Accurate Depth Recovery Method Based on the Fusion of Time-of-Flight and Dot-Coded Structured Light

Feifei Gu [1,2], Huazhao Cao [1], Pengju Xie [1] and Zhan Song [1,2,3,*]

1 Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; ff.gu@siat.ac.cn (F.G.); huazhao.cao@siat.ac.cn (H.C.); pj.xie@siat.ac.cn (P.X.)
2 Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen 518052, China
3 The Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong 999077, China
* Correspondence: zhan.song@siat.ac.cn

Abstract: 3D vision technology has been gradually applied to intelligent terminals ever since Apple Inc. introduced structured light on iPhoneX. At present, time-of-flight (TOF) and laser speckle-based structured light (SL) are two mainstream technologies applied to intelligent terminals, both of which are widely regarded as efficient dynamic technologies, but with low accuracy. This paper explores a new approach to achieve accurate depth recovery by fusing TOF and our previous work—dot-coded SL (DCSL). TOF can obtain high-density depth information, but its results may be deformed due to multi-path interference (MPI) and reflectivity-related deviations. In contrast, DCSL can provide high-accuracy and noise-clean results, yet only a limited number of encoded points can be reconstructed. This inspired our idea to fuse them to obtain better results. In this method, the sparse result provided by DCSL can work as accurate "anchor points" to keep the correctness of the target scene's structure, meanwhile, the dense result from TOF can guarantee full-range measurement. Experimental results show that by fusion, the MPI errors of TOF can be eliminated effectively. Dense and accurate results can be obtained successfully, which has great potential for application in the 3D vision task of intelligent terminals in the future.

Keywords: depth recovery; time-of-flight; structured light; data fusion

1. Introduction

3D vision technology has been recognized as one of the most important technologies for mobile intelligent terminals to enter the new era of intelligent human-computer interaction [1–3]. Up to now, 3D vision technology has been successfully applied to mobile terminals in many fields such as 3D live detection, virtual reality, augmented reality, and so on. The two mainstream technologies applied at mobile terminals are time-of-flight (TOF) and laser speckle-based structured light (SL), which are widely known as having low accuracy [4–6]. Therefore, they cannot be used to obtain accurate 3D information, which significantly limits the application scope of 3D vision at intelligent terminals.

The major error sources of TOF come from multipath interference (MPI) and reflectivity-related distance variations, which have been the major bottleneck restricting the application of TOF in the field of precise measurement. The elimination of these major error sources has always been a research hot spot in the field of TOF [7–13]. At present, most studies attempt to solve MPI problems from the TOF's projection model itself, such as the multiple frequencies approach [8], the sparse deconvolution approach [9], the iterative computational approach [10,11], the direct and global separation approach [12], and so on. However, most of these approaches have tried to eliminate MPI from the receiving waveform of TOF; therefore, their effect was rather limited. In recent years, deep learning methods have been introduced to solve MPI [13]. Though these can remove the MPI and TOF data noise to

some extent, their effect on real data may be limited because the training dataset can only be collected from simulated scenes.

As for the structured light (SL) technique, it can be generally divided into the temporally coded SL [14–17] and spatially coded SL (SCSL) [18–20], depending on different codification strategies. Compared with the former, the latter can work with a single shot, which makes it a more efficient and applicable tool for 3D vision task on mobile intelligent terminals. Among the SCSL, laser speckle-based SL is currently one of the most popular 3D dynamic imaging solution for mobile terminals. For example, the Intel Realsense series depth cameras, Apple's iPhone X smartphone, vivo's Find X smartphone, and Orbbec's Astra series depth cameras are all based on this technique. However, in laser speckle-based SL, the stochastic speckles are used for stereo correlation matching to provide depth information. The measurement accuracy is rather low due to block matching. Moreover, the statistically random pattern has the shortcomings of low spatial resolution and high sensitivity to noise. To improve its measurement accuracy, we proposed a dot-coded SL (DCDL) in our previous work [20]. By adding coding information into the distribution of random dots, feature points can be located in sub-pixel precision and hence high-accuracy measurement can be achieved. However, limited by its coding strategies, of which the codeword at every image pixel is determined by its neighboring pattern elements, DCDL can only conduct sparse reconstruction.

From above analysis, it can be seen that TOF and DCSL have complementary characteristics. DCSL can provide high-accuracy and noise-clean point clouds, but it is plagued by coding density. In contrast, TOF can provide complete and dense point clouds, yet perplexed by MPI and reflectivity-related error sources. Therefore, to achieve accurate depth recovery, this paper explored fusing DCSL with TOF. Their fusion can make up for each other's shortcomings and lead to accurate and dense depth recovery. The sparse points provided by DCSL can work as accurate "anchor points" to keep the correctness of the target scene's structure, meanwhile, the result from TOF can guarantee dense and full-range measurement. In this way, high-density and high-accuracy depth measurements can be conducted in real time. To our knowledge, the fusion approach of TOF and DCSL has not yet been published or revealed.

This paper is structured as follows. Section 2 exhibits the complementary characteristics of TOF and DCSL by measuring a typical object. Section 3 presents the detailed introductions of the proposed technology and algorithm. Section 4 describes the major experimental results. Section 5 is the discussion. Finally, the conclusion is drawn in Section 6.

2. Characteristic Comparison between TOF and DCSL

As a start, it is necessary to exhibit the complementary characteristics of TOF and DCSL. As mentioned above, the major error sources of TOF come from MPI and reflectivity-related distance variations. Therefore, we take a ball with multi-reflective properties as the target object, and put it in a corner of the wall, as shown in Figure 1. The TOF camera here adopted a commercially available camera-Pico Zense DCAM710. The DCSL here used our system in [20]. We reconstruct the ball with TOF and DCSL, respectively, and put the results in Figure 1.

Consider TOF's reconstruction data at first. It can be seen that it is very dense, and shows superior performance in depth discontinuous regions, such as the region between the ball's edge and the background wall. However, because of MPI, the corner of a wall cannot be recovered correctly. More specifically, the right angle is reconstructed into an obtuse angle. In addition, the ball's surface is deformed in regions that have different reflective properties. In contrast, the DCSL can reconstruct the surface of the ball and the background wall correctly. This is mainly because the reflected light in DCSL was encoded, while in TOF, the reflected light may possibly be disturbed by multiple reflections. Nevertheless, limited by the spatial codification strategies, the reconstruction data of DCSL is discrete. In most spatially coded SL, researchers try to achieve dense results through the

interpolation method. However, the interpolation results are less reliable. Moreover, this method cannot handle the discontinuous regions well. Considering the complementary characteristics of TOF and DCSL, fusing them to obtain better results becomes an interesting possibility. The main challenge in fusing them is accurate data alignment method and effective fusion strategy.
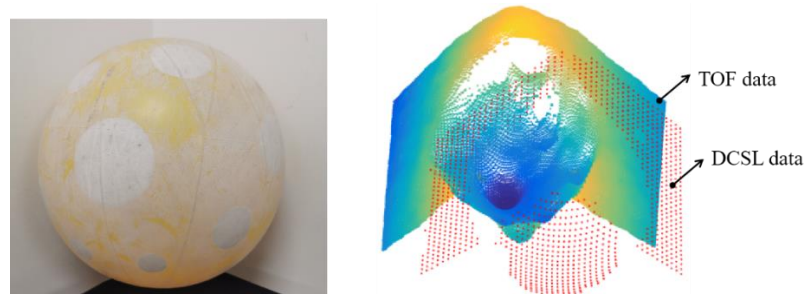
Figure 1. Comparison of a ball's reconstruction results by TOF and DCSL. Note that the ball's surface is of multi-reflective properties, and it was placed in a corner of the wall.

3. Key Technology and Algorithm

The overall algorithm pipeline is shown in Figure 2, which includes three key procedures: data alignment, data fusion, and post processing. First, the two data sets from TOF and DCSL were initially aligned based on the calibrated parameters and ICP (iterative closest point). Second, to fuse them elaborately, the whole scene was segmented into different regions based on the morphological operation. After that, the final fusion was conducted region by region. At last, the necessary post-processing method was conducted to provide better result.
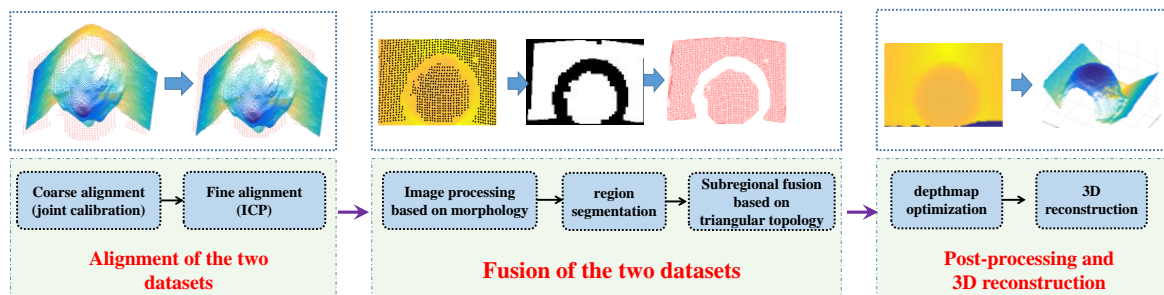
Figure 2. Pipeline of the proposed algorithm.

3.1. Data Alignment

Data alignment is the premise of data fusion. In our system, the reconstruction results of TOF and DCSL were recovered under different coordinate systems. Therefore, the first step is to align them in the same coordinate system. Let A be TOF's 3D point set with N_a points denoted $\vec{a}_i$: $A = \left\{ \vec{a}_i \right\}$ for $i = 1, \ldots, N_a$. Similarly, let B be DCSL's 3D point set with N_b points denoted $\vec{b}_i$: $B = \left\{ \vec{b}_i \right\}$ for $i = 1, \ldots, N_b$. The goal of data alignment is to make A and B as close as possible. In this paper, two steps of data alignment were adopted: coarse alignment and fine alignment.

Coarse alignment. Coarse alignment was conducted based on the calibrated stereo parameters between the TOF camera and DCSL camera. This is a typical stereo calibration procedure, and the optimal stereo calibration method in [21] was adopted here. Through this method, the relative transformation $(\Re, t)$ between these two systems can be determined. When both datasets have been retrieved, they can be initially unified by transforming DCSL data into the TOF camera's coordinate system. Notably, it is necessary and better to put the DCSL camera and the TOF camera as close as possible in practice. In this

way, the 3D information recovered by two different systems will have similar perspectives, which allows for a better fusion effect.

Fine alignment. After coarse alignment, the two data sets have been initially aligned, yet their positions may still mismatch. Therefore, fine alignment procedure is necessary. Here the accelerated ICP (iterative closest point) [22] was adopted to estimate the finer position and orientation relationship $(\Re', t')$ between two data sets. Through this method, B can be finely aligned to A by: $B' = \Re' B + t'$.

3.2. Data Fusion

After data alignment, data fusion can be conducted. However, because of the fact that large deviations may possibly exist in TOF data, it is a big challenge to fuse two datasets in 3D space directly. In this work, we tried to fuse them in 2.5D space, that is, the depth map space. The main procedures include depth remapping, region segmentation and sub-regional data fusion.

3.2.1. Calculate the Remapping Depth Information of the Aligned Point Clouds

First, calculate the remapping depth information of the aligned TOF and DCSL data through the following Equation (1),

$$
\begin{aligned}
u_i &= [f_x \cdot \vec{\omega}_i^x / \vec{\omega}_i^z + u_0] \\
v_i &= [f_y \cdot \vec{\omega}_i^y / \vec{\omega}_i^z + v_0] \quad (\vec{\omega}_i = \vec{a}_i, \vec{b}_i) \\
d_i &= \vec{\omega}_i^z
\end{aligned} \tag{1}
$$

In Equation (1), the symbol $[\cdot]$ denotes the round-up operator. $(\vec{\omega}_i^x, \vec{\omega}_i^y, \vec{\omega}_i^z)$ is the 3D coordinates of the point $\vec{\omega}_i$. For TOF data, $i = 1, \ldots, N_a$. For the DCSL data, $i = 1, \ldots, N_b$. (f_x, f_y) and (u_0, v_0) denote the focal length and the image plane's principal point of the TOF camera respectively. The depth value at the pixel (u_i, v_i) is d_i. By Equation (1), the depth map of TOF data and the aligned DCSL data can be achieved. Take the target scene in Figure 1 as an example, the remapping depth maps are shown in Figure 3.

(**a**)   (**b**)   (**c**)
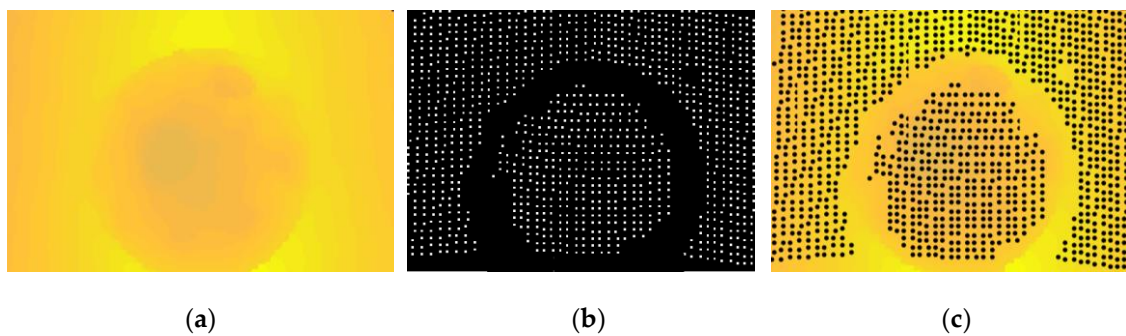
Figure 3. Remapped depth maps of (**a**) TOF data; (**b**) DCSL data; and (**c**) the aligned TOF and DCSL data.

As can be seen from Figure 3, there are some difficulties in fusing depth maps. First, limited by the working principle of DCSL, only continuous surfaces have effective data, such as the spherical surface and the background wall as displayed in Figure 3b. In other words, DCSL cannot handle discontinuous regions. Yet thanks to the opposite characteristics of TOF, the depth information in discontinuous regions can be retrieved smoothly. Considering that, we developed the basic fusion strategy: in continuous regions, DCSL data was utilized to correct TOF data to improve accuracy, and in depth-discontinuous regions, TOF data was filled to guarantee completeness of the whole target scene. From this point, we adopted a sub-regional fusion strategy, which conducts fusion upon continuous and discontinuous regions separately. In continuous regions, TOF data can be corrected by

DCSL data to achieve better accuracy, while in discontinuous regions where DCSL data was missing, TOF data was preserved to keep the completeness of the scene.

3.2.2. Region Segmentation in Depth Maps

$$R = \cup(\bigcup_{i=1}^{M} C_i, \ \bigcup_{k=1}^{N} D_k), C_i \cap D_k = \Phi \tag{2}$$

In Equation (2), the symbol $\cup(\cdot)$ and $\cap(\cdot)$ denote the union and intersection sets respectively. Φ denotes null set. C_i and D_k satisfies $C_i \cap C_j = \Phi (i \neq j)$ and $D_k \cap D_p = \Phi(k \neq p)$ respectively.

Segmentation of continuous regions. As shown in Figure 3b, the depth data of DCSL was discrete; therefore, it is inappropriate to adopt common image segmentation algorithms directly. In this case, the morphological operation was adopted to connect adjacent pixels before region segmentation. Considering the discrete characteristic of DCSL's depth data, a closing operation, followed by an opening operation, was conducted. The closing of the discrete depth data X by S, denoted by $CL_S(X)$, is defined as the dilation of X by a structuring element S followed by the erosion by $\breve{S}$. In addition, the opening operation $OP_{S'}(CL_S(X))$ is defined as the erosion of $CL_S(X)$ by a structuring element S' followed by the dilation by $\breve{S}'$. That is:

$$CL_S(X) \equiv ER_{\breve{S}}\{DI_S(X)\}$$
$$OP_{S'}(CL_S(X)) \equiv DI_{\breve{S}'}\{ER_{S'}(CL_S(X))\} \tag{3}$$

See Figure 4 for an example of this morphological operation on discrete data.
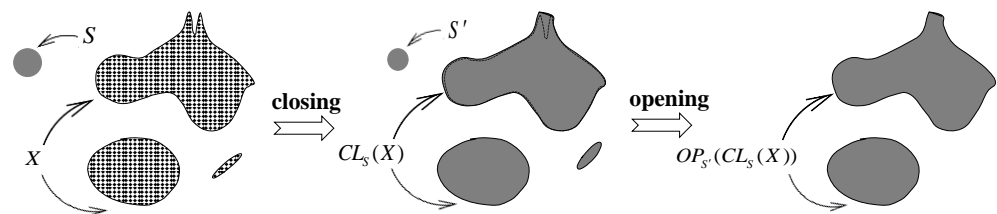
Figure 4. Schematic diagram of the preprocessing of discrete data by using morphological operation.

For DCSL's sparse data, the close operation above can help fill small holes and cracks, and the open operation can help remove isolated points, burr, and small bridges. It is noteworthy that the size and shape of structural elements have an important influence on the segmentation effect; therefore, it should be designed carefully by considering the sparsity of DCSL data in practice. Take the depth map in Figure 3b as an example and its segmentation result with an 8×8 structural element in the closing operation and a 3×3 structural element in the opening operation was displayed in Figure 5a. It can be observed that the whole depth map can be separated into two continuous regions, which are labeled 1 and 2 for convenience.

Segmentation of discontinuous regions. It is noteworthy that the discontinuous regions cannot be classified simultaneously by the above operations. In our fusion strategy, TOF data in continuous regions can be corrected by DCSL data, while in discontinuous regions, TOF data was preserved to keep the completeness of the whole scene. However, it is impossible to fill TOF data into the discontinuous regions as a whole, because TOF data may be distorted a lot in practice and great gaps may occur along the edge of filling. In this case, it is necessary to divide discontinuous regions into different sub-regions. Here we adopted a scanline-based method to finely segment different regions.
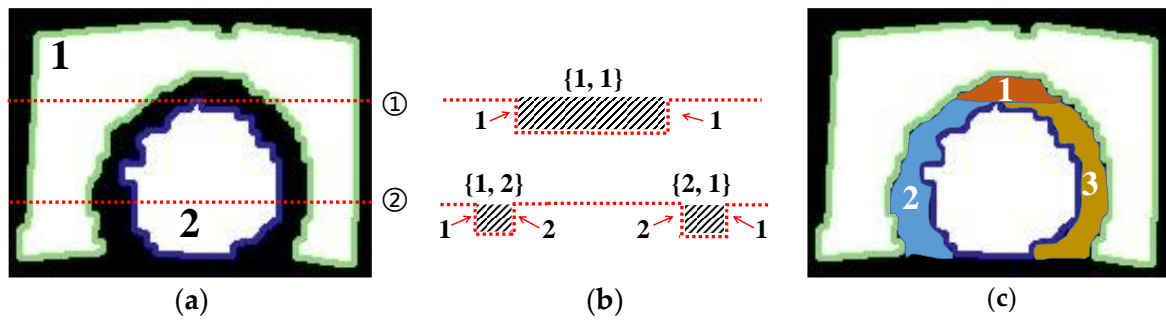
Figure 5. Region segmentation (**a**) Results of initial region segmentation and edge detection; (**b**) the scanline-based segmentation of discontinuous regions; (**c**) segmentation results of discontinuous regions.

First, the canny operator was adopted to find corresponding contours of current regions, which were marked in green and blue lines as in Figure 5a. Then, the whole image was scanned, as in Figure 5a, from up to down. For each scanline, brightness variations will occur when meeting the detected contours, which is also the boundary between continuous and discontinuous regions. Here we denote the positions of brightness variation by the labels of the involved continuous region. Take the scanline ① in Figure 5a,b as an example. Both the two positions where brightness variation occurs are labelled 1. Therefore, the definition of a discontinuous region can be set to be $D\{p, q\}$, with p and q denoting D's nearest left/right brightness variation positions respectively. Based on this definition, there are three discontinuous regions, $D\{1, 1\}$, $D\{1, 2\}$, $D\{2, 1\}$ as displayed in Figure 5c. In this way, all continuous and discontinuous regions of the image R as illustrated in Equation (2) can be achieved.

3.2.3. Sub-Regional Data Fusion

Since the depth maps of TOF and DCSL have been aligned in advance, they can be segmented synchronously into a series of continuous regions C_i and discontinuous regions D_k by the proposed method in Section 3.2.2. After that, the data fusion can be conducted region by region. Note that DCSL data was missing in discontinuous regions; therefore, we preserve TOF data directly to keep the completeness of the whole scene, just as mentioned above. In continuous regions C_i, the key challenge is how to correct dense TOF data based on accurate yet discrete DCSL data. Considering the characteristics of TOF and DCSL data, a topology-based method was proposed. Take one of the continuous regions as an example. The basic procedure of fusion is displayed in Figure 6.
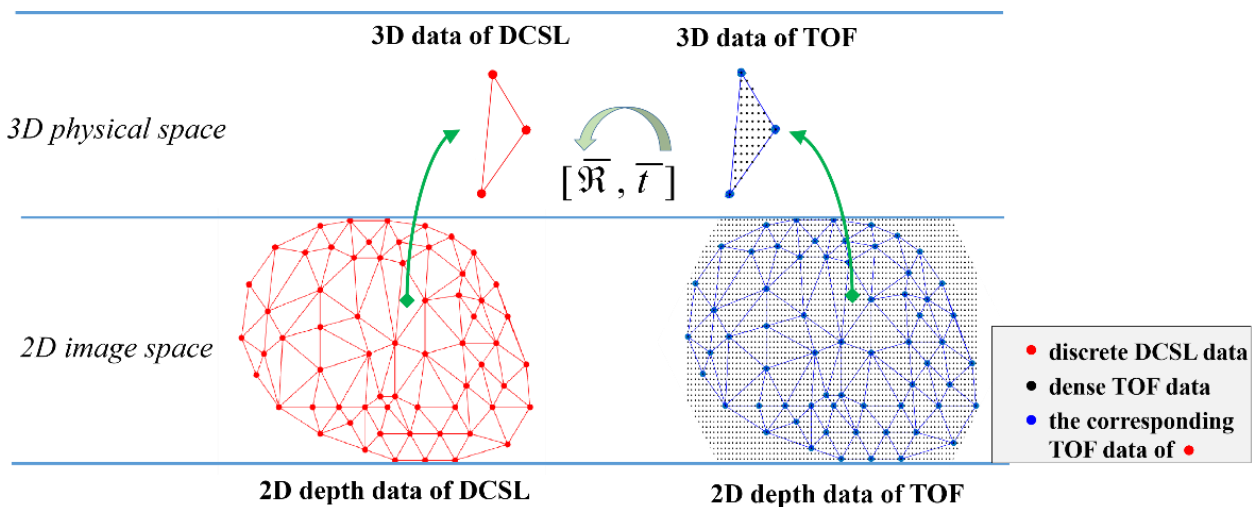
Figure 6. Subregional fusion of TOF and DCSL data based on triangular topology.

First, the matching counterpart of each DCSL datum in TOF data was found. The matching result was determined to be the closest TOF point against each DCSL point in Euclidean distance. The matching point pairs of DCSL and TOF were denoted as $\{S\} \leftrightarrow \{T\}$ for simplicity. In Figure 6, the point in $\{S\}$ and $\{T\}$ was marked as red and blue dots, respectively. Second, establish the Delaunay triangular topology based on $\{S\}$ and $\{T\}$. Note that both $\{S\}$ and $\{T\}$ share the same topology structure and therefore each triangle is one to one correspondence. Then, we can improve the accuracy of the whole TOF data triangle by triangle. For every triangle in $\{S\}$ and $\{T\}$, which were denoted as S_Δ and T_Δ for convenience, the best-fitting transformation that aligns T_Δ against S_Δ can be calculated based on the three vertices by adopting the method in [23]. At last, with the calculated transformation matrix, all TOF data inside the triangle of T_Δ can be synchronously aligned against S_Δ. It can be seen that since their fusion was conducted triangle by triangle, with the overall precision being guaranteed by DCSL data, high-accuracy and high-density reconstruction can be accomplished.

3.3. Post-Processing of the Fused Depth Map

After the fusion strategy mentioned above, initial high-density and high-accuracy depth information can be achieved. However, even if the fusion is conducted delicately in sub-regions, there will still be small holes, cracks, noisy outliers, and so on. In this case, post-processing of the fused depth map was necessary. An edge-preserving filter was first adopted to smooth the depth noise while attempting to preserve edges [1]. One-dimensional exponential moving average (EMA), with a coefficient α that determines the amount of smoothing, was used. Then, filtering was conducted by raster scanning the depth map in both axes recursively. The specific recursive equation is shown below.

$$S_t = \begin{cases} Y_1 & t = 1 \\ \alpha Y_t + (1-\alpha)S_{t-1} & t > 1 \ \& \ \Delta < \delta \\ Y_t & t > 1 \ \& \ \Delta > \delta \end{cases} \tag{4}$$

Here Y_t denotes the newly recorded instantaneous depth value and S_{t-1} denotes the EMA at any time t. The symbol Δ implies depth difference: $\Delta = |S_t - S_{t-1}|$. EMA was adopted only when neighboring pixels have a depth difference Δ less than a pre-defined threshold δ. We empirically set $\alpha = 0.1$ and $\delta = 3$ in practice. Then the spatial hole-filling method was used to fill small holes/cracks by making use of the neighboring left or right valid pixels within a specified radius. In this way, an optimal depth map can be achieved, and therefore more accurate reconstruction results can be obtained.

To sum up, by the fusion of TOF and DCSL data, the TOF data can be corrected by the accurate DCSL data, and the depth errors caused by MPI and surface reflectivity of the TOF camera can be effectively eliminated. At the same time, TOF data is used to fill the sparse structure of DCSL data, especially in discontinuous areas, which can greatly improve the density of reconstructed point clouds and keep the integrity of the target scene. In this way, high-accuracy and high-density 3D reconstruction can be accomplished.

4. Experimental Results

The prototype of the fusion system is shown in Figure 7, which consists of a commercially available TOF camera (Pico Zense DCAM710, Pico Technology Co., Ltd., Cambridgeshire, UK) and a DCSL system. The TOF camera includes an infrared depth camera (640 × 480, 30 FPS) and an RGB camera (1920 × 1080, 30 FPS). The working distance of this system was 0.2~5.0 m. The DCSL system consists of an industrial camera of HIKVISION MV-CA003-21UM (640 × 480, 817 FPS, RMA Electronics, Inc., Hingham, MA, USA) and a projector of BENQ w1700s, BenQ, Taipei, Taiwan. The projector is of resolution 3840 × 2160, which has a projection ratio of 1.47~1.76:1. The calibration results of the system are shown in Table 1.
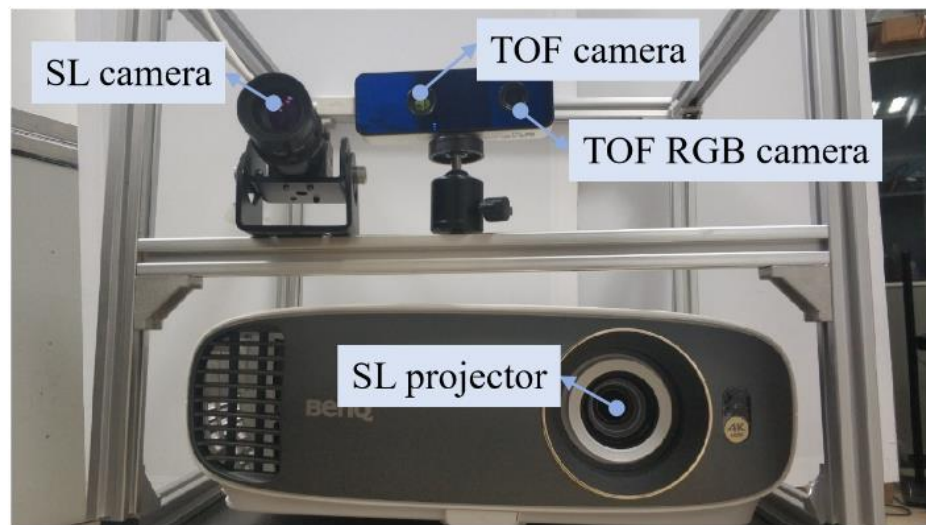
Figure 7. Experimental setup.

Table 1. Calibration results.

	Focal Length/Pixels	Principal Points/Pixels	Lens' Distortion Coefficients
SL camera	(1277.81, 1276.70)	(343.88, 237.42)	(-0.12, -0.14, 0, 0, 0)
SL projector	(6562.87, 6560.45)	(1937.94, 3010.70)	(-0.01, -0.19, 0, 0, 0)
TOF camera	(464.30, 463.92)	(319.01, 252.31)	(0.50, -0.08, 0, 0, 0)
SL camera vs. SL projector	Translation vector/mm: (168.79, -146.43, 124.43) Rotation vector: (0.17, -0.14, 0.04)		
TOF camera vs. SL camera	Translation vector/mm: (-30.97, -18.31, -56.08) Rotation vector: (-0.09, 0.07, 0.06)		
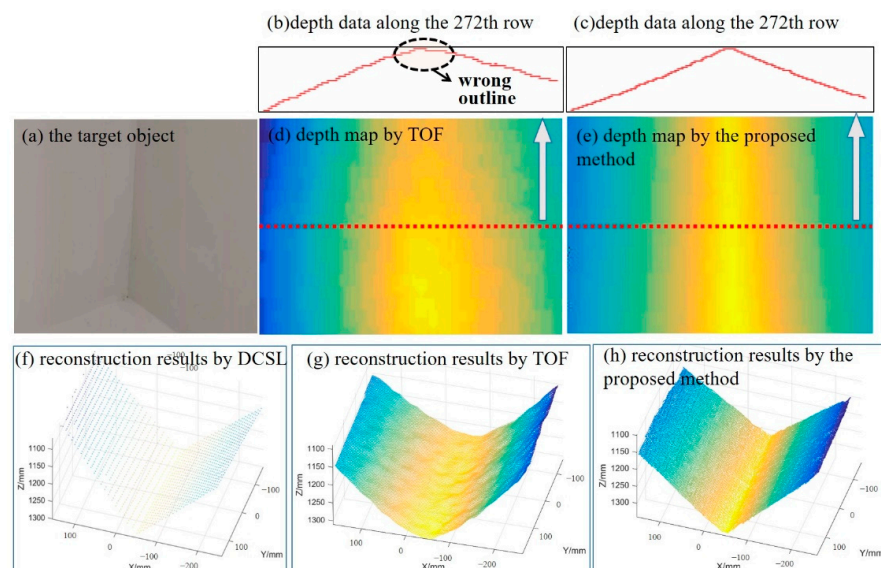
Figure 8. Reconstruction results of two perpendicular planes. (**a**) the target object; (**b**) estimated depth by TOF along the 272nd row; (**c**) estimated depth by the proposed method along the 272nd row; (**d**) estimated depth map by TOF; (**e**) estimated depth map by the proposed method; points clouds reconstructed by (**f**) DCSL; (**g**) TOF; and (**h**) the proposed method respectively.

4.1. Performance of the Proposed Fusion Method

To demonstrate the benefits of the proposed fusion approach, we evaluate the accuracy of TOF, DCSL, and the proposed method comparatively in our first experiment. Two standard planes perpendicular to each other and a standard spherical surface with complex reflective properties were utilized as the measurement targets. Their reconstruction results are displayed in Figures 8 and 9, respectively.
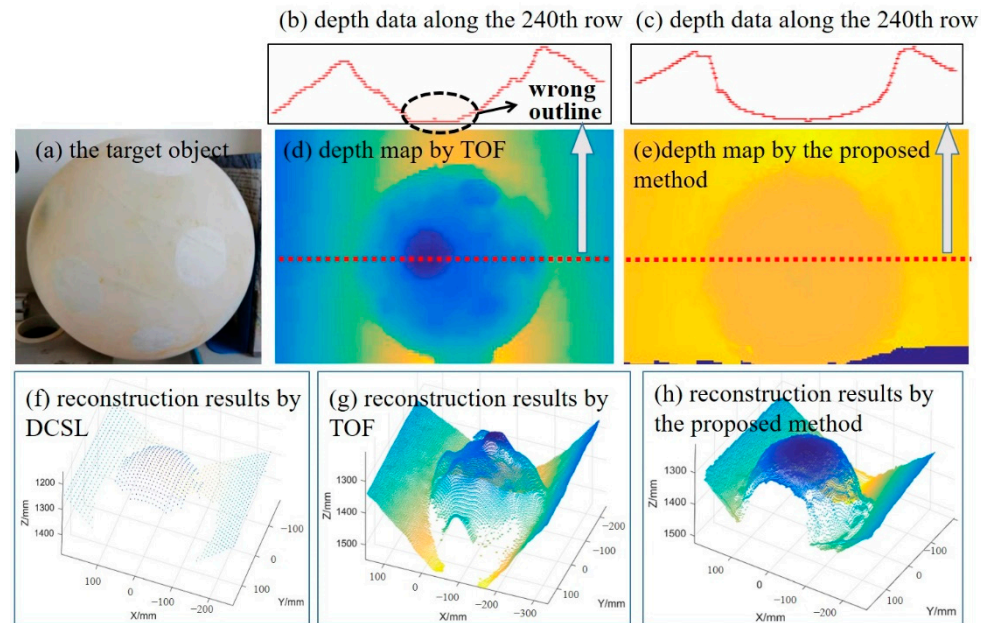
Figure 9. Reconstruction results of the spherical surface. (**a**) the target object; (**b**) estimated depth by TOF along the 240th row; (**c**) estimated depth by the proposed method along the 240th row; (**d**) estimated depth map by TOF; (**e**) estimated depth map by the proposed method; points clouds reconstructed by (**f**) DCSL; (**g**) TOF; and (**h**) the proposed method respectively.

To evaluate the results quantitatively, experimental data were analyzed and are provided in Table 2. Note that for two perpendicular planes, the number of reconstructed point clouds, the angle between the two planes, and the plane fitting errors were designated as the key indicators to evaluate the reconstruction effect. As for the spherical surface, the number of reconstructed point clouds, the sphere fitting error, and the diameter's estimation error were designated as the key indicators.

Table 2. Reconstructed results of TOF, DCSL and the proposed method.

Methods	Perpendicular Planes				Spherical Surface		
	Number of Point Clouds/k [a]	Angle Errors/°	Plane Fitting Error/mm		Number of Point Clouds/k	Sphere Fitting Error/mm	Diameter's Estimation Error/mm
			Left Plane	Right Plane			
TOF	~48	10.05	4.40	6.60	~37	13.48	9.06
DCSL	~1.5	0.40	0.52	0.61	~1.2	0.54	0.92
our method	~47	0.62	0.89	0.87	~34	2.20	1.60

[a] k denotes one thousand.

Several observations can be made from this experiment. First, as shown in Figure 8 and Table 2, TOF can obtain much more dense point clouds, but it cannot measure perpendicular planes correctly due to MPI. Not only were the two planes reconstructed with large errors, but the intersection angle between them also shows obvious deviations. As shown in Table 2, the angle error of TOF camera is as high as $10.05°$. As a comparison, DCSL can only obtain a limited number of point clouds, but it can recover accurate 3D structures of the target objects. DCSL can handle MPI effect well, and the measured angle error was only $0.40°$. The angle error of the proposed method is $0.62°$. Though the accuracy is a little

lower than DCSL, it is much better than TOF. Our method combines the advantages of TOF and DCSL, which can not only realize high-density point clouds, but also solve the problem of MPI. Second, it can be observed in Figure 9 that the accuracy of TOF was largely affected by the reflectivity characteristic of target surfaces. Take the spherical surface as an example. The number of point clouds built by TOF was about 37000, while the DCSL can only build 1200 points. At the same time, the sphere fitting error of TOF data was 13.48 mm, which is mainly due to the multi-reflectivity-related deviations, while the DCSL can handle this problem well and the fitting error was only 0.54 mm. This clearly reflects their complementary characteristics. By fusing them, the proposed method obtained about 34000 points, and the sphere fitting error was 2.20 mm. This means that in our method, the final error of 3D reconstruction is greatly reduced compared with TOF under the condition of ensuring high-density performance.

4.2. Performance of the Proposed Method in Different Working Distances

In the second experiment, the measurement accuracy of the proposed method was tested in different working distances. A standard plane which is of size 50 cm × 50 cm and owns the machining accuracy of ±5 μm was utilized as the target object. It was placed at different distances against the experimental setup and the standard deviation of the reconstructed plane was taken as the precision evaluation index. The comparative results of TOF, DCSL, and the proposed method are displayed in Figure 10.
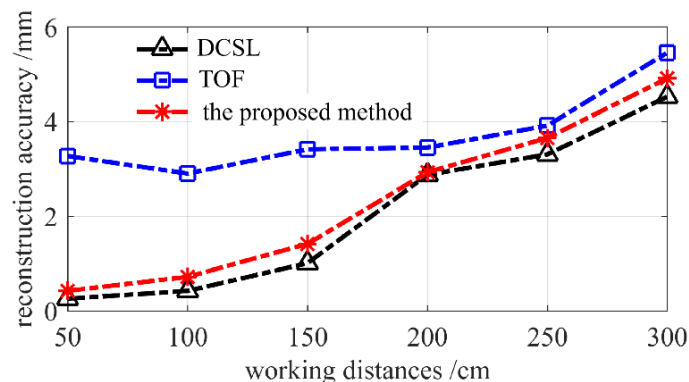
Figure 10. Reconstruction accuracy comparison of TOF, DCSL and the proposed method in different working distances.

It can be observed in Figure 10 that compared with TOF, the presented technique's accuracy improves greatly at the working range from 0.5~3.0 m. Its advantage was especially obvious at the close range. Taking 1.0 m as an example, the accuracy of TOF and our method were 2.90 mm and 0.72 mm, respectively. The accuracy has been improved up to 75.17%. It is safe to conclude that by introducing DCSL into the TOF camera, submillimeter accuracy can be achieved at close range. With the increase of working distances, the advantages of DCSL become less and less obvious. This is the shortcoming of all methods that working based on the triangular reconstruction principles. Therefore, in long range, TOF begins to play an important role in our fusion method. In this way, both DCSL and TOF's advantages can be combined effectively in real measurement.

It is noteworthy that the accuracy of the proposed method is slightly lower that DCSL method. At the working distance of 1.0 m, the accuracy of DCSL and the proposed method were 0.42 mm and 0.72 mm respectively. This is mainly because in the fusing process, the discrete data of DCSL were used to maintain the correct structure of 3D data, while local holes were filled with TOF's dense data. Generally, the sparser the DCSL's data, the lower the fusion accuracy. Therefore, it is easy to conclude that the sparsity of DCSL played an important role in determining the fusion accuracy. Further research and analysis was conducted in the next experiment.

4.3. Performance of the Proposed Method w.r.t. the Sparsity of DCSL

The performance of our fusion method with regard to the sparsity of DCSL's point clouds was explored in this experiment. The standard plane and sphere in Figures 8a and 9a were taken as the target objects. The related TOF data remain unchanged, yet the DCSL data were sampled to simulate different sparsities. The original DCSL data were assumed to be X. Then, X was sampled with different ratios: 1/4, 1/2, 1, and 2. The results are denoted by 1/4X, 1/2X, 1X, and 2X accordingly. Note that 1/4X and 1/2X refer to the downing sampling operations and 2X refer to up sampling by data interpolation.

It can be seen from Figures 11 and 12 that the sparsity of DCSL did have obvious effects on the final fusion accuracy. In the case of 1/4X density, the corner error of two perpendicular planes was 0.68°, while in the case of 2X density, the error was decreased to 0.41°. As for the sphere fitting errors, the error in the case of 1/4X density was as high as 2.92 mm, yet was only 0.80 mm in the case of 2X density. It is appropriate to conclude that the denser the DCSL data, the higher the fusion accuracy. This is easy to understand. The accurate DCSL data were designated as "anchor points" to keep the whole structure of target objects. More anchor points lead to higher precision. However, because the coding window of the DCSL pattern cannot be compressed to a very small size, the density of DCSL data was limited. In the practical design of the DCSL pattern, the optimal coding window should be chosen by considering both coding density and each pattern blocks' recognition accuracy. Details refer to [20].
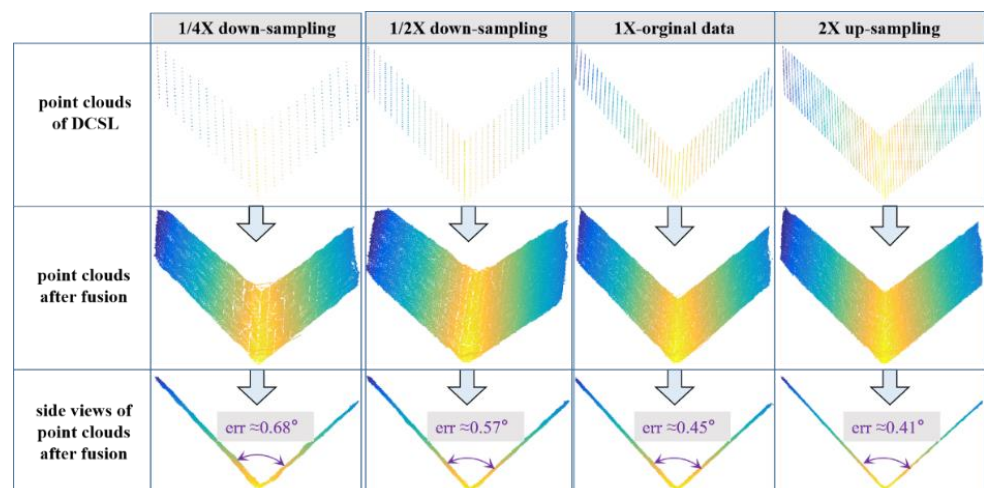
Figure 11. Corner reconstruction errors of two perpendicular planes with regard to the sparsity of DCSL's data.

4.4. Performance of the Proposed Method in Complex Scenes

In the last experiment, we validate the performance of our fusion system in measuring complex scenes. The scenario is shown in Figure 13a. Corners of the scene detected by using DCSL method were displayed in Figure 13e. Point clouds reconstructed by DCSL, TOF, and the proposed method were put in Figure 13b–d respectively. As shown in Figure 13g, the cardboard box below the ball and the background wall show great deviations due to the MPI effect of TOF, while by introducing the accurate DCSL data, the whole structure of the target scene can be kept effectively. In this way, 3D depth information can be recovered densely and accurately.
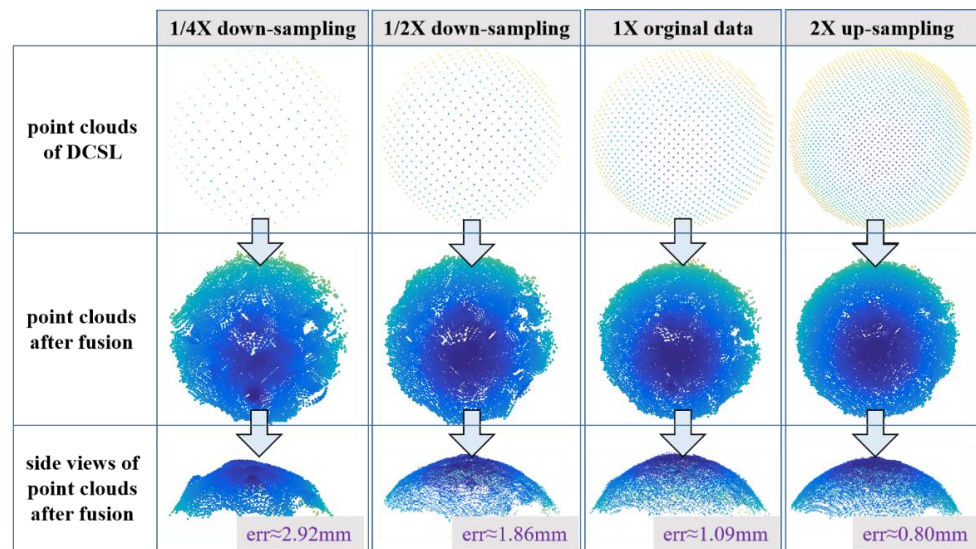
Figure 12. Sphere fitting errors w.r.t. the sparsity of DCSL's data.
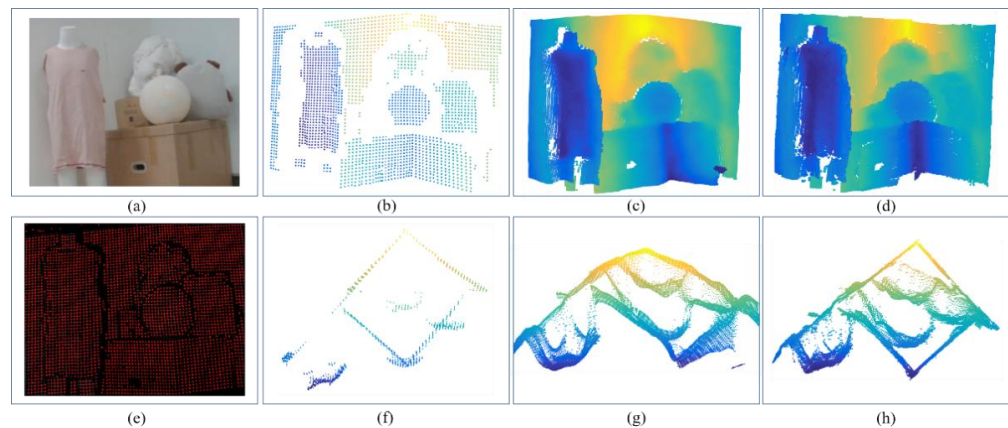
Figure 13. Reconstruction results of complex scenes. (**a**) the target scene, (**e**) detected corner by DCSL; Point clouds by (**b**) DCSL, (**c**) TOF, (**d**) the proposed method, and the side view of point clouds by (**f**) DCSL, (**g**) TOF, (**h**) the proposed method respectively.

5. Discussion

With the advent of the 5G era, mobile intelligent terminals such as mobile phones have become crucial portals for 3D visual data and AI perception. TOF and structured light, as the two most popular technologies applied at intelligent terminals, have been deeply and widely researched in recent years. However, both of them have their advantages and disadvantages. This paper explored a new 3D vision technology for intelligent terminals, which is a combination of these two mainstream depth cameras. Extensive experimental results validated that in our method, the MPI and reflectivity-related distance variations of TOF can be effectively eliminated. This shows great potential to accomplish high-accuracy and high-quality 3D measurement task at intelligent terminals, for example, 3D face accurate recognition, and 3D visual guidance and so on. In our next step, we will focus on developing miniaturized devices that can be inserted into intelligent terminals. Possible improvements include making TOF and DCSL share a common camera, designing miniaturized projection devices based on diffractive optical elements etc. In addition, a new fusion strategy based directly on the point clouds will also be explored. It should be pointed out that there may be new challenges and problems in developing miniaturized devices. For example, if TOF and DCSL share a camera, interference phenomenon may affect the reconstruction results, which will be considered in our future research.

6. Conclusions

We propose and experimentally demonstrate a new fusion technique for 3D dense and accurate reconstruction. The present technique can realize the complementary advantages of two mainstream 3D reconstruction technologies: TOF and structured light. By using the high-precision depth information of spatially-encoded structured light, and the high-density depth information of TOF, the overall depth information of high accuracy, as well as high density, can be retrieved. In addition, the MPI and reflectivity-related deviations of TOF can be effectively eliminated. Since the presented algorithm is the fusion of two mature techniques that have been widely adopted in mobile intelligent terminals, it has great potential to be transplanted to the Android/iOS platform on Smartphones or iPads in the future.

Author Contributions: Conceptualization, F.G. and Z.S.; methodology, F.G.; software, F.G., H.C. and P.X.; validation, F.G., H.C. and P.X.; writing—original draft preparation, F.G.; writing—review and editing, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key-Area Research and Development Program of Guangdong Province, grant number 2019B010149002, the National Natural Science Foundation of China (NSFC), grant number 62105352; the Natural Science Foundation of Guangdong Province, grant number 2020A1515010486; the Natural Science Foundation of Shenzhen, grant number JCYJ20190806171403585, the Shenzhen Science and Technology Program, grant number RCBS20200714114921207, and the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy(SZ), grant number GML-KF-22-25.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

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

References

1. Zabatani, A.; Surazhsky, V.; Sperling, E.; Moshe, S.B. Intel® RealSense™ SR300 Coded light depth Camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *99*, 2333–2345. [CrossRef] [PubMed]
2. Sarbolandi, H.; Sarbolandi, D.H.; Lefloch, D.; Kolb, A. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Comput. Vis. Image Underst.* **2015**, *139*, 1–20. [CrossRef]
3. Zhang, S. High-speed 3D shape measurement with structured light methods: A review. *Opt. Lasers Eng.* **2018**, *106*, 119–131. [CrossRef]
4. Lindner, M.; Schiller, I.; Kolb, A.; Koch, R. Time-of-Flight sensor calibration for accurate range sensing. *Comput. Vis. Image Underst.* **2010**, *114*, 1318–1328. [CrossRef]
5. Heng, S.G.; Samad, R.; Mustafa, M.; Abdullah, N.R.H.; Pebrianti, D. Analysis of performance between Kinect v1 and Kinect v2 for various facial part movements. In Proceedings of the IEEE International Conference on System Engineering and Technology, Shah Alam, Malaysia, 7 October 2019; pp. 17–22.
6. Feng, S.; Zuo, C.; Zhang, L.; Tao, T.; Hu, Y.; Yin, W.; Chen, Q. Calibration of fringe projection profilometry: A comparative review. *Opt. Lasers Eng.* **2021**, *143*, 106622. [CrossRef]
7. Whyte, R.; Streeter, L.; Cree, M.; Dorrington, A. Review of methods for resolving multi-path interference in time-of-flight range cameras. In Proceedings of the IEEE Sensors, Valencia, Spain, 2–5 November 2014; pp. 629–632.
8. Bhandari, A.; Kadambi, A.; Whyte, R.; Barsl, C. Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization. *Opt. Lett.* **2014**, *39*, 1705–1708. [CrossRef] [PubMed]
9. Kadambi, A.; Whyte, R.; Bhandari, A.; Streeter, L. Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.* **2013**, *32*, 167. [CrossRef]
10. Fuchs, S.; Suppa, M.; Hellwich, O. Compensation for multipath in TOF camera measurements supported by photometric calibration and environment integration. In *Computer Vision Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 31–41.
11. Jiménez, D.; Pizarro, D.; Mazo, M.; Palazuelos, S. Modelling and correction of multipath interference in time of flight cameras. In Proceedings of the CVPR, Providence, RI, USA, 16–21 June 2012; pp. 893–900.
12. Whyte, R.; Streeter, L.; Cree, M.J.; Dorrington, A.A. Resolving multiple propagation paths in time of flight range cameras using direct and global separation methods. *Opt. Eng.* **2015**, *54*, 113109. [CrossRef]
13. Agresti, G.; Zanuttigh, P. Deep Learning for Multi-Path Error Removal in TOF Sensors. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland, 2018.

14. Qian, J.; Tao, T.; Feng, S.; Chen, Q.; Zuo, C. Motion-artifact-free dynamic 3D shape measurement with hybrid Fourier-transform phase-shifting profilometry. *Opt. Express* **2019**, *27*, 2713. [CrossRef] [PubMed]

15. Wu, Z.; Zuo, C.; Guo, W.; Tao, T.; Zhang, Q. High-speed three-dimensional shape measurement based on cyclic complementary gray-code light. *Opt. Express* **2019**, *27*, 1283–1297. [CrossRef] [PubMed]

16. Song, Z.; Jiang, H.L.; Lin, H.B.; Tang, S.M. A high dynamic range structured light means for the 3D measurement of specular surface. *Opt. Lasers Eng.* **2017**, *95*, 8–16. [CrossRef]

17. Fu, B.; Li, F.; Zhang, T.; Jiang, J.; Li, Q.; Tao, Q.; Niu, Y. Single-shot colored speckle pattern for high accuracy depth sensing. *IEEE Sens. J.* **2019**, *19*, 7591–7597. [CrossRef]

18. Gu, F.F.; Song, Z.; Zhao, Z.L. Single-shot structured light sensor for 3D dense and dynamic reconstruction. *Sensors* **2020**, *20*, 1094. [CrossRef] [PubMed]

19. Zhan, S.; Tang, S.M.; Gu, F.F.; Chu, S.; Feng, J.Y. DOE-based structured-light method for accurate 3D sensing. *Opt. Lasers Eng.* **2019**, *120*, 21–30.

20. Gu, F.F.; Cao, H.Z.; Song, Z.; Xie, P.J.; Zhao, J.; Liu, J. Dot-coded structured light for accurate and robust 3D reconstruction. *Appl. Opt.* **2020**, *59*, 10574–10583. [CrossRef] [PubMed]

21. Gu, F.; Zhao, H.; Ma, Y.; Bu, P.; Zhao, Z. Calibration of stereo rigs based on the backward projection process. *Meas. Sci. Technol.* **2016**, *27*, 085007. [CrossRef]

22. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. Sensor fusion IV: Control paradigms and data structures. *Int. Soc. Opt. Photonics* **1992**, *1611*, 586–606.

23. Sorkine, O. Least-squares rigid motion using svd. *Tech. Notes* **2017**, *120*, 52.