electronics

MDPI

Article

Real–Virtual 3D Scene-Fused Integral Imaging Based on Improved SuperPoint

Wei Wu [1,2,*], Shigang Wang [2,*], Wanzhong Chen [2], Hao Wang [1,*] and Cheng Zhong [3]

[1] College of Computer Science and Engineering, Changchun University of Technology, No. 2055 Yan'an Street, Changchun 130012, China
[2] College of Communication Engineering, Jilin University, No. 5372 Nanhu Road, Changchun 130012, China; chenwz@jlu.edu.cn
[3] College of International Education, Changchun University of Technology, No. 2055 Yan'an Street, Changchun 130012, China; zhongcheng@ccut.edu.cn
* Correspondence: wuwei@ccut.edu.cn (W.W.); wangsg@jlu.edu.cn (S.W.); cc716997@163.com (H.W.)

Abstract: To enrich 3D scenes, a real–virtual fusion-based integral imaging method is proposed. It combines the Softargmax function with Gaussian weighting coefficients for sub-pixel feature point extraction from SuperPoint detection results. SIFT is also used for feature point detection and matching, along with the improved SuperPoint. Subsequently, based on the multi-view 3D reconstruction, the real object is reconstructed into a 3D model. A virtual model is then fused with the 3D reconstructed model of the real object to generate a real–virtual fusion elemental image array based on the display platform's optical parameters. The experimental results demonstrate that the proposed method can optically reconstruct more realistic and vivid real–virtual fusion 3D images. This method can enrich a scene's content, enhance visualization and interactivity, save costs and time, and provide flexibility and customization.

Keywords: integral imaging; 3D reconstruction; real–virtual fusion

Citation: Wu, W.; Wang, S.; Chen, W.; Wang, H.; Zhong, C. Real–Virtual 3D Scene-Fused Integral Imaging Based on Improved SuperPoint. *Electronics* **2024**, *13*, 970. https://doi.org/10.3390/electronics13050970

Academic Editor: Manohar Das

Received: 12 January 2024
Revised: 6 February 2024
Accepted: 21 February 2024
Published: 3 March 2024

1. Introduction

Unlike traditional two-dimensional (2D) imaging, three-dimensional (3D) imaging technology has the potential to capture the depth and angle information of objects. Three-dimensional imaging and display systems have brought revolutionary changes in many fields, providing more comprehensive, accurate, and visualized data, thus promoting innovation and development. Through 3D imaging technology, we can understand and analyze objects and scenes in the real world in a more in-depth and comprehensive way, bringing tremendous potential and opportunities to various industries [1]. Integral imaging is a true 3D display technology characterized by its simple architecture, low manufacturing cost, and good imaging effect. It has been developed for more than a century. Many researchers have dedicated their efforts to improving the 3D display performance in various aspects [2–5]. Lim et al. used depth-slice images and orthogonal-view images to optimize element image arrays and enhance the display resolution [6]. Navarro et al. proposed a smart pseudo-orthogonal transformation method to optimize the depth of field [7]. Kwon et al. improved the resolution by reconstructing and rapidly processing the intermediate view through the image information of adjacent elements [8]. They also enhanced the depth-of-field range by using a multiplexed structure combining a dual-channel beam splitter with a microlens array [9]. Zhang et al. proposed a target contour extraction method and a block-based image fusion method to generate a reconstructed image with extended degrees of freedom and improve the depth of field [10]. Sang et al. utilized a pre-filter function array (PFA) to preprocess elemental images, aiming to improve the reconstruction fidelity of integral imaging [11]. Ma et al. employed a conjugate pinhole camera and a pinhole-based projection model to mitigate distortion and expand the view angle of integral

imaging [12]. Wang et al. utilized a bifocal lens array to achieve a 3D image display with a depth of field twice that of conventional integral imaging [13]. Cao et al. proposed a high-resolution integral imaging display with an enhanced point light source to improve the image resolution of the vertical and horizontal dimensions [14]. The above research improved the quality and efficiency of integral imaging content generation and display by optimizing and improving the resolution, viewing angle, and depth of field range of integral imaging.

With developments in computer science, real–virtual fusion based on integral imaging has recently become a research hotspot field, which overlays 3D images in real environments [15,16]. Hong et al. introduced an integral floating 3D display that adopts a concave lens to combine 3D images and real-world scenes [17]. Hua and Javidi et al. used the microscopic integral imaging to create a mini-3D image and combined it with the freeform optical to design a see-through 3D display [18]. Yamaguchi and Takaki et al. used two integral imaging systems to achieve a background occlusion capability [19]. H. Deng et al. proposed a method for creating a magnified augmented reality (AR) 3D display using integral imaging. The method involves the use of a convex lens and a half-mirror to merge a magnified 3D image with a real-world scene [20]. To increase real-time performance, Wang et al. proposed GPU-based computer-generated integral imaging and presented an AR device for surgical 3D image overlay [21]. Furthermore, developing a high-definition, real–virtual fusion 3D display is also a problem that needs to be resolved [22–24]. However, these methods can only present one real scene at a time. In order to achieve a more immersive and rich augmented reality experience, as well as capture, process, and combine different real scenes simultaneously, this study proposes a method to import real objects into virtual spaces. By utilizing virtual capture technology based on integral imaging, it is possible to generate a real–virtual fusion 3D display that combines multiple real scenes in a single process. Since modeling real objects degrades the object display quality, it is necessary to enhance the 3D reconstruction effect.

The Structure from Motion (SFM) is a classical 3D reconstruction algorithm [10], first proposed by Longuet-Higgins [25], that can accurately reconstruct real objects and generate computerized 3D models. The Scale Invariant Feature Transform (SIFT) has demonstrated excellent application effectiveness and high robustness [26]. It is also a widely recognized feature extraction algorithm. However, it tends to perform poorly on image regions with weak textures. With the rise of deep learning, there has been a rapid advancement and increase in iteration speeds in image feature point detection. The fast corner detection is the first algorithm to obtain image corners through machine learning [27]. Yi et al. proposed the learned invariant feature transform (LIFT), an end-to-end convolutional neural network-based feature point detection [28]. Detone et al. proposed a self-supervised interest point (SuperPoint) algorithm for feature point extraction that is trained in a self-supervised manner [29]. This algorithm has a good detection effect in areas of the image with a weak texture. However, the SuperPoint algorithm can only extract integer coordinates of feature points, which limits its accuracy. Wu added a sub-pixel module based on the Softargmax function to implement the subpixel detection of feature points [30]. However, in dense feature maps, the edges of the pixel blocks extracted by Softargmax often contain isolated points with high pixel values. These isolated points can affect the accuracy of the bias values calculated by Softargmax. Since the calculation of Softargmax relies on the weights of individual pixels within the pixel block, if there are isolated high pixel value points, their weights may be excessively amplified, leading to inaccuracies in the bias values. This inaccuracy further affects the results of subsequent tasks and reduces the overall accuracy of the algorithm. On the other hand, if the size of the extracted pixel blocks is reduced to avoid isolated points, the pixels at the boundaries will not be fully utilized. This is because boundary pixels usually carry important edge information, and reducing the size of the pixel blocks may result in the neglect or loss of boundary pixels, thereby affecting the accuracy of Softargmax calculations.

Therefore, this paper proposes to improve the SuperPoint function by adding Gaussian weights and performing sub-pixelization to enhance the matching accuracy of image feature point detection and improve the mapping effect in real scenes. By reducing the weight of the edge area, the interference caused by isolated points with high pixel values can be minimized. Meanwhile, increasing the weight of the central area can better capture the position of the main feature points, improving the accuracy of feature point detection. In addition, sub-pixelization processing can further improve the positioning accuracy of feature points, thereby improving the quality and authenticity of scene reconstruction.

In conclusion, this paper combines integral imaging and 3D reconstruction techniques and proposes a method for generating elemental image arrays for real–virtual fusion scenes. The 3D information of multiple real scenes can be combined in a single process. The SuperPoint feature point detection is improved with an added Gaussian weight and combined with the SIFT algorithm to obtain a rich number of accurate image feature points. Then, the SFM algorithm is used to complete the 3D model reconstruction. Finally, following the principle of integral imaging, the elemental image array of the real–virtual fusion scene can be generated. This eliminates the need to consider the size and matching issues of the scene. The model only needs to be reconstructed once and can be used multiple times, which saves the cost and time of modeling and rendering. This enables the rapid generation of diverse scenes while minimizing the need for additional equipment and resources. Users can modify the real–virtual fusion model anytime according to their needs, achieving a more personalized scene presentation.

This article is organized as follows: in Section 2, the content generation for real–virtual fusion in integral imaging is introduced. Section 2.1 presents a joint feature point detection method aimed at improving the detection effectiveness. Section 2.2 focuses on how to generate real–virtual fused elemental image arrays. Section 3 presents the experimental results and provides a comprehensive discussion. Finally, in Section 4, conclusions are drawn based on the conducted research.

2. Integral Imaging Content Generation for Real–Virtual Fusion

An improved SuperPoint and the SIFT algorithm are used for feature point detection and matching, followed by the utilization of the SFM-based method to achieve the 3D reconstruction of the object. By combining the 3D model of the real object with the virtual model, a mixed-reality scene is created. This allows the integral imaging virtual capture method to capture both real and virtual models simultaneously. The process is shown in Figure 1.

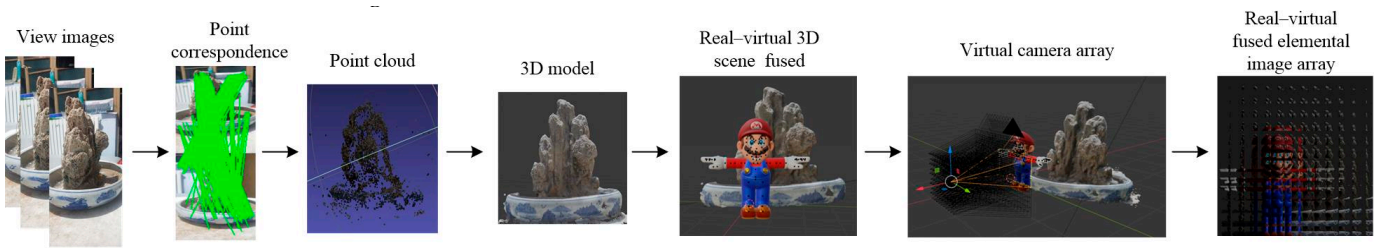


Figure 1. The overall approach of our proposed method.

2.1. Joint Feature Point Detection

The SIFT algorithm is renowned for its ability to detect a large number of feature points accurately. However, it exhibits an uneven distribution of feature points and performs poorly in low-texture regions. On the other hand, the SuperPoint feature detection, based on deep learning networks, excels in detecting feature points in low-texture regions but lacks precision in positioning. By combining the strengths of both algorithms, SIFT and SuperPoint, we can effectively enhance the capability of image feature point detection and matching. This combination allows us to leverage the advantages of each algorithm while mitigating their respective shortcomings. The SIFT algorithm provides a robust and accu-

rate localization of feature points while the SuperPoint algorithm enhances the performance in low-texture regions. As a result, the overall capability for detecting and matching image feature points is significantly improved. This integration of SIFT and SuperPoint brings a complementary approach that addresses the limitations of each algorithm individually, leading to a more reliable and comprehensive feature point detection.

SuperPoint is a deep learning-based algorithm used for self-supervised feature point detection and description. It extracts image features through deep learning and generates pixel-level image feature points along with their corresponding descriptors. The SuperPoint network can be divided into four main parts: the VGG-style shared encoder, the feature point decoder, the descriptor subdecoder, and the error construction, as shown in Figure 2.

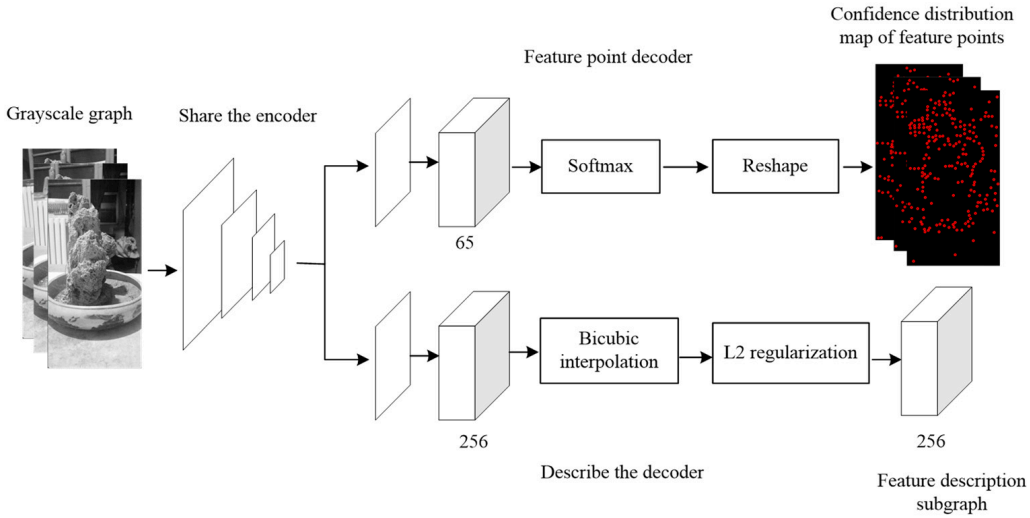


Figure 2. The architecture of the SuperPoint network.

The SuperPoint feature detection network outputs the feature point confidence distribution map and the feature descriptor map. Each pixel value on the feature point confidence distribution map represents the probability that the corresponding point is a feature point in the image. SuperPoint obtains feature points using the Non-Maximum Suppression (NMS) function. It detects a set of pixels above a certain threshold and then uses these pixels as centers to obtain bounding boxes of size N × N. The overlap degree between each bounding box is calculated, and feature points that are too close to each other are removed to obtain the final set of feature points. However, these feature points have integer coordinates, and the detection results are not precise [31]. A study [30] added a sub-pixel module based on the Softargmax function to the SuperPoint feature point detection algorithm to achieve the sub-pixel detection of feature points. Considering that the NMS function only retains the maximum value within an N × N pixel size and sets all other pixel points to 0, it is easy to have isolated points with larger pixel values on the edges of the pixel blocks extracted by Softargmax in densely featured images, which will affect the accuracy of the bias value calculated by Softargmax. On the other hand, if the size of the extracted pixel block is reduced, the pixel points at the boundary cannot be utilized, which will also affect the calculation accuracy of Softargmax.

To solve the above problem, we introduce Gaussian weight coefficients into the SuperPoint sub-pixel feature point detection algorithm. It starts by obtaining the integer coordinate feature point set through the feature point decoder of the SuperPoint algorithm. Then, the pixel blocks extracted by the Softargmax function are processed with the addition of Gaussian weight coefficients, typically with a block size of 5 × 5. This approach effectively reduces the weight of the edge regions within the pixel blocks while increasing the weight of the central regions. As a result, it mitigates the impact of isolated points with large pixel values at the edges on the accuracy of the Softargmax calculation and improves the precision of the feature point detection.

Let G be the Gaussian weight coefficient, the size is consistent with the pixel block extracted by Softargmax, and $G(i,j)$ is the weight value at (i,j); then, the expectation of offset $(\delta x, \delta y)$ in the x and y directions is calculated as follows:

$$\delta x = \frac{\sum_j \sum_i e^{G(i,j)S(x_i,y_j)} i}{\sum_j \sum_i e^{G(i,j)S(x_i,y_j)}}, \delta y = \frac{\sum_j \sum_i e^{G(i,j)S(x_i,y_j)} j}{\sum_j \sum_i e^{G(i,j)S(x_i,y_j)}}, \quad (1)$$

where $S(x,y)$ represents the probability that the pixel at position (x,y) in the image is a feature point, while $S(x_i,y_j)$ represents the pixel value at position i,j in the extracted pixel block.

According to the obtained offset $(\delta x, \delta y)$, the final feature point coordinates are calculated as follows:

$$(x,y) = (x_0 + \delta x, y_0 + \delta y) \quad (2)$$

where x_0 and y_0 are the integer coordinates of feature points, which also represent the center position of the extracted pixel block. The above method increases the algorithm's complexity due to the addition of the Gaussian weight coefficient, but it can effectively reduce the error generated in the process of feature point sub-pixelization in the feature-point-dense image or region, resulting in more accurate final feature points.

The sub-pixel feature descriptors of the improved SuperPoint feature point detection algorithm are obtained through the descriptor decoder. The input to the descriptor decoder is convolved to generate a descriptor matrix of size $H/8 \times W/8 \times 256$. This matrix is then subjected to bicubic interpolation and L2 normalization to obtain a 256-dimensional feature descriptor vector for the sub-pixel feature points.

The feature points extracted by SIFT are high-precision sub-pixel feature points, and those extracted by the improved SuperPoint algorithm are also at the sub-pixel level. This means that the feature points extracted by these two algorithms at the same position will not completely coincide. Therefore, in order to merge the feature point set extracted by SIFT and the improved SuperPoint, it is necessary to calculate the distance between each feature point in the improved SuperPoint and each feature point in the SIFT set. If the distance exceeds a threshold, the feature point is retained; otherwise, it is discarded. The formula for this calculation is as follows:

$$\begin{cases} \sqrt{(x_{1i} - x_{2j})^2 + (y_{1i} - y_{2j})^2} > \lambda, S_{(x_{1i},y_{1i})} = 1 \\ \sqrt{(x_{1i} - x_{2j})^2 + (y_{1i} - y_{2j})^2} < \lambda, S_{(x_{1i},y_{1i})} = 0 \end{cases} \quad (3)$$

where $(x_{1,i}, y_{1,i})$ and $(x_{2,j}, y_{2,j})$ are the feature points in the improved SuperPoint and SIFT feature point sets. The value of $S_{(x,y)}$ indicates whether the feature point (x,y) in the feature point set of the improved SuperPoint is discarded, and λ is a constant.

Considering that the dimensions of the feature point description extracted by the SIFT and improved SuperPoint algorithms are 128 and 256, respectively, this paper first uses feature point matching and then detects distance thresholds to merge feature point collections. The feature point matching algorithm calculates the distance ratio between the closest and the second closest neighboring features and matches the SIFT feature point set and the improved SuperPoint feature point set on two images, respectively.

2.2. Virtual–Real Fusion Element Image Array Generation

The generation of elemental image arrays mainly involves two methods: real-scene capture and computer-based virtual generation. The real-scene capture can acquire the 3D information of the actual scene but requires camera arrays or lens arrays. The calibration and synchronization of camera arrays are complex and precise processes, while lens arrays are limited in size and shooting range due to manufacturing constraints. The computer-based virtual generation method uses 3D modeling and animation software such as Maya,

3DsMax, and Blender to simulate the shooting process of real lens arrays or camera arrays, establishing an optical mapping model to generate elemental image arrays. This method is not limited by capturing devices, allowing for an infinitely small virtual camera spacing and shooting from within virtual objects [32]. It can achieve various shooting structures such as parallel, convergence, and shift. However, the captured 3D models are virtual models. Therefore, addressing the issues of a high cost, complex processes, limited shooting structures, and a single virtual shooting model in real-scene capture, this section proposes a method for generating a real–virtual fusion elemental image array. It utilizes SFM to establish the 3D models of real objects and merge them with the virtual models. The ray tracing technique is used to render the real–virtual fusion elemental image array. This method enables the capture of real scenes without being limited by capturing devices, providing a more diverse source of true 3D videos for integral imaging system research.

Compared to other 3D reconstruction methods, SFM is based on the principles of multi-view geometry to restore the 3D structural information of a scene through images captured from different angles [33]. By using images from different angles, it is possible to calculate the fundamental matrix and essential matrix, which can then be decomposed to extract the relative camera pose information and determine the position and orientation of each camera when capturing the images. Subsequently, using triangulation techniques, the 3D point cloud of the scene can be reconstructed. In the reconstruction process, bundle adjustment is performed to minimize errors considering error accumulation. Bundle adjustment is a method that simultaneously adjusts all camera parameters to optimize camera poses and the positions of 3D points by minimizing reprojection errors. These steps are repeated, incorporating new images, until all multi-view images are reconstructed to obtain the final sparse point cloud. The advantages of SFM are its ability to capture camera motion freely without prior camera calibration and its ability to fuse multi-view information to generate sparse but accurate 3D reconstruction results. It can be used in various scenarios, including indoor, outdoor, static, and dynamic scenes. Therefore, this paper adopts SFM for the 3D reconstruction of real objects. The specific process is shown in Figure 3.

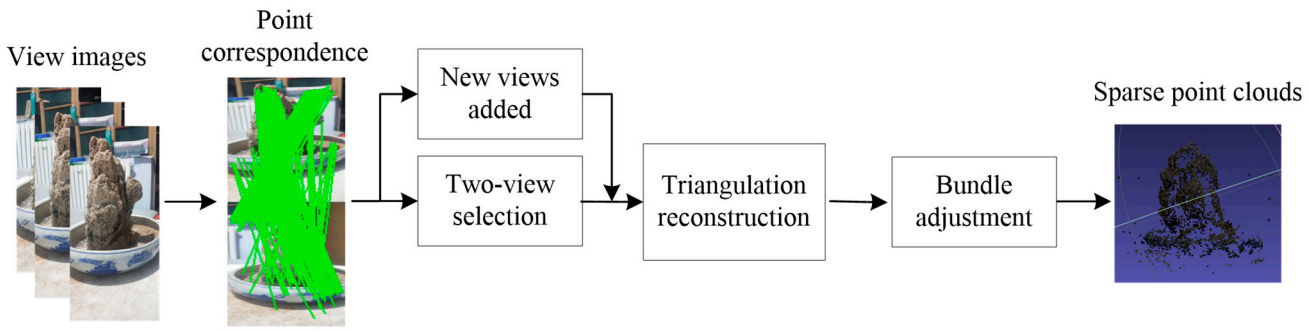


Figure 3. SFM algorithm.

Import the real object model into virtual space and combine it with a virtual model, and a virtual reality fusion scene is created. After that, using integral imaging virtual acquisition, a real–virtual 3D scene-fused elemental image array can be generated. The element image array serves as a bridge connecting the integral imaging acquisition and reconstruction process.

The process of collecting and visualizing spatial scene information using an integral imaging system can be summarized as consisting of two stages: acquisition and display, as shown in Figure 4a [5]. During the acquisition process, media with information recording capabilities are placed in parallel behind the lens array. In this way, when light emitted from an object passes through a lens array and is projected onto a capture device, the image information of the object at a certain spatial angle is saved at the elemental image array. Integral imaging is an optically symmetrical system, where the acquisition and display processes are optically reversible. In the display process, the elemental image array captured

on the recording media is projected onto the lens array through high-definition display devices such as liquid crystal displays (LCDs). According to the reversible principle of light paths, the light emitted from the display screen, after optical decoding by the lens array, converges at the original location of the captured object, generating a real three-dimensional image. Therefore, in the integral imaging system, both the recording and reproducing processes need to satisfy the lens imaging principle—the Gaussian imaging formula.

$$\frac{1}{g} + \frac{1}{L} = \frac{1}{f} \tag{4}$$

where g and L are the distance between the captured or display device and the lenslet array; f is the focal length of the lens array.

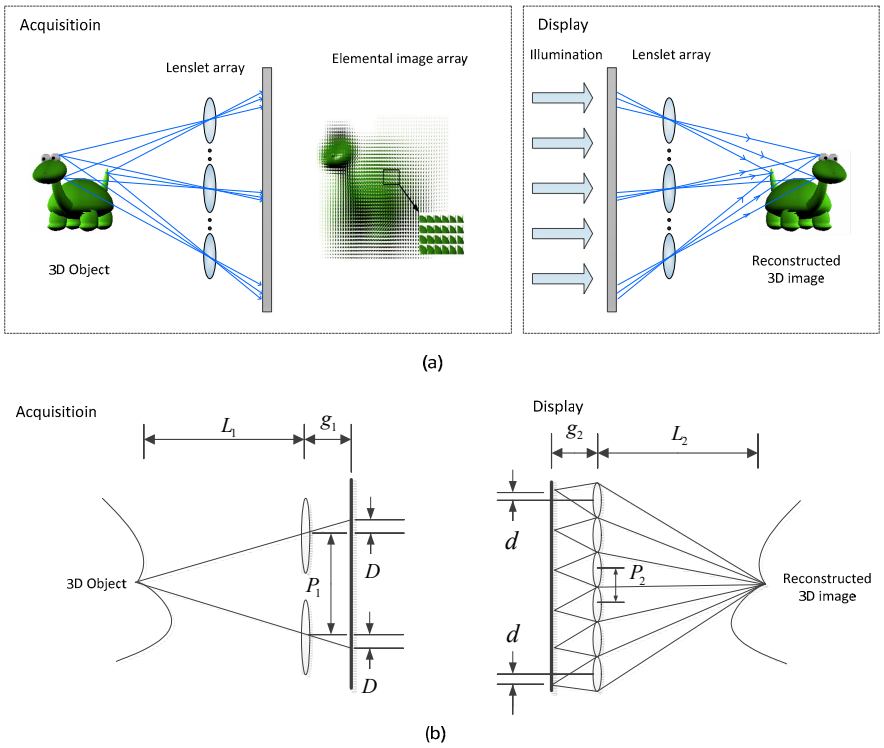


Figure 4. Integral imaging acquisition and display process. (**a**) Fundamental principles; (**b**) optical mapping model.

In order to make the acquired elemental image array efficiently displayed in the actual light field environment, it is important that the camera array parameters in the virtual acquisition process must be set according to the display platform. The optical mapping model is shown in Figure 4b.

$$\begin{aligned} \frac{L_1}{g_1} &= \frac{P_1}{D} \\ \frac{L_2}{g_2} &= \frac{P_2}{d} \\ \frac{D}{d} &= \beta \end{aligned} \tag{5}$$

where L_1, g_1, P_1, and D are the object distance, camera focal length, camera array pitch, and disparity between adjacent image pairs in acquisition; L_2, g_2, P_2, and d are the image distance, distance between the lens and the display plane, lens pitch, and disparity between adjacent elemental image pairs in the display; β is the magnification factor.

After establishing the optical mapping model from the acquisition to display, the virtual camera array is set according to the display parameters, as shown in Figure 5.

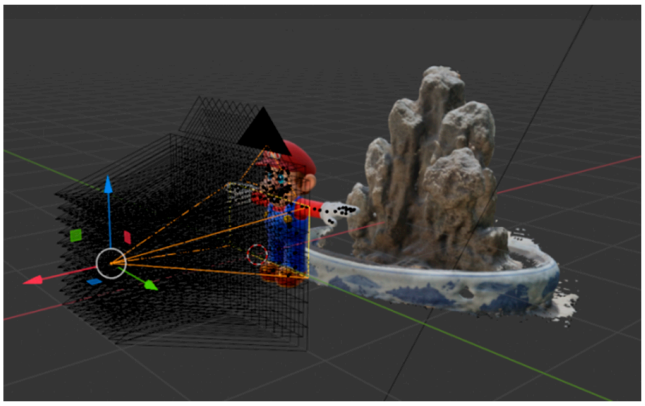


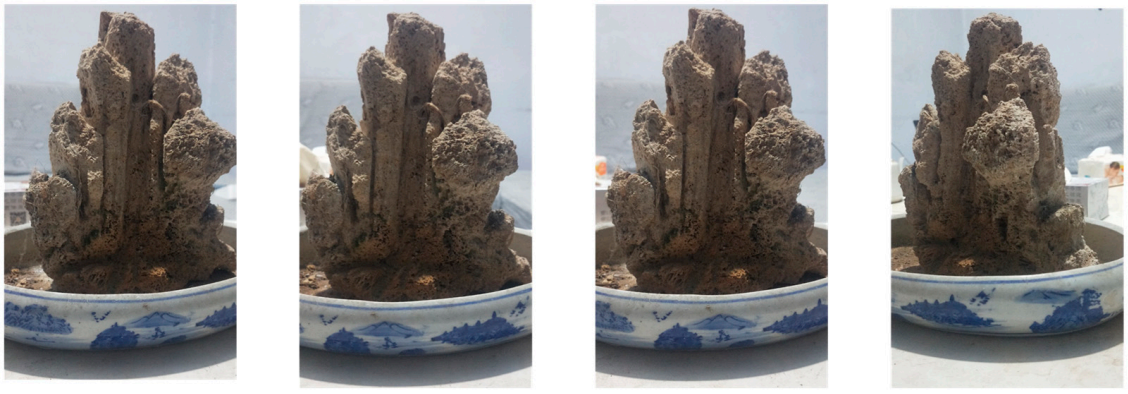
Figure 5. Virtual camera array.

3. Experimental Results

The experiment first verified that the improved SuperPoint feature point detection algorithm can reconstruct the 3D model of the object more accurately and realistically. On this basis, the real–virtual fusion elemental image array combining the real object reconstruction and virtual models and its optical 3D reconstruction effect were given experimentally.

A standard test dataset of large-scale castle-P30 and a self-captured small-scale rockery scene with more textures and details were selected as the real scenes. The image sets consisted of 30 buildings and 23 rockery images captured from different angles, with sizes of 3072 × 2048 and 2306 × 4608 pixels. Figure 6 shows a subset of different view images.

(**a**)



(**b**)

Figure 6. A subset of different view images. (**a**) Castle-P30; (**b**) rockery.

In order to verify the effectiveness of the proposed method, three sparse point cloud reconstructions were carried out according to different feature point detection algorithms. The reconstruction result is shown in Figure 7. Table 1 shows the number of point clouds generated.

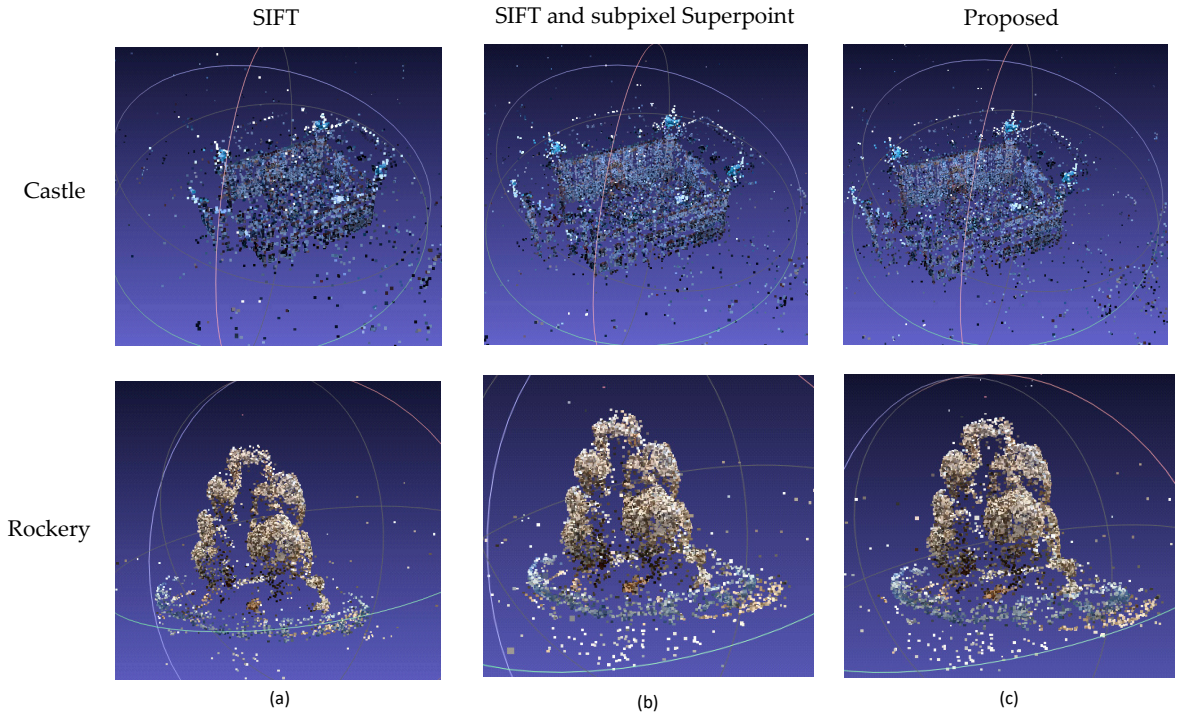


Figure 7. Sparse point cloud reconstruction of the castle-P30 image set and rockery generated by (**a**) SIFT feature point detection, (**b**) SIFT and subpixel SuperPoint detection [30], and (**c**) the proposed method.

Table 1. The total number of points in the sparse point cloud.

	Castle-P30	**Rockery**
SIFT	19,249	13,833
SIFT and subpixel SuperPoint detection	27,207	16,141
Proposed method	27,663	16,185

Figure 7 and Table 1 demonstrate that the proposed method generates approximately one-third more point clouds compared to the single SIFT algorithm, and compared to [30], it also has significant improvements. Furthermore, our proposed method exhibits superior accuracy in feature point detection and a stronger capability to generate 3D sparse point clouds, particularly in areas with weak textures such as the base and center of the rockery. This improvement is attributed to the incorporation of Gaussian weight coefficients to process the pixel blocks extracted by the Softargmax function. By reducing the weight of the edge regions and increasing the weight of the central regions, this approach mitigates the impact of isolated high pixel values at the edges of the pixel blocks on the results of the Softargmax calculation.

For the obtained sparse point cloud, this study performed patch reconstruction using the patch-based multi-view stereo (PMVS). Through the Poisson surface reconstruction operation in Meshlab, a smooth 3D model of the real object was obtained. The surface texture information of the model was recovered using the captured image of the object. As a result, a highly restored, detailed, and vivid 3D model was achieved, as shown in Figure 8.

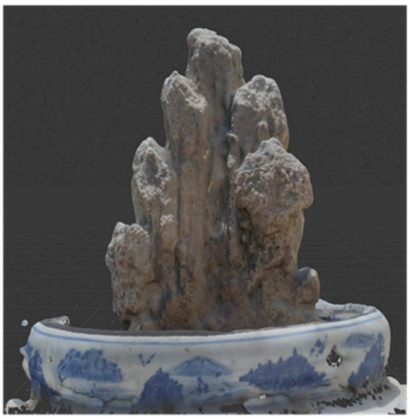


Figure 8. Reconstructed 3D model of the rockery.

The 3D model and texture map of the real object rockery were imported into Blender to form a real–virtual fusion scene with the virtual model Mario, as shown in Figure 9. Then, a virtual camera array was established straight ahead of the scene to generate real–virtual fused elemental image array. The parameters of the virtual camera array and optical lens array are shown in Table 2, and the real–virtual fused elemental image array is shown in Figure 10.

Figure 9. Virtual reality fusion scene.

Table 2. Configuration of the experiment.

Virtual camera array	Distance between 3D object and camera array	500 mm
	Number of cameras	12 × 12
	Resolution per camera	67 × 67
Optical lens array	Number of lenses	12 × 12
	Focal length	60 mm
	Pitch of lens	10 mm

To assess the quality of the displayed real–virtual fused elemental image array, the virtual and optical reconstruction for the left, right, center, top, and bottom views, respectively, were established, as shown in Figures 11 and 12.

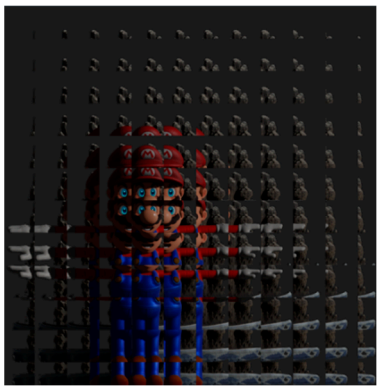


Figure 10. Real–virtual fused element image array.

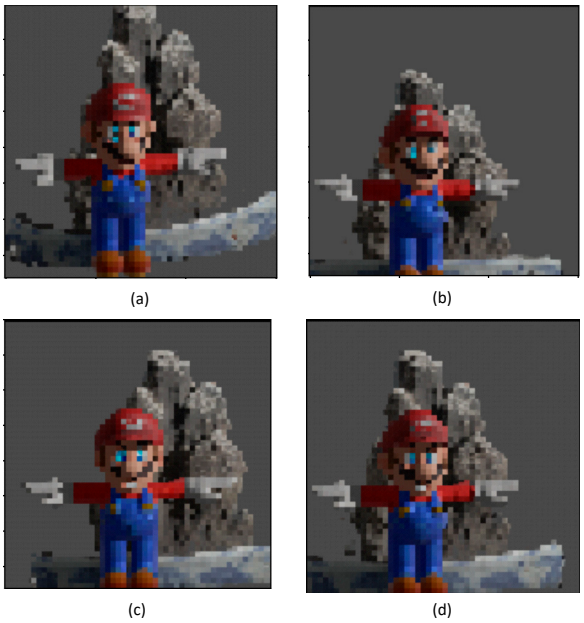


Figure 11. Virtual reconstruction of (**a**) top, (**b**) bottom, (**c**) left, and (**d**) right views.

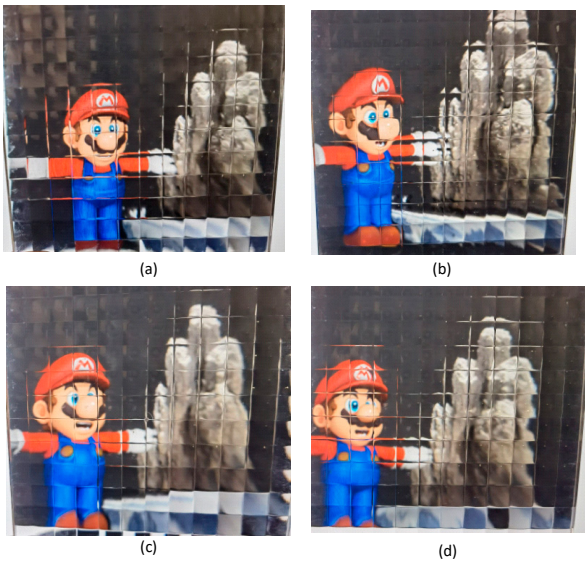


Figure 12. Optical reconstruction of (**a**) top, (**b**) bottom, (**c**) left, and (**d**) right views.

The experimental results show that combining real scenes with virtual models and utilizing integral imaging virtual capture can generate an element image array with both real and virtual contents in a single step, greatly enriching the 3D scene. The introduction of Gaussian weighting coefficients helps to better address the edge region issue in the SuperPoint sub-pixel feature detection algorithm. The edge region often contains isolated points with large pixel values, which can significantly affect the accuracy of computations. By reducing the weight of the edge region, the interference caused by these isolated points on the computational results can be minimized. At the same time, increasing the weight of the central region allows for a better capture of the positions of the main feature points, thus improving the accuracy of feature point detection. Therefore, the proposed method in this paper is able to adapt to feature point distributions in various scenarios, enhancing the quality and authenticity of reconstruction results. As seen in Figures 8 and 12, the detailed features of the rockery are well preserved, enabling a more realistic and vivid optical reconstruction of the 3D images. This means that during the reconstruction process, we are able to capture more detailed information, making the final reconstruction result more realistic, which is crucial for the visualization and interactivity of real scenes.

4. Conclusions

In this paper, the process of generating real–virtual fused 3D scenes based on integral imaging is introduced. The improved SuperPoint algorithm enhances the accuracy of feature point detection by adding Gaussian weighting coefficients to softeargmax to perform sub-pixel feature point detection. Through the reconstruction of sparse, dense point clouds and a Poisson surface, a highly restored, detailed, and vivid 3D model is achieved. Then, integral imaging is used to generate and reproduce the real–virtual fused 3D images by combining them with the virtual model. The experimental results show that more realistic and vivid real–virtual fusion 3D images can be optically reconstructed. This method enhances user immersion and experience by providing more intuitive and interactive visual effects, allowing them to better understand and explore scenes. It also helps save costs and time in modeling and rendering, enabling the rapid generation of diverse scenes while reducing the need for additional equipment and resources. Users can modify virtual models or replace real objects according to their needs, achieving a more personalized scene presentation. Overall, fusing real object modeling with virtual models can provide more realistic, rich, and interactive 3D scenes, offering users a better visual experience and immersion. The proposed method may be widely used in the fields of virtual reality (VR), augmented reality (AR), and training simulations in various industries, allowing learners to gain a more realistic experience closer to actual scenarios. This is particularly important for industries that involve complex operations or high-risk work environments, such as aerospace, medical surgery, hazardous chemical handling, and more. In the future, we will explore the combination of other deep learning methods to further improve the 3D reconstruction effect.

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

Funding: This research was supported by the National Natural Science Foundation of China (NSFC) (61631009 and 61901187) and the Science and Technology Development Plan of Jilin Province (20220101127JC and 20210201027GX).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors would like to acknowledge support from the State Key Laboratory of New Communications Technologies at Jilin University.

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

References

1. Martínez-Corral, M.; Javidi, B. Fundamentals of 3D imaging and displays: A tutorial on integral imaging, light-field, and plenoptic systems. *Adv. Opt. Photonics* **2018**, *10*, 512–566. [CrossRef]
2. Xiao, X.; Javidi, B. Advances in three-dimensional integral imaging sensing, display, and applications. *Appl. Opt.* **2013**, *52*, 546–560. [CrossRef] [PubMed]
3. Hui, R.; Li, X.N. Review on tabletop true 3D display. *J. Soc. Inf. Display* **2020**, *28*, 75–91.
4. Wu, W.; Wang, S.G.; Piao, M.L.; Zhao, Y.; Wei, J. Performance metric and objective evaluation for displayed 3D images generated by different lenslet arrays. *Opt. Commun.* **2018**, *426*, 635–641. [CrossRef]
5. Wu, W.; Wang, S.G.; Zhong, C.; Piao, M.L.; Zhao, Y. Integral Imaging with Full Parallax Based on Mini LED Display Unit. *IEEE Access* **2019**, *7*, 32030–32036. [CrossRef]
6. Lim, Y.T.; Park, J.H.; Kwon, K.C.; Kim, N. Resolution-enhanced integral imaging microscopy that uses lens array shifting. *Opt. Express* **2009**, *17*, 19253–19263. [CrossRef] [PubMed]
7. Navarro, H.; Martínez-Cuenca, R.; Saavedra, G.; Martínez-Corral, M.; Javidi, B. 3D integral imaging display by smart pseudoscopic-to-orthoscopic conversion(SPOC). *Opt. Express* **2010**, *18*, 25573–25583. [CrossRef] [PubMed]
8. Kwon, K.C.; Jeong, J.S.; Erdenebat, M.U.; Piao, Y.L.; Yoo, K.H.; Kim, N. Resolution-enhancement for an orthographic-view image display in an integral imaging microscope system. *Biomed. Opt. Express* **2015**, *6*, 736–746. [CrossRef]
9. Kwon, K.C.; Erdenebat, M.U.; Alam, M.A.; Lim, Y.T.; Kim, K.G.; Kim, N. Integral imaging microscopy with enhanced depth-of-field using a spatial multiplexing. *Opt. Express* **2016**, *24*, 2072–2083. [CrossRef]
10. Zhang, M.; Wei, C.Z.; Piao, Y.R.; Liu, J.Q. Depth-of-field extension in integral imaging using multi-focus elemental images. *Appl. Opt.* **2017**, *56*, 6059–6064. [CrossRef]
11. Zhang, W.L.; Sang, X.Z.; Gao, X.; Yu, X.B.; Yan, B.B.; Yu, C.X. Wavefront aberration correction for integral imaging with the pre-filtering function array. *Opt. Express* **2018**, *26*, 27064–27075. [CrossRef]
12. Ma, S.T.; Lou, Y.M.; Hu, J.M.; Wu, F.M. Enhancing integral imaging performance using time-multiplexed convergent backlight. *Appl. Opt.* **2020**, *59*, 3165–3173. [CrossRef] [PubMed]
13. Wang, L.; Deng, H.; Zhong, F.Y.; Cheng, C.; Li, Q. Integral imaging display with enhanced depth of field based on bifocal lens array. *J. Soc. Inf. Display* **2021**, *29*, 689–696. [CrossRef]
14. Cao, Y.L.; Chagnaadorj, B.O.; Dalkhaa, N.E.; Baasantseren, G. Aberration Compensated Point Light Source Display with High-Resolution. *Front. Phys* **2022**, *3*, 919050. [CrossRef]
15. Shen, X.; Javidi, B. Large depth of focus dynamic micro integral imaging for optical see-through augmented reality display using a focus-tunable lens. *Appl. Opt* **2018**, *57*, B184–B189. [CrossRef]
16. Javidi, B.; Carnicer, A.; Arai, J.; Fujii, T.; Hua, H.; Liao, H.; Martínez-Corral, M.; Pla, F.; Stern, A.; Waller, L.; et al. Roadmap on 3D integral imaging: Sensing, processing, and display. *Opt. Express* **2020**, *28*, 32266–32293. [CrossRef]
17. Hong, J.; Min, S.W.; Lee, B. Integral floating display systems for augmented reality. *Appl. Opt.* **2012**, *51*, 4201–4209. [CrossRef]
18. Hua, H.; Javidi, B. A 3D integral imaging optical see-through head-mounted display. *Opt. Express* **2014**, *22*, 13484–13491. [CrossRef]
19. Yamaguchi, Y.; Takaki, Y. See-through integral imaging display with background occlusion capability. *Appl. Opt.* **2016**, *55*, A144–A149. [CrossRef]
20. Deng, H.; Wang, Q.H.; Xiong, Z.L.; Zhang, H.L.; Xing, Y. Magnified augmented reality 3D display based on integral imaging. *Optik* **2016**, *127*, 4250–4253. [CrossRef]
21. Wang, J.; Suenaga, H.S.; Liao, H.; Hoshi, K.; Yang, L.; Kobayashi, E.; Sakuma, I. Real-time computer-generated integral imaging and 3D image calibration for augmented reality surgical navigation. *Comput. Med. Imaging Graph.* **2015**, *40*, 147–159. [CrossRef]
22. Li, Q.; He, W.; Deng, H.; Zhong, F.Y.; Chen, Y. High-performance reflection-type augmented reality 3D display using a reflective polarizer. *Opt. Express* **2021**, *29*, 9446–9453. [CrossRef] [PubMed]
23. Deng, H.; Chen, C.; He, M.Y.; Li, J.J.; Zhang, H.L.; Wang, Q.H. High-resolution augmented reality 3D display with use of a lenticular lens array holographic optical element. *J. Opt. Soc. Am. A* **2019**, *36*, 588–593. [CrossRef] [PubMed]
24. Huang, H.K.; Hua, H. High-performance integral-imaging-based light field augmented reality display using freeform optics. *Opt. Express* **2018**, *26*, 17578–17590. [CrossRef] [PubMed]
25. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections. *Rds. Comp. Vis.* **1987**, *293*, 61–62. [CrossRef]
26. Lowe, D.G. Object recognition from local scale-invariant features. *IEEE Int. Conf. Comput. Vis.* **1999**, *2*, 1150–1157.
27. Rosten, E.; Drummond, T. Matchine learning for high-speed corner detection. In Proceedings of the Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 430–443.
28. Yi, K.M.; Trulls, E.; Lepetit, V.; Fua, P. LIFT: Learned invariant feature transform. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 467–483.

29. Detone, D.; Malisiewic, T.; Rabinovich, A. SuperPoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 337–349.
30. Wu, Y.X. Research on Key Technologies of 3D Reconstruction Based on Visible Light Multi-View Images. Master Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2022.
31. Meza, J.; Romero, L.A.; Marrugo, A.G. MarkerPose: Robust Real-time Planar Target Tracking for Accurate Stereo Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 1282–1290.
32. Alpaslan, Z.Y.; Sawchuk, A.A. Multiple camera image acquisition models for multi-view 3D display interaction. In Proceedings of the IEEE 6th Workshop on Multimedia Signal Processing, Siena, Italy, 29 September–1 October 2004; pp. 256–262.
33. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.