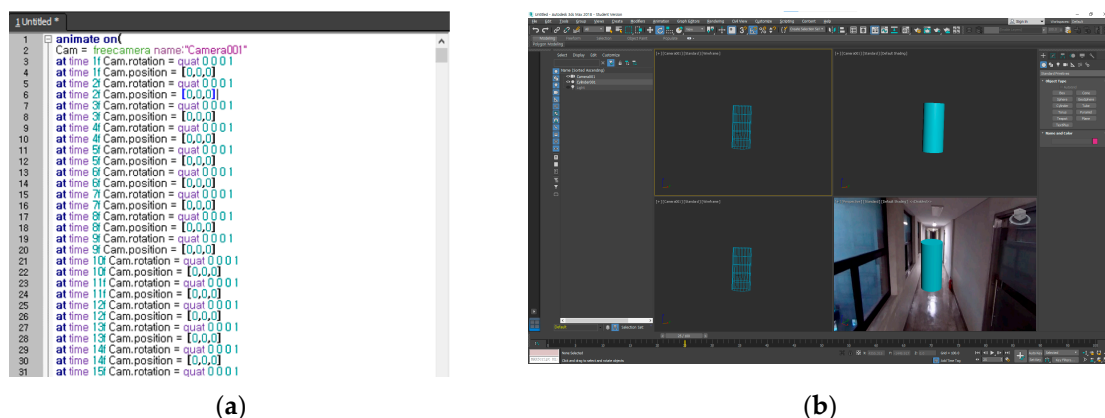


Figure 5. 3D composition process in 3ds Max: (a) Max script and (b) 3ds Max scene.

To measure the accuracy of the camera trajectory with a Kalman filter, the traditional tracking method using an RGB camera was set to the ground truth, in order to compare the applied Kalman filter and raw data of the camera trajectory. The use of the traditional tracking method as a ground truth—even if it is not the best—allows us to show that the proposed method has the same camera trajectory accuracy as the traditional method.

3.2. Pre-Visualization

The purpose of the pre-visualization system is to confirm the composition result while recording the 360° video. For this purpose, we connect the 360° camera and stereo-vision ZED to a PC through a USB 3.0 port to send a video signal and trajectory data within the 3D program. In this study, we used the Unity game engine, which synchronizes the external parameters using the virtual camera from the ZED and generates a 360° virtual space for streaming the 360° camera video feed of the texture of a spherical object in real-time. The spherical object is set to 2.5 m in radius so as not to interfere with the placement of the virtual object. It also follows the virtual camera. It streams the video feed at 4K resolution at 60 fps, with a delay of 0.212 s. If the frame rate and time code do not match, the 3D composition will fail. To avoid this, the update function in Unity is set to 60 updates per second using FixedUpdate which has a static update rate, and a 0.212 s delay is given to the ZED data to match the time code.

The pre-visualization system uses simple 3D objects such as a box, cylinder, and a human-shaped figure. The real-time lighting and texture composition mentioned in various studies can be applied to our proposed method, although the purpose of our system is to confirm the possibility of such composition, and not perfect its application. Therefore, our system does not consider real-time lighting and texture composition techniques. Figure 6 shows the pre-visualization system and a simple 3D object.

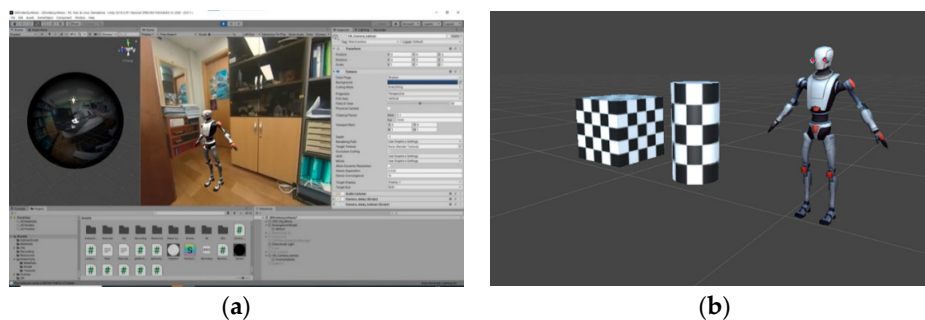


Figure 6. Pre-visualization system: (a) pre-visualization Unity scene and (b) simple 3D object.

4. Experimental Results

In our proposed system, in order to measure the camera trajectory and verify the composition of the pre-visualization system, we recorded two different 360° video clips, indoors and outdoors. The scenes were captured for duration of 26 s and 19 s at rate of 60 fps. Figure 7 shows the 360° images recorded.



Figure 7. Experimental space for recording of 360° video: (a) indoor and (b) outdoor.

4.1. Camera Trajectory

The camera trajectory experiment was undertaken to show the efficiency of the proposed system through comparison with the traditional method of extracting camera trajectory, and additionally to show the improved accuracy of camera trajectory using the Kalman filter. Therefore, the proposed system and an RGB camera were used simultaneously for extracting each camera trajectory. The camera trajectory of the traditional method was set as the ground truth. For various camera movements, we used only hands without special equipment such as a stabilizer. Figure 8a,c shows the camera trajectory extracted from the ZED in comparison with the ground truth, which was recorded using the RGB camera. Figure 8b,d shows the camera trajectory extracted from the ZED with a Kalman filter in comparison with the ground truth. The deviations in percentage error calculated for both raw trajectory data and trajectory data with a Kalman filter, in comparison with the ground truth, are shown in Table 1. From Figure 8 and Table 1, it can be seen that the camera trajectory extracted from the ZED with a Kalman filter is mostly aligned with the ground truth, with a percentage error of less than 3.1%. In addition, the raw camera trajectory data extracted from the ZED is also mostly aligned with the ground truth. However, position X indoors shows a percentage error of 11.8%. By contrast, the Kalman filter shows a percentage error of 2.6%, which is less than that of the raw data.

As a result, it can be seen that the data extracted from the ground truth using the traditional method and the stereo-vision approach do not show a significant difference. This indicates that the proposed method achieved significant results for real-time composition. However, as can be seen in Table 1, the trajectory data following application of the Kalman filter show a lower difference from the ground truth when compared to the traditional method for all data. This indicates that applying the Kalman filter is more effective in preventing noise in the stereo-vision sensor and obtaining stable data.

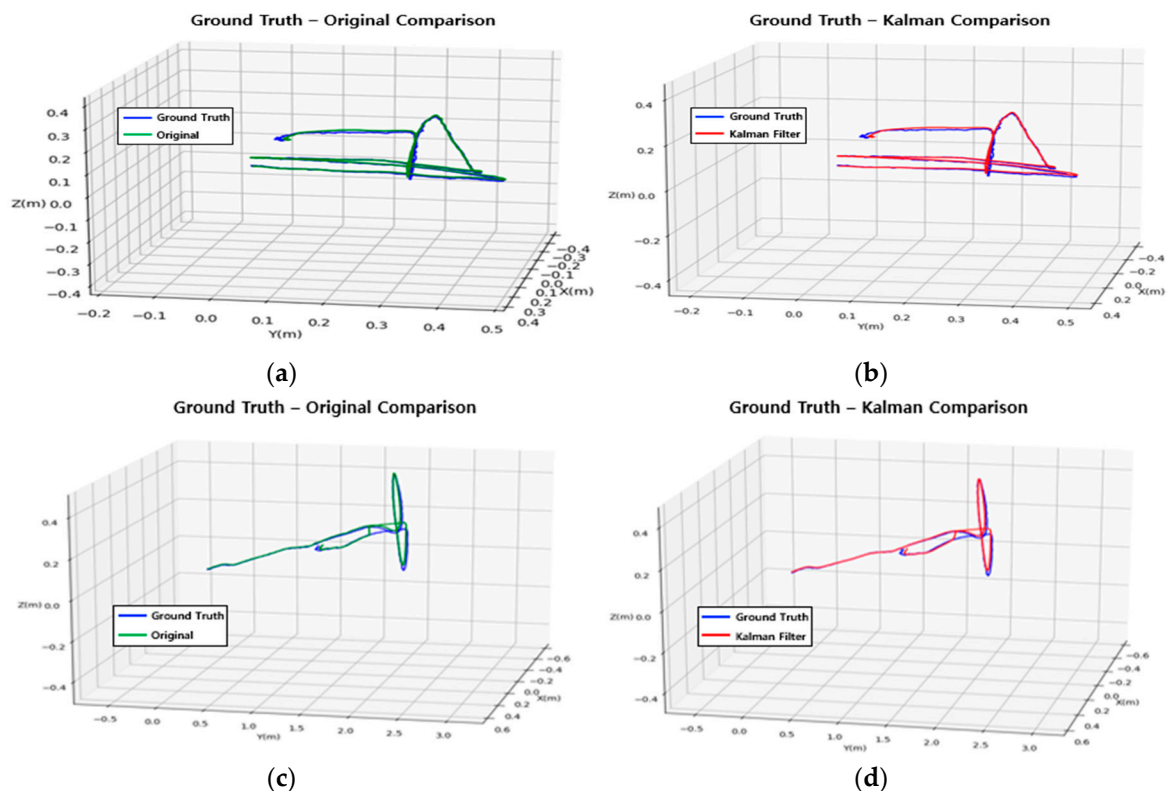


Figure 8. Accuracy evaluation of the camera trajectory extracted from (a,c) the ZED and (b,d) the applied Kalman filter against the ground truth (Boujou).

Table 1. Standard deviation in the comparison of the ground truth, raw trajectory data, and trajectory with the Kalman filter.

		Position X (%)	Position Y (%)	Position Z (%)
Indoor	Ground truth–raw trajectory	11.81081	1.875021	0.547256
	Ground truth–trajectory with Kalman filter	2.6780439	0.432794	0.112748
Outdoor	Ground truth–raw trajectory	1.740435	3.604414	0.147483
	Ground truth–trajectory with Kalman filter	1.537084	3.093616	0.077608

4.2. 3D Composition Using Pre-Visualization System

At the same time as the recording, a 360° video clip and the external parameters of the stereo vision were transmitted to the Unity 3D game engine to create a virtual camera for pre-visualization. Figure 9 shows the results of the pre-visualization of the indoor and outdoor scenes while recording the 360° video. The result displayed through the pre-visualization system was used to confirm the composition result. For the final video, further composition processes such as lighting, shadowing, and texturing in 3D software are needed.

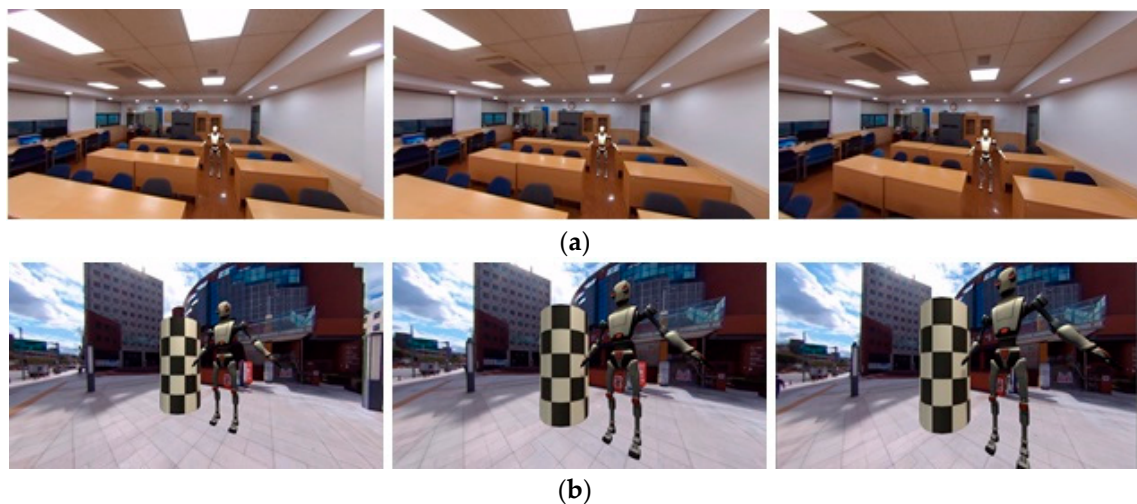


Figure 9. Results of pre-visualization: (a) indoor and (b) outdoor.

The final composition was conducted in 3ds Max 2018. When the recording was finished, the 3ds Max script, which included the trajectory information of the stereo vision, was immediately generated. It was used to create a virtual camera in the 3ds Max virtual space. Figure 10 shows the rendered images and the final 3D composition images. No difference can be seen in the camera trajectory because it uses the same trajectory data saved from a real-time pre-visualization system. As a result, it does not need an extra process for extracting the camera-tracking data, and thus our proposed system is more time efficient than the traditional method.

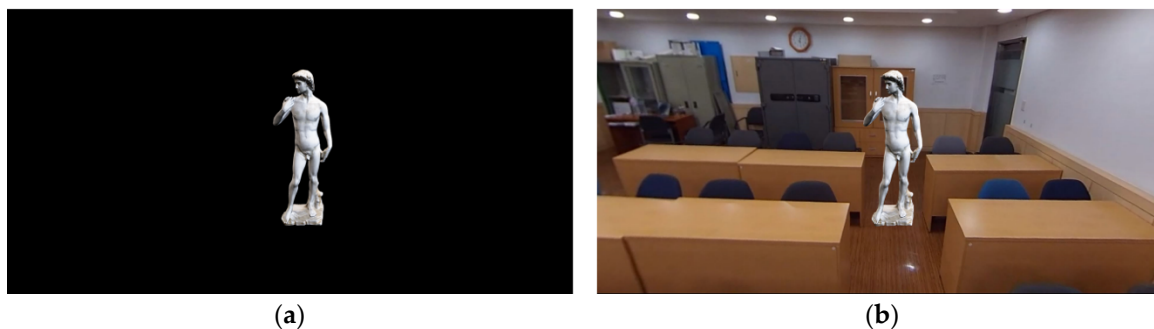


Figure 10. Cont.

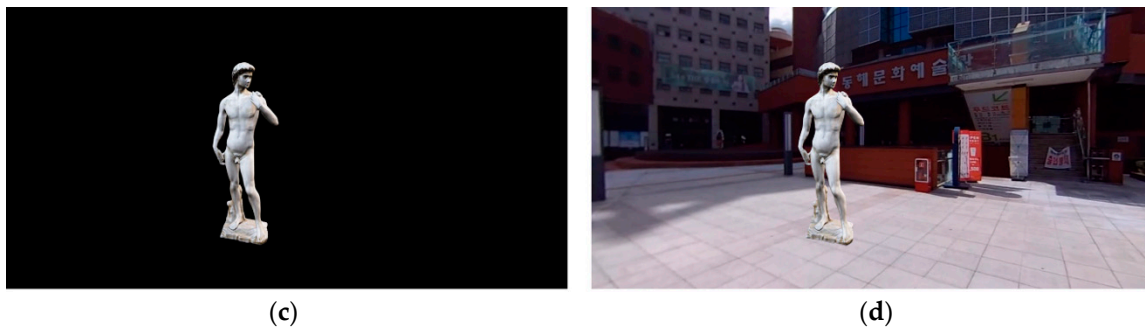


Figure 10. Rendered images and final 3D compositing images: (a,b) indoor and (c,d) outdoor.

5. Conclusions

In this paper we proposed a real-time 3D composition method for 360° video production. The proposed system consists of two subsystems. Firstly, a stereo-vision ZED is used to obtain the parameters of the external cameras, which are mounted together to estimate the camera trajectory in real-time. Secondly, an efficient pre-visualization system is implemented to preview the results of the 3D composition during the recording.

In this study, we exploited a system that overcomes the limitations of the traditional method, which uses camera tracking after video recording. Our experimental results show that the 3D composition results of the proposed system are not significantly different than the results obtained using the traditional method. In addition, we implemented a stable trajectory by applying a Kalman filter to the raw data obtained from the ZED. The Kalman filter achieved better trajectory results than the raw data. Our system has an advantage over the traditional method because it does not need to extract feature points from the captured images. It can save the data of the external parameters during the recording process, and this was also verified in the composition results. However, as a limitation of the proposed system, it works using a USB port and not a network. In the future, the authors plan to implement a network communication system by installing a network device that will be able to send video and transformed data to a PC for further processing.

It can be predicted that, with the advancement of the virtual reality industry, interest in the 3D composition of 360° video will also increase, and therefore a more efficient system will be required. We expect that the system presented herein will be applicable for the effective 360° video production of 3D composition in low-budget production companies.

Author Contributions: Conceptualization, J.L.; methodology, J.L., P.G.; software, J.L.; validation, J.L., L.H., and S.K.; formal analysis, J.L. and S.H.; investigation, J.L.; resources, S.K., S.H. and S.L.; data curation, J.L., P.G. and L.H.; writing—original draft preparation, J.L. and L.H.; writing—review and editing, S.K., S.H. and S.L.; visualization, J.L., L.H. and P.G.; supervision, S.K. and S.H.; project administration, S.L., P.G.; funding acquisition, S.L., S.H., and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2020-2015-0-00448) & (IITP-2020-01846) supervised by the Institute of Information & Communications, Technology, Planning, & Evaluation (IITP). This research was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-00922, Development of holographic stereogram printing technology based on multi-view imaging).

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Munafo, J.; Diedrick, M. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Exp. Brain Res.* **2017**, *235*, 889–901. [CrossRef] [PubMed]
2. Jung, S.; Whangbo, T. Study on inspecting VR motion sickness inducing factors. In Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Kuta, Bali, 8–10 August 2017.
3. Erica, H. *The Art and Technique of Matchmoving: Solutions for the VFX Artist*, 1st ed.; Elsevier: New York, NY, USA, 2010; pp. 1–14, ISBN 9780080961132.
4. Dobbert, T. The Matchmoving Process. In *Matchmoving: The Invisible Art of Camera Tracking*, 1st ed.; Sybex: San Francisco, CA, USA, 2005; pp. 5–10, ISBN 0782144039.
5. Pollefeij, M.; Van, G.L.; Vergauwen, M.; Verbiest, F.; Cornelis, K.; Tops, J. Visual modeling with a hand-held camera. *Int. J. Comput. Vis.* **2004**, *59*, 207–232. [CrossRef]
6. Davison, A.J. Real-Time Simultaneous Localisation and Mapping with a Single Camera. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003.
7. Klein, G.; Murray, D. Parallel Tracking and Mapping for Small AR Workspaces. In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 13–16 November 2007; pp. 225–234.
8. Davison, A.J. SLAM++: Simultaneous Localisation In addition, Mapping at the Level of Objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1352–1359.
9. Bao, S.Y.; Savarese, S. Semantic Structure from Motion. In Proceedings of the IEEE CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011.
10. Hafeez, J.; Jeon, H.J.; Hamacher, A.; Kwon, S.C.; Lee, S.H. The effect of patterns on image-based modelling of texture-less objects. *Metrol. Meas. Syst.* **2018**, *25*, 755–767.
11. Lee, J.; Hafeez, J.; Kim, K.; Lee, S.; Kwon, S. A novel real-time match-moving method with HoloLens. *Appl. Sci.* **2019**, *9*, 2889. [CrossRef]
12. Huang, K.-C.; Chien, P.-Y.; Chien, C.-A.; Chang, H.-C.; Guo, J.-I. A 360-degree panoramic video system design. In *Proceedings of the Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2014; pp. 1–4.
13. Michiels, N.; Jorissen, L.; Put, J.; Bekaert, P. Interactive Augmented Omnidirectional Video with Realistic Lighting. *Public Key Cryptogr. PKC 2018* **2014**, *8853*, 247–263. [CrossRef]
14. Huang, J.; Chen, Z.; Ceylan, D.; Jin, H. 6-DOF VR videos with a single 360-camera. In Proceedings of the 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 18–22 March 2017.
15. Cruz-Mota, J.; Bogdanova, I.; Paquier, B.; Bierlaire, M.; Thiran, J. Scale invariant feature transform on the sphere: Theory and applications. *Int. J. Comput. Vis.* **2012**, *98*, 217–241. [CrossRef]
16. Zhao, Q.; Feng, W.; Wan, L.; Zhang, J. SPHORB: A fast and robust binary feature on the sphere. *Int. J. Comput. Vis.* **2015**, *113*, 143–159. [CrossRef]
17. Leutenegger, S.; Lynen, S.; Bosse, M.; Siegwart, R.; Furgale, P. Keyframe-based visual-inertial odometry using nonlinear optimization. *Int. J. Robot. Res.* **2015**, *34*, 314–334. [CrossRef]
18. Sun, K.; Mohta, K.; Pfrommer, B.; Watterson, M.; Liu, S.; Mulgaonkar, Y.; Taylor, C.J.; Kumar, V. Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Rob. Autom. Lett.* **2018**, *3*, 965–972. [CrossRef]
19. Vit, A.; Shani, G. Comparing RGB-D sensors for close range outdoor agricultural phenotyping. *Sensors* **2018**, *18*, 4413. [CrossRef] [PubMed]
20. ZED. Available online: <https://www.stereolabs.com/zed/> (accessed on 10 October 2020).
21. Ibragimov, I.Z.; Afanasyev, I.M. Comparison of ROS-based Visual SLAM Methods in Homogeneous Indoor Environment. In Proceedings of the 2017 14th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, 25–26 October 2017; pp. 1–6.
22. Alapetite, A.; Wang, Z.; Hansen, J.P.; Zajaczkowski, M.; Patalan, M. Comparison of three off-the-shelf visual odometry systems. *Robotics* **2020**, *9*, 56. [CrossRef]
23. Iorns, T.; Rhee, T. Real-Time Image Based Lighting for 360-Degree Panoramic Video. In *Image and Video Technology—PSIVT 2015 Workshops*; Huang, F., Sugimoto, A., Eds.; PSIVT 2015, Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9555.

24. Rhee, T.; Petikam, L.; Allen, B.; Chalmers, A. MR360: Mixed reality rendering for 360° panoramic videos. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1379–1388. [[CrossRef](#)] [[PubMed](#)]
25. Iorns, T.; Rhee, T.H. Real-Time Image Based Lighting for 360-Degree Panoramic Video. In Proceedings of the PSIVT Workshops, Auckland, New Zealand, 23–27 November 2015; pp. 139–151.
26. Kim, L.H.; Lee, J.H.; Kim, K.J.; Lee, S.H. A study on motion graphics virtual camera using real-time position tracking in post-production. *J. Mov. Image Technol. Assoc. Korea* **2019**, *1*, 133–149.
27. Z1. Available online: <https://theta360.com/en/about/theta/z1.html> (accessed on 10 October 2020).
28. Kalman, R.E. A new approach to linear filtering and prediction problem. *J. Basic Eng.* **1960**, *82*, 34–45. [[CrossRef](#)]
29. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; Lecture; University North Carolina: Chapel Hill, NC, USA, 2001.
30. Prabhu, U.; Seshadri, K.; Savvides, M. Automatic Facial Landmark Tracking in Video Sequences using Kalman Filter Assisted Active Shape Models. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010.
31. Chen, S.Y. Kalman filter for robot vision: A survey. *IEEE Trans. Ind. Electron.* **2012**, *59*, 4409–4420. [[CrossRef](#)]
32. Smeulders, A.W.M.; Chu, D.M.; Cucchiara, R.; Calderara, R.; Dehghan, A.; Shah, M. Visual tracking: An experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1442–1468. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).