

Real-time and accurate monocular 3D sensor using the reference plane calibration and an optimized SGM based on opencl acceleration

Wei Yin^{a,b,c}, Lu Cao^{a,b,c}, Hang Zhao^{a,b,c}, Yan Hu^{a,b,c}, Shijie Feng^{a,b,c}, Xiaolei Zhang^b, Detong Shen^b, Huai Wang^d, Qian Chen^c, Chao Zuo^{a,b,c,*}

^a Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

^b Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210019, China

^c Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing, Jiangsu Province 210094, China

^d Suzhou Abham Intelligent Technology Co., Ltd., Suzhou, Jiangsu Province 215000, China

ARTICLE INFO

Keywords:

Speckle projection profilometry
3D Imaging
Stereo matching
Opencl

ABSTRACT

Speckle projection profilometry (SPP), as a promising structured light projection technique, can achieve global unambiguous 3D measurement by projecting a single random speckle pattern. In addition, the projected speckle pattern is usually etched into the microstructure of highly integrated Vertical-Cavity Surface-Emitting Laser (VCSEL), which makes the hardware system compact enough to be mounted on mobile devices such as robots. However, since the stereo matching algorithm used in SPP involves high computational overhead, it usually runs in real-time on specially customized hardware platforms such as ASIC/FPGA, rather than general-purpose mobile platforms. In this paper, we propose a real-time and accurate 3D measurement method using a monocular 3D sensor based on the infrared speckle projection. Similar to Kinect v1, our sensor mainly consists of an IR dot projector and one IR camera for projecting and capturing speckle images synchronously. Low-cost and high-quality speckle projection is achieved by customizing the projection pattern of VCSEL and using the beam copy function of Diffractive Optical Elements (DOE). Based on the 3D imaging principle of monocular 3D sensors, a reference plane calibration method is proposed to obtain a high-quality reference speckle image for improving the monocular matching accuracy. Then, benefited from the local memory mechanism and multiple operating synchronizations on the OpenCL environment, an optimized semi-global matching (SGM) algorithm using GPU is presented to achieve efficient and accurate depth reconstruction dynamically. Within the measurement range of 0.8m (length) × 0.5m (width) × 1m (depth), the proposed method can achieve real-time and single-shot 3D imaging with an accuracy of 1.277 mm at 75 FPS on GTX 1060 and 15 FPS on ARM Mail G52(mobile platform).

1. Introduction

Optical 3D measurement has become currently one of the most promising 3D sensing technologies and is extensively applied in industry inspection and scientific research. Various optical 3D measurement approaches have been developed, such as time-of-flight (TOF) [1–3], binocular/multiview stereo vision [4–11], and structured light projection [12–18]. For TOF, a pulse signal is emitted from the near-infrared light source and reflected by the measured object, and then received by the detector. The depth information can be estimated directly by recording the time difference from sending to receiving the light signal. However, due to the limitation of the manufacturing technique, TOF is mainly suitable for long-distance depth sensing, and its measurement accuracy and resolution at short distances are relatively low.

Currently, driverless cars are equipped with TOF-based 3D LiDAR to perceive the surrounding 3D scenes at long distances [19]. Stereo vision is a passive 3D measurement method based on the principle of triangulation. Multiple images of the tested scene are acquired by using two or more cameras from different perspectives. The same object point on each image is found through various stereo matching techniques, and then its disparity can be calculated to obtain the corresponding depth information. The measurement system based on stereo vision has the advantages of simple hardware configuration and easy implementation. However, due to the occlusion or shadows in practical applications, there may be obvious differences in visual information between multiple views, which leads to a higher mismatch rate and missing rate in the disparity maps. In addition, for weak texture regions in the measured scenes, it is also difficult to find corresponding points

* Corresponding author.

E-mail addresses: geniushshijie@163.com (S. Feng), chenqian@njust.edu.cn (Q. Chen), zuochoao@njust.edu.cn (C. Zuo).

from multiple perspectives that limit the 3D measurement accuracy [20–23].

In contrast, the structured light-based 3D measurement methods can be regarded as an improved form of stereo vision, which is realized by replacing a camera with a light source generator (such as a projector) in the system configuration. The light source generator projects a series of special structured light patterns onto the object to be measured [24]. Compared with stereo vision-based methods, the structured light-based methods, as an active 3D sensing technique, can easily overcome the problem of low matching accuracy for weak texture regions. Among 3D measurement methods based on structured light projection, common structured light patterns mainly include fringe patterns and speckle patterns, which have been developed into two mainstream methods accordingly: fringe projection profilometry (FPP) [25–31] and speckle projection profilometry (SPP) [32–35]. For FPP, the projector projects a series of fringe patterns onto the measured scene. The phase information proportional to the surface profile is embedded in the fringe images recorded by the camera and successfully recovered by using various phase retrieval techniques, such as Fourier transform profilometry (FTP) using only a single fringe pattern [36–38] or Phase-shifting profilometry (PSP) at least three fringe patterns [39,40]. But, these methods both perform the arctangent function for phase retrieval thus resulting in the wrapped phase with 2π phase jumps. The operation of phase unwrapping is necessary to eliminate the phase ambiguities and convert the wrapped phases into the absolute ones [41–50]. Therefore, it is difficult to obtain high-precision and absolute phase information from a single fringe image in FPP, which limits its applications in dynamic 3D measurement [51,52].

Different from FPP, the structured light pattern projected in SPP is designed according to various spatial coding strategies: Non-formal codification [53,54], De Bruijn sequences [55–57], and M-arrays [58]. These design methods ensure that the whole measurement space has a global uniqueness by projecting only a single pattern, and assist in the establishment of an accurate global correspondence between stereo images, which makes SPP have the advantage of single-shot 3D reconstruction. The efficient 3D measurement methods based on SPP are popular in consumer electronics, such as Microsoft's motion-sensing device (Kinect), Apple's face recognition (FaceID), intel's stereo camera (RealSense), and orbbec's depth camera (Astra). But, due to the measured objects with complex reflection characteristics and the perspective differences between the stereo camera, it is still difficult to guarantee that every pixel in the whole measurement space has perfect global uniqueness by only projecting a speckle pattern [29,33,34]. In order to solve the common mismatching in SPP, some classical stereo matching algorithms, such as SGM [5,22], ELAS [8], and PatchMatch [9], are used to achieve robust absolute 3D measurement by smoothing the disparity map, but at the cost of matching accuracy. Gu et al. [59] proposed an improved SGM algorithm to increase the measurement accuracy of speckle-based structured light sensor by adopting a new penalty term. In addition, it is more important that the expensive computational overhead required by dense stereo matching brings an enormous challenge for real-time and accurate 3D reconstruction in mobile devices [60].

In the early days, due to the limitation of stereo matching methods and computational performance, some GPU-accelerated local matching methods were proposed to improve the matching accuracy and running speed by using different cost calculation methods or multi-scale matching windows [61–63]. However, in Middlebury, the matching accuracy of these methods is lower than the global stereo matching method. The global stereo matching method achieves accurate pixel-level stereo matching through a global energy function, but it has higher computational complexity and is not suitable for real-time applications. Different from global methods, semi-global matching (SGM) methods have the same algorithmic complexity as local methods by combining one-dimensional cost aggregation results from all directions, achieving efficient and accurate stereo matching [5,22]. A lot of works have proposed different hardware solutions of SGM to implement quasi-real-time, real-

time, and ultra-fast stereo matching. In 2008, Ernst et al. [64] pioneered a fully GPU implementation of SGM with mutual information at 4.2 FPS on GeForce 8800 ULTRA for VGA images (640×480 pixel) and 128 pixel disparity range. In 2011, Banz et al. [65] presented a real-time SGM method with non-parametric rank transform based on GPU acceleration, which reaches 27 FPS for images (1024×768 pixel) with 128 pixel disparity range. In addition, Banz et al. [66] provided the FPGA parallelization scheme at 30 FPS for VGA images. In 2016, benefited from the shared memory mechanism, Hernandez-Juarez et al. [67] proposed a real-time disparity estimation method on embedded GPU devices. Their method reaches 27 FPS on a Tegra X1 for an image size of 640×480 and 128 disparity levels using SGM method including matching with center-symmetric census transform but without any post-processing. Even though some stereo vision approaches utilize high-performance GPU or FPGA platforms to achieve real-time stereo matching, considering that speckle projection-based 3D sensors are generally mounted on mobile devices with weak GPU performance, these methods cannot be directly applied to stereo matching in SPP for meeting the requirements of accuracy and speed at the same time. Therefore, it is necessary to re-evaluate the GPU-based SGM method in SPP and develop new parallelization schemes.

In this work, for enhancing the performance of single-shot 3D imaging of SPP in mobile devices, we propose an infrared speckle projection-based monocular 3D sensor using the reference plane calibration and an optimized SGM based on OpenCL acceleration. In terms of hardware, our 3D sensor mainly consists of an IR dot projector and one IR camera for projecting and capturing speckle images synchronously. Low-cost and high-quality speckle projection is achieved by customizing the projection pattern of the Vertical-Cavity Surface-Emitting Laser (VCSEL) and using the beam copy function of Diffractive Optical Elements (DOE). In terms of software, according to the 3D imaging principle of monocular 3D sensors, a calibration method is proposed to obtain a high-quality reference speckle image for improving the measurement accuracy. In addition, different from the common CUDA-based GPU development environment, we choose the OpenCL framework [68] which is more suitable for mobile devices as the implementation environment of stereo matching. An optimized SGM algorithm using OpenCL acceleration is proposed to obtain efficient, dense, and accurate matching results. The experiment results demonstrate that our 3D sensor can achieve high-quality 3D reconstruction dynamically with the millimeter accuracy at the speed of 75 FPS on GTX 1060 and 15 FPS on ARM Mail G52.

2. Principle

2.1. The infrared speckle projection-based monocular 3D sensor

The monocular 3D sensor developed in this paper is shown in Fig. 1, its main components include an infrared speckle projector (IR dot projector, Crystal Optech C-PM-STR001A-C), illuminators of IR and RGB, a distance detector, a color camera (RGB camera, OmniVision OV02K10), and an infrared camera (IR camera, OmniVision OV9282). The distance between the IR camera and the IR dot projector is 35 mm, which is also the baseline of monocular 3D sensor.

For 3D sensors based on speckle projection, the quality of the projected speckle pattern is critical to its 3D measurement results. The specific implementation of speckle projection is as follows: based on a near-infrared light source, the dot-like distribution of the required speckle pattern is realized by opening holes at specific positions on the top layer of a Vertical-Cavity Surface-Emitting Laser (VCSEL) [69,70] in Fig. 1(a). The infrared light emitted from the VCSEL passes through a collimating lens, which collimates and converges the diffraction beam with a wide lobe into an approximately parallel format based on the principle of light refraction, thus generating a speckle-like structured light pattern. It is worth noting that the high-density speckle pattern is one of the necessary prerequisites for achieving dense and accurate 3D imaging. Therefore, the number of the etched holes in the top layer of the VCSEL should

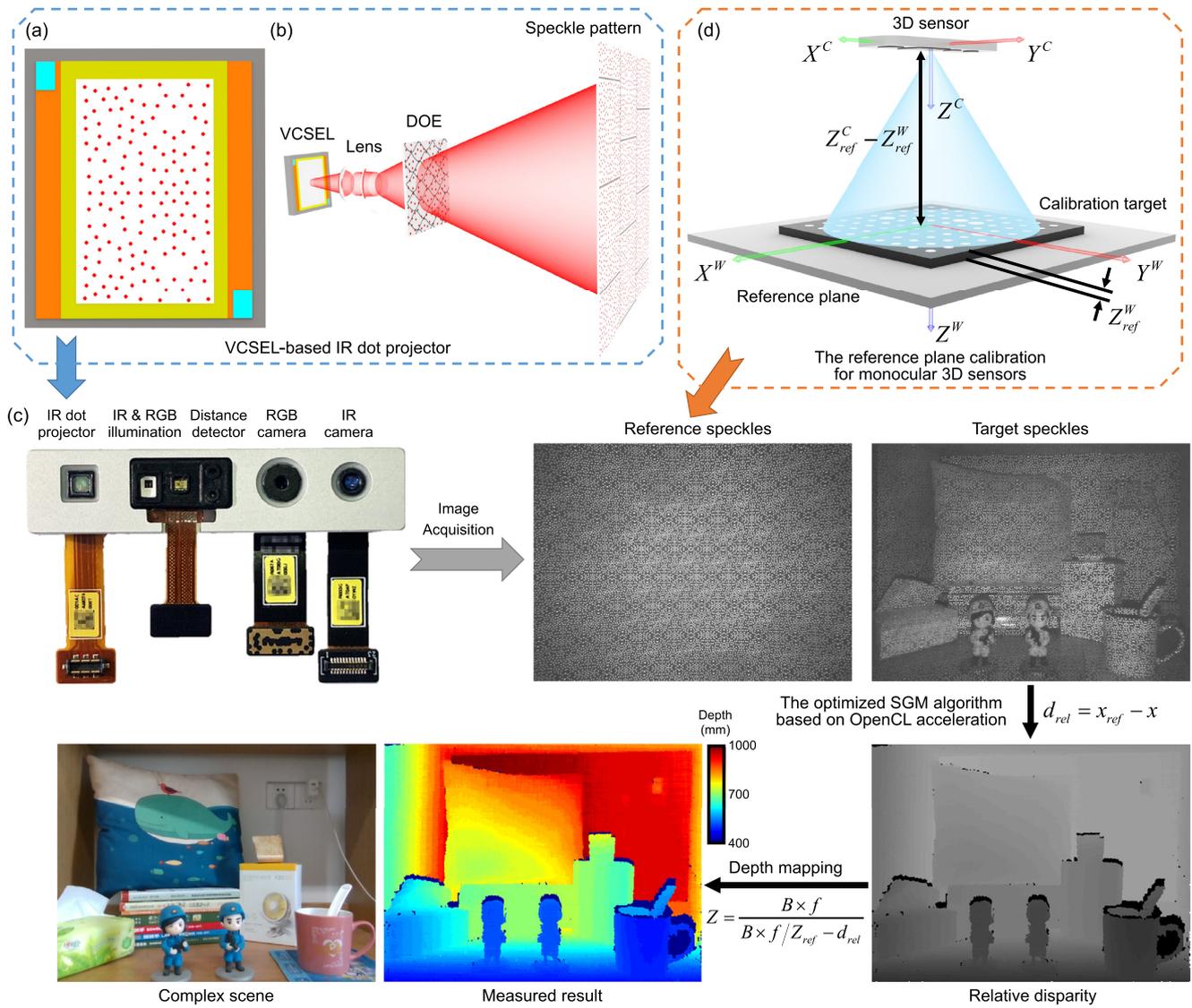


Fig. 1. Overview of the infrared speckle projection-based monocular 3D sensor. (a)-(b) The diagram of infrared speckle projection scheme including VCSEL and DOE. (c) Hardware system and 3D imaging procedure of monocular 3D sensor. (d) The diagram of the proposed calibration method for monocular 3D sensors with the reference speckle image.

be as numerous and dense as possible, but is subject to two aspects: on the one hand, the holes must be large enough to ensure the amount of light passing through to enhance the brightness of the speckle, thus increasing the signal-to-noise ratio (SNR) of the whole projected pattern; on the other hand, if the hole size is too large, the missing reflective area in the top layer of the VCSEL will make it difficult for the VCSEL to resonate, thus affecting the proper operation of the projection module. In Fig. 1(b), the specific structure of DOE is used to disperse the illumination beam and diffract tens or hundreds of sub-patterns that are exactly the same as the incident pattern, thus increasing the number of speckles as well as the measurable field. As shown in Fig. 1(a), according to the customized VCSEL-based miniaturized speckle projection scheme, the overall size of the projection module is less than $1\text{mm} \times 1\text{mm}$, and the emission angle is 25° , which can project a total of 35 thousand clear and uniform speckle spots to the measured scene, enabling a low-cost and high-quality speckle projection.

2.2. The 3D imaging principle of monocular 3D sensors

In a typical 3D imaging system based on stereo vision, two cameras placed along the horizontal direction simultaneously capture stereo im-

age pairs of the target scene. Stereo matching is implemented to obtain the disparity map d , which represents the pixel-wise correspondence between the rectified stereo image pair. The depth map Z of the scene can be calculated as follows [71]:

$$Z = \frac{B \times f}{d - x_c^l + x_c^r}, \quad (1)$$

where B and f are the system baseline and the focal length of the camera, and d represents the horizontal disparity between corresponding points in the two images. x_c^l and x_c^r are the horizontal coordinates of the principal points in the left and right cameras. The speckle projection-based monocular 3D sensor can be regarded as an improved form of stereo vision, which replaces a camera with an IR dot projector to improve 3D imaging results in weakly textured areas. Based on the reversibility of the projection and imaging optical paths, the projector can be regarded as an inverse camera, and the captured image is the projected speckle pattern. However, due to the manufacturing technique and the limited module size, the miniaturized projection module cannot be considered as an ideal camera, and its normal model parameters are not available.

Therefore, another feasible 3D imaging principle is proposed [72]: the monocular 3D sensor is translated along the Z-axis to acquire a series of speckle images of the reference plane at the known depth Z_{ref} within the measured range in Fig. 1(c). Based on Eq. (1), the speckle image of the reference plane will provide a constant and known reference disparity d_{ref} :

$$d_{ref} = \frac{B \times f}{Z_{ref}} + x_c^l - x_c^r. \quad (2)$$

Since the speckle projector is placed horizontally relative to the camera, there is only an approximately horizontal shift between the reference speckle image and the target speckle image:

$$(x_{ref}, y_{ref}) = (x + d_{rel}, y), \quad (3)$$

where d_{rel} is the relative disparity. Then, the actual disparity d can be obtained:

$$d = d_{ref} - d_{rel}. \quad (4)$$

Combining Eqs. (1), (2), and (4), the depth map can be calculated:

$$Z = \frac{B \times f}{d - x_c^l + x_c^r} = \frac{B \times f}{B \times f / Z_{ref} - d_{rel}}. \quad (5)$$

Assuming that the measured depth range of the monocular 3D sensor is $[Z_{min}, Z_{max}]$, the range of d_{rel} can be obtained:

$$\frac{B \times f}{Z_{ref}} - \frac{B \times f}{Z_{min}} \leq d_{rel} \leq \frac{B \times f}{Z_{ref}} - \frac{B \times f}{Z_{max}}. \quad (6)$$

Under the above description, if the relative disparity map and the depth of the reference speckle image are known, the depth of the target scene can be obtained immediately, enabling monocular 3D imaging.

2.3. The calibration method for monocular 3D sensors with the reference speckle image

According to Eq. (5), it can be found that Z_{ref} is closely related to the depth Z of the target image, which affects the accuracy of monocular 3D sensors. Therefore, it is crucial to obtain a reference speckle image with high quality, which requires that the distance Z_{ref} between the reference plane and the IR camera is high-precision and reliable. Fanello et al. at Google exploited a calibration procedure to recover the projected reference pattern successfully [73]. However, calibrating such a monocular 3D sensor is nontrivial due to the complicated procedure. In this subsection, a calibration method is proposed for monocular 3D sensors with the reference speckle image. By approximately replacing the reference plane with a standard calibration board, the extrinsic parameters between the reference plane and the IR camera can be calibrated to obtain a high-quality reference speckle image in Fig. 1(d).

Specifically, the IR camera captures an image of the calibration board, which is placed on the ground (the reference plane). After obtaining the 2D feature coordinates $\mathbf{p}^c = [x^c, y^c]^T$ of the calibration image, according to the standard pin-hole model, we have:

$$s \begin{bmatrix} x^c \\ y^c \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X^W \\ Y^W \\ Z^W \\ 1 \end{bmatrix}, \quad (7)$$

where s is a scaling factor, \mathbf{A} is a 3×3 intrinsic matrix, $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]$ is a 3×3 rotation matrix, \mathbf{t} is a 3×1 translation vector, and $\mathbf{p}^W = [X^W, Y^W, Z^W]^T$ are the world coordinates of features. With the known \mathbf{A} , $\mathbf{p}^c = [x^c, y^c]^T$ can be converted to the normalized coordinates $\mathbf{p}^n = [x^n, y^n]^T$:

$$\begin{bmatrix} x^n \\ y^n \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X^W \\ Y^W \\ Z^W \\ 1 \end{bmatrix}. \quad (8)$$

Assuming that the world coordinate system is defined on the calibration target with $Z^W = 0$, Eq. (8) can also be rewritten as:

$$\begin{bmatrix} x^n \\ y^n \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X^W \\ Y^W \\ 1 \end{bmatrix}. \quad (9)$$

SVD can be implemented to provide the exact solution for Eq. (9), and the last column vector of \mathbf{V} obtained using SVD is the initial guess of $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}]$. Then the Levenberg-Marquardt algorithm is used to solve the nonlinear minimization problem in Eq. (8). Benefited from the uniform thickness Z_{ref}^W of the calibration board with high quality, combining the world coordinates \mathbf{P}^W and the extrinsic matrix $[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{t}]$, the reference plane on 3D camera coordinate system can be represented as:

$$\begin{bmatrix} X_{ref}^C \\ Y_{ref}^C \\ Z_{ref}^C \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X^W \\ Y^W \\ Z_{ref}^W \\ 1 \end{bmatrix}. \quad (10)$$

Z_{ref} can be obtained by calculating the distance from the origin of the 3D camera coordinate system to the reference plane, the normal vector of the reference plane is \mathbf{r}_3 . Finally, by continuously adjusting the pose of the monocular 3D sensor to make \mathbf{r}_3 approach to $[0, 0, -1]^T$, a high-quality reference speckle image at Z_{ref} can be acquired.

2.4. The optimized SGM algorithm based on opencl acceleration

In this section, different from the CUDA-based parallel computing environment with GPUs, an optimized SGM algorithm based on OpenCL acceleration is proposed to obtain efficient and accurate matching results d_{rel} between the reference speckle image and the target speckle image, enabling high-quality 3D reconstruction dynamically for universal mobile devices. There is generally a four-step pipeline for stereo matching, including matching cost calculation, cost aggregation, disparity computation, and disparity refinement. The specific diagram of the proposed OpenCL-based SGM algorithm is shown in Fig. 2.

The matching cost calculation aims to obtain the similarity between a pixel in the target image and its candidates in the reference image. Among different kinds of similarity metrics or cost functions, it has been proved that the census transform is an efficient and robust operator, which provides better matching performance even in the presence of strong radiometric differences [22]. In our method, census transform with a 5×5 window is adopted to extract the local feature vectors of speckle images, which can be described as:

$$Census(x, y) = \bigotimes_{i=-2}^2 \bigotimes_{j=-2}^2 T(I(x, y), I(x+i, y+j)), \quad (11)$$

$$T(I(x, y), I(x+i, y+j)) = \begin{cases} 0, & I(x, y) \leq I(x+i, y+j), \\ 1, & I(x, y) > I(x+i, y+j), \end{cases} \quad (12)$$

where $Census(x, y)$ is a 25-bit feature vector of the central pixel (x, y) , \otimes represents a bit-wise concatenation operator. Since the census transform is a pixel-independent algorithm, the feature vectors of all pixels in the whole target image can be calculated simultaneously. For the simplest OpenCL implementation, $W \times H$ work items are created, each of which reads the intensity values of pixels in a 5×5 window to compute the feature vector of an individual pixel. However, the speckle images are stored in the global memory of GPU, and there is a large access latency for the read operation of the global memory with a total of $W \times H \times 25$ times. Unlike the global memory, the local memory, which is on-chip memory, is shared and accessible to all work items of the workgroup, and its access latency is an order of magnitude lower than the global memory. Considering the high overlap between the windows read by the census transform of adjacent pixels, a local memory-based efficient implementation scheme is proposed that the speckle image is deposited

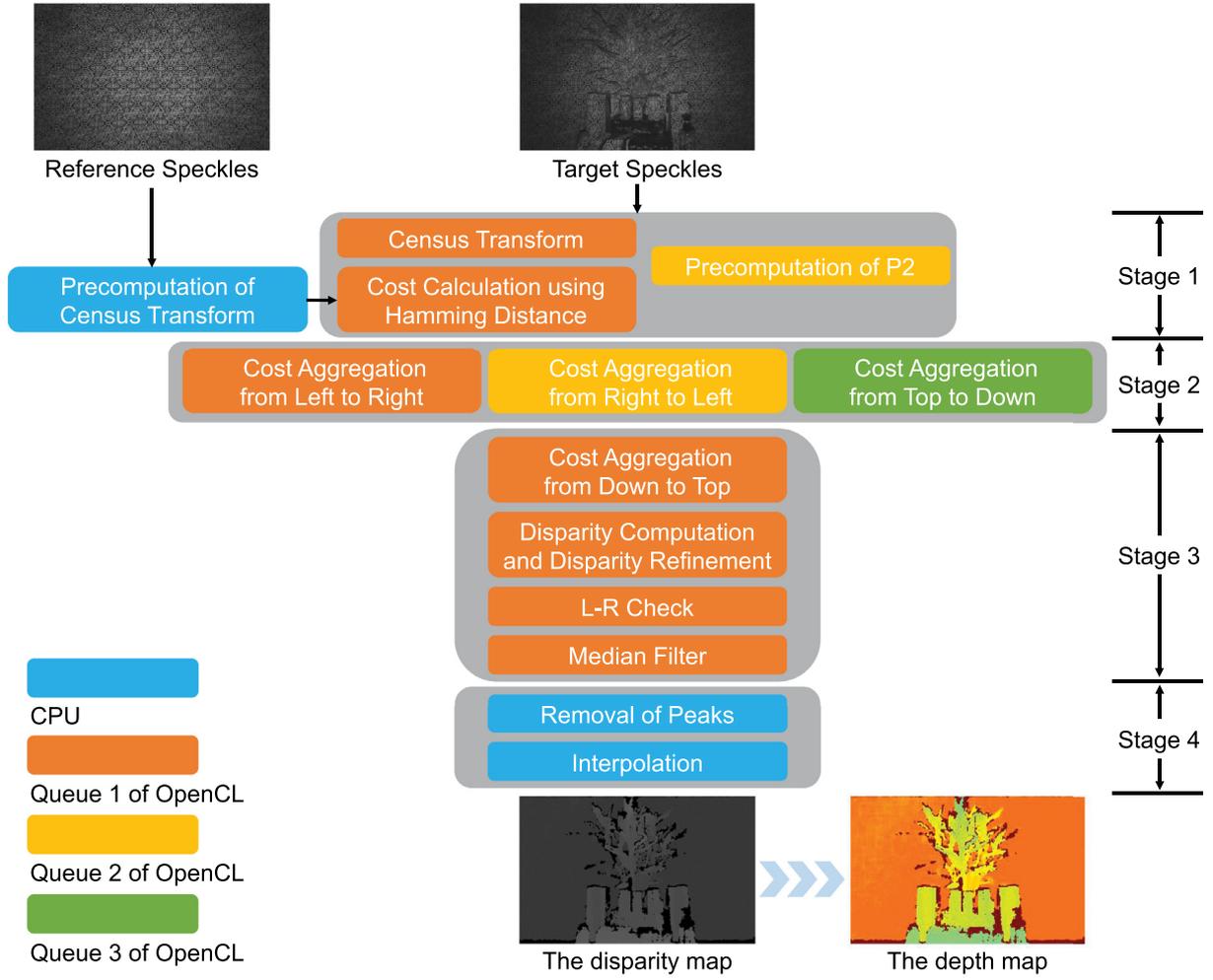


Fig. 2. The schematic diagram of the optimized SGM algorithm based on OpenCL acceleration.

in slices into the local memory of each workgroup. Specifically, the kernel program of OpenCL has $W/32 \times H/32$ workgroups, each of which is divided into 32×32 work items. Each workgroup is responsible for the census transform of 32×32 pixels. In order to make all work item in the workgroup complete the census transformation of a 5×5 window at the same time, each work item will performs one or two read operations of the global memory, thereby writing 36×36 pixels in the speckle image into the local memory. After synchronizing the work items of same workgroup, each work item reads the corresponding 5×5 window to compute the feature vector of a pixel. There are only the read operation of the global memory with a total of $W \times H \times 2$ times and the local memory with a total of $W \times H \times 25$ times.

And then, based on the disparity range $[D_{min}, D_{max}]$ of d_{rel} in Eq. (5), the initial matching cost can be obtained by calculating the Hamming distance between the feature vector of each pixel in the target image and all candidates of the reference image:

$$Cost(x, y, d) = BC(Census(x, y) \oplus Census_{ref}(x + d, y)), \quad (13)$$

where \oplus is a XOR operation, $BC(\bullet)$ is used to count the number of 1 in the XOR result. Since the cost calculation using the Hamming distance is a row-independent algorithm, the initial matching cost of each row in the target image can be calculated simultaneously. Considering the high overlap between candidate reference pixels corresponding to the neighboring pixels in the target image, the feature vectors of each row are stored in a segmented manner into the local memory of each workgroup. Specifically, the kernel program of OpenCL has H work-

groups, each of which is divided into D work items. Each workgroup is responsible for the cost calculation of all pixels in a row, and each work item performs the cost calculation of a target pixel with a total of W/D times. In each time, the feature vectors of D pixels in the target image and the corresponding $2 \times D$ pixels in the reference image are first written into the local memory. After synchronizing the work items of same workgroup, each work item computes the matching cost of a pixel by reading the feature vectors of $D + 1$ pixels in the local memory.

The initial matching cost $Cost(x, y, d)$ is coarse and needs to be further optimized using cost aggregation or cost filtering. In our method, the SGM-based cost aggregation approximates the global solution by aggregating 1D matching costs along 4 independent paths. The aggregated cost $L_i(\mathbf{p}, d)$ of the pixel $\mathbf{p}(x, y)$ at disparity d along a path $\mathbf{r}_i = (r_x, r_y)$ is defined recursively as:

$$L_i(\mathbf{p}, d) = Cost_i(\mathbf{p}, d) + \min \begin{cases} L_i(\mathbf{p} - \mathbf{r}_i, d) \\ L_i(\mathbf{p} - \mathbf{r}_i, d - 1) + P_1 \\ L_i(\mathbf{p} - \mathbf{r}_i, d + 1) + P_1 \\ \min_k L_i(\mathbf{p} - \mathbf{r}_i, k) + P_2 \end{cases} - \min_k L_i(\mathbf{p} - \mathbf{r}_i, k), \quad (14)$$

$$S(\mathbf{p}, d) = \frac{1}{4} \sum_{i=1}^4 L_i(\mathbf{p}, d), \quad (15)$$

where $S(\mathbf{p}, d)$ is final aggregated cost. P_1 is a constant penalty, P_2 is a penalty varies with the intensity gradient:

$$P_1 \leq P_2 = \frac{P_3}{|I(\mathbf{p}) - I(\mathbf{p} - \mathbf{r}_i)|} \leq P_3, \quad (16)$$

where P_3 is another constant penalty. Since the four paths of cost aggregation are $\mathbf{r}_1 = [1, 0]^T$, $\mathbf{r}_2 = [-1, 0]^T$, $\mathbf{r}_3 = [0, 1]^T$, and $\mathbf{r}_4 = [0, -1]^T$, the cost aggregation is a row-independent or column-independent algorithm, i.e., four aggregated costs $L_i(\mathbf{p}, d)$ can be calculated simultaneously. Furthermore, the cost aggregation is performed sequentially along the path for each pixel in a row or column, i.e., the aggregated cost of the current pixel is only related to the aggregated cost of all disparities of the previous pixel. It is obvious that each disparity of the current pixel can be cost aggregated at the same time. In addition, the parallel computing speed can be further improved by computing the aggregated cost of multiple disparities for one pixel simultaneously in each work term. Take $\mathbf{r}_1 = [1, 0]^T$ as an example, the kernel program of OpenCL has H workgroups, each of which is divided into $D/2$ work items. Each workgroup is responsible for the cost aggregation of all pixels in a row, and each work item performs the cost aggregation of two disparities of a target pixel with a total of W/D times. In each time, each work item will perform two write operations of the global memory, thereby writing the aggregated cost of D disparities of the previous pixel into the local memory. After synchronizing the work items of same workgroup, all aggregated costs are ranked using the reduction algorithm to obtain the minimum aggregated cost. Finally, each work item computes the aggregated cost of two disparities of a pixel independently and stores them in the local memory for the cost aggregation of the next pixel.

In disparity computation, the disparity at each pixel is selected as the index of the minimum cost in $S(\mathbf{p}, d)$:

$$D(\mathbf{p}) = \underset{d}{\operatorname{argmin}} S(\mathbf{p}, d), \quad (17)$$

Sub-pixel disparity estimation is implemented by fitting a parabola using neighboring costs:

$$D_{\mathbf{p}}^{\text{sub}} = D_{\mathbf{p}} - \frac{S(\mathbf{p}, D_{\mathbf{p}} + 1) - S(\mathbf{p}, D_{\mathbf{p}} - 1)}{2S(\mathbf{p}, D_{\mathbf{p}} + 1) + 2S(\mathbf{p}, D_{\mathbf{p}} - 1) - 4S(\mathbf{p}, D_{\mathbf{p}})}, \quad (18)$$

In addition, the disparity map of the reference image is obtained using $S(\mathbf{p}, d)$ by a diagonal search for the minimum [64]:

$$D_{\text{ref}}(\mathbf{p}) = \underset{d}{\operatorname{argmin}} S(\mathbf{p} + [d, 0]^T, d). \quad (19)$$

The disparity computation and sub-pixel disparity estimation are pixel-independent algorithms, which are only related to the aggregated cost of the current pixel. Specifically, the kernel program of OpenCL has $W \times H$ workgroups, each of which is divided into $D/2$ work items. Each workgroup is responsible for disparity computation and subpixel disparity estimation of an individual pixel. First, each work item performs two write operations to store the aggregated cost of the current pixel in the local memory. After synchronizing the work items of same workgroup, the reduction algorithm is used to rank the aggregated cost of the current pixel and obtain the disparity with the minimum cost. Then, one work item of the workgroup performs the sub-pixel disparity estimation of the current pixel.

After obtaining the disparity map from the reference and target speckle images, a modified Left-Right consistency check (L-R Check) is used to identify invalid pixels in the disparity map, including occluded, mismatched, and background areas:

$$D^{\text{sub}}(\mathbf{p}) = \begin{cases} -10, & S(\mathbf{p}, D(\mathbf{p})) > P_4, \\ -20, & x + D(\mathbf{p}) \leq 0 \parallel x + D(\mathbf{p}) > W, \\ -30, & D(\mathbf{p}) - D_{\text{ref}}(\mathbf{p} + D(\mathbf{p})) > 1, \end{cases} \quad (20)$$

where P_4 is a parameter related to $S(\mathbf{p}, d)$, which determines some pixels with larger aggregated costs as invalid background areas. According to the disparity value $D^{\text{sub}}(\mathbf{p})$, some pixels whose corresponding points are beyond the reference image are set as occlusion areas. Based on the

disparity maps of the reference and target images, some pixels with unequal disparity values are determined as mismatch points by L-R Check. Since the modified L-R Check is a pixel-independent algorithm, each work item is responsible for a single pixel.

For disparity optimization, the disparity map is first filtered by GPU-based median filtering with a 3×3 window, and converted to the depth map. The disparity and depth maps from the GPU are then transferred to the CPU. In the CPU environment, the 4-connected-based image segmentation method is exploited to process the disparity map to identify and remove segments with fewer pixels. The mismatched points in the disparity map are interpolated by selecting the second-lowest disparity value from their 8-neighborhood points.

There are several means to further improve the efficiency of the proposed OpenCL-based SGM algorithm shown in Fig. 4, as follows:

- (1) According to the 3D imaging principle of monocular 3D sensors, the reference image is a known speckle image at Z_{ref} and its census transform can be calculated in advance.
- (2) The local memory mechanism of OpenCL is widely used to improve computational efficiency in the census transform, cost calculation, cost aggregation, disparity calculation, sub-pixel disparity estimation, and median filtering.
- (3) The GPU-accelerated reduction algorithm is used to quickly find the minimum value of a set of data.
- (4) OpenCL can execute multiple queues for different operations simultaneously. While performing the census transform and cost calculation, another queue pre-computes the penalty parameter P_2 of the target image. Considering that the same aggregation cost is used in the disparity calculation of the reference image and the target image, for cost aggregation, it is more efficient to first launch three queues simultaneously to calculate the aggregated cost of three paths, and then invoke one queue to aggregate another path while averaging the total cost from all paths.

3. Experiments

To reveal the actual performance of the proposed monocular 3D sensor, some experiments are carried out, including precision analysis, experimental comparisons of different 3D sensors, and dynamic scenes measurement. In our sensor, the resolution of the IR camera is 640×480 , the baseline between the camera and the projector is about 35mm . First, a square LED with a wavelength of 940nm is used to uniformly illuminate the calibration board, and the calibration targets with different orientations at 16 different distances are continuously captured to obtain the camera intrinsic parameters as shown in Fig. 3. Considering the geometric parameters of the proposed 3D sensor, the absolute disparity range is suitably set to 64 pixels to measure scenes with a depth range of 0.3m to 3m .

In the proposed OpenCL-based SGM algorithm, census transform with a 5×5 window is adopted to obtain the initial matching cost, which ranges from 0 to 25. After the SGM-based cost aggregation, the aggregated cost ranges from 0 to 300. In order to enhance the accuracy of speckle matching, the preset thresholds of the parameters P_1 , P_3 , and P_4 are set as 30, 120, and 120 based on the matching quality metric after an exhaustive empirical search.

3.1. Precision analysis

To verify the feasibility of the proposed calibration method for monocular 3D sensors with the reference speckle image, five high-quality speckle reference images at different distances (400mm , 500mm , 600mm , 700mm , and 800mm) are collected to implement plane measurements. Precision analysis results are obtained by measuring the planar target at distances between 400mm and 1500mm . In this experiment, not only the 3D imaging accuracy of the monocular 3D sensor can be quantitatively evaluated, but also the best reference speckle image is selected.

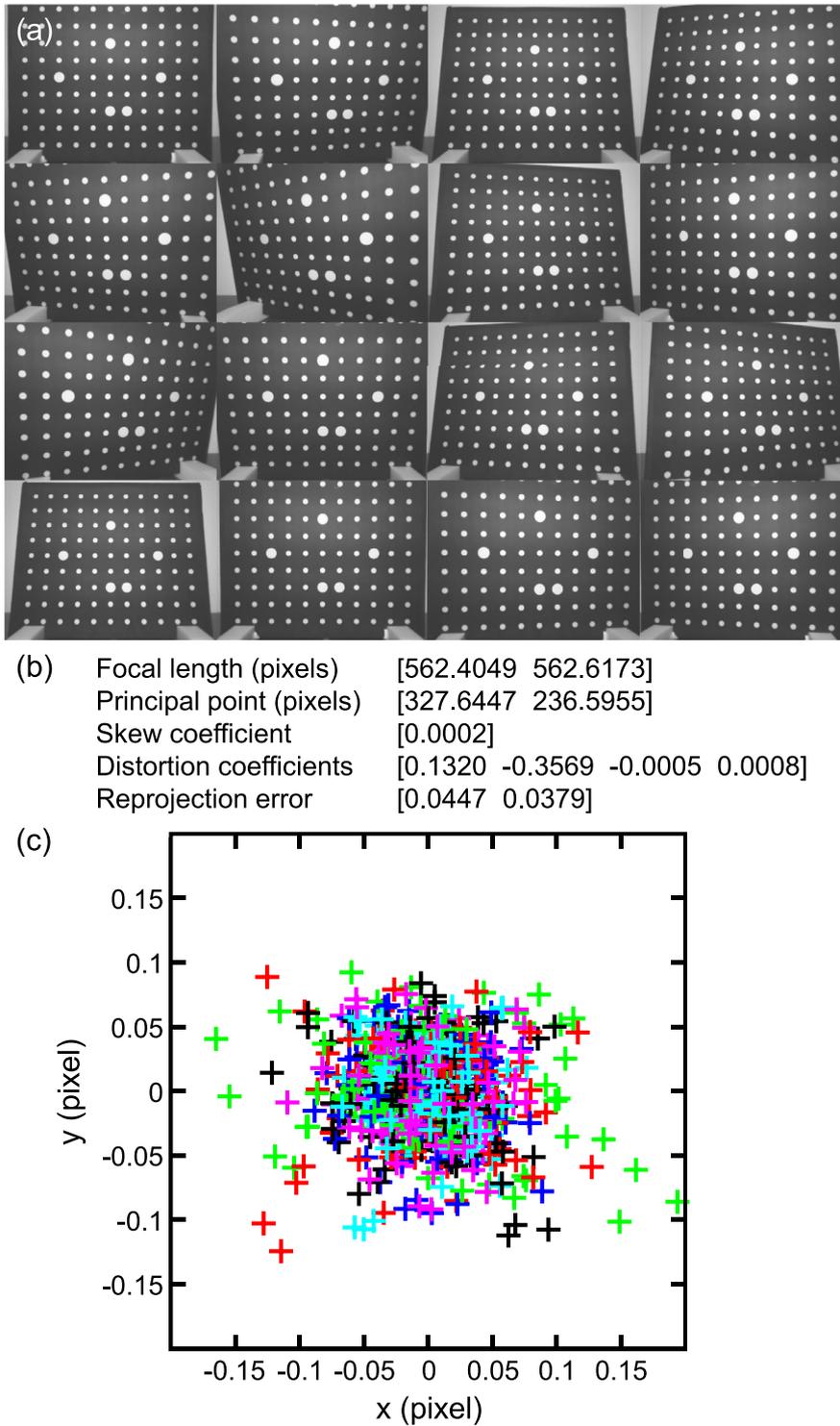


Fig. 3. Calibration results of IR camera. (a) 16 images of the calibration board with different poses. (b) The calibrated camera parameters. (c) Reprojection error distributions of the camera.

However, due to the limited disparity accuracy of stereo matching and the quantization error of the depth map, the measurement accuracy of planar targets is not only related to the 3D measurement algorithm, but also affected by the angle between the planar and the Z-axis of the 3D sensor [74]. As the measurement angle gradually increases, the 3D measurement error of the plane increases significantly in Fig. 4. Therefore, for each distance, it needs to continuously adjust the pose of the monocular 3D sensor to obtain accurate precision analysis results by multi-frame averaging of the plane fitting, as shown in Fig. 5. It is easy to find that the measurement precision of the planar target is slightly

different at distances between 400mm and 1500mm using different reference images, but their trend is the same, and the following conclusions can be drawn:

- (1) As the measurement distance increases, the measurement precision gradually deteriorates, which is determined by the triangulation of stereo vision.
- (2) If the 3D measurement is performed using the reference speckle image at Z_{ref} , the measurement accuracy of the planar target close to Z_{ref} is relatively high, which is determined by the measurement principle of monocular 3D sensors. According to

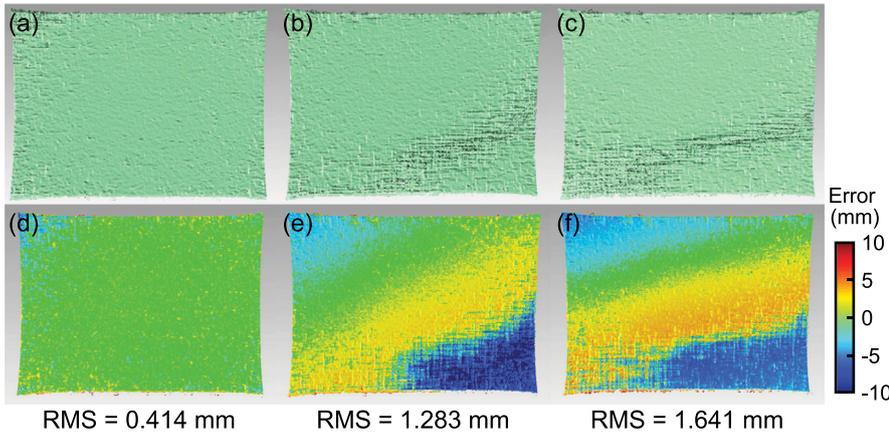


Fig. 4. The 3D measurement error of the plane increases significantly as the measurement angle gradually increases. (a)-(c) The 3D measurement results. (d)-(f) The 3D measurement errors of (a)-(c).

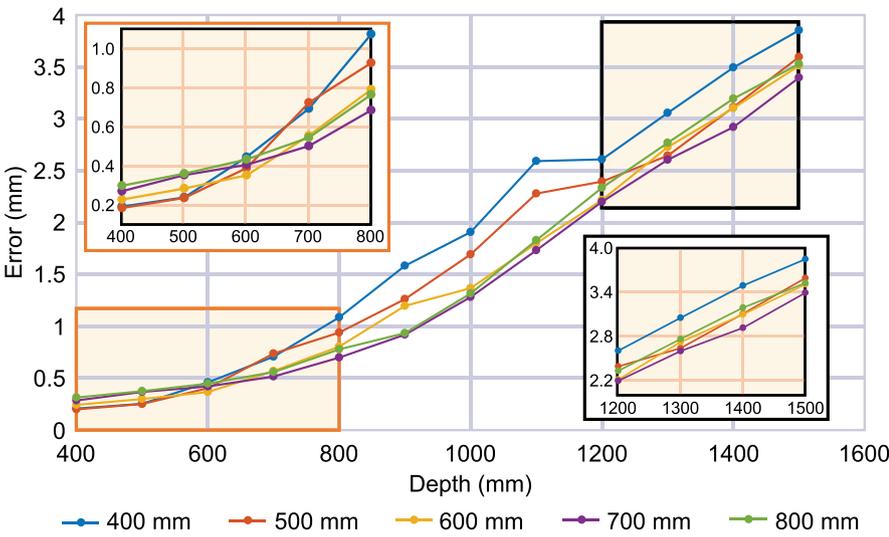


Fig. 5. Precision analysis of 3D measurement results for the planar target at distances between 400mm and 1500mm using reference speckle patterns at different distances.

Eq. (5), as the planar target is closer to the reference plane, the relative disparity d_{rel} is more accurate, thereby improving the measurement accuracy.

- (3) In Fig. 5, the measurement accuracy (within 700mm, 1000mm, and 1500mm) is less than 0.8mm, 2mm and 4mm. This experimental result indirectly proves the high quality of the reference speckle image, which confirms that the proposed calibration method for monocular 3D sensors with the reference speckle image is practical and feasible.
- (4) While quantitatively evaluating the 3D imaging accuracy of the monocular 3D sensor, the measurement accuracy at distances between 400mm and 1500mm is higher using the reference speckle image at 700mm, which is selected as the best reference speckle image.

3.2. Experimental comparison of 3D sensors

To verify the actual performance of our monocular 3D sensor, the experiments of precision analysis and complex scene measurement are carried out for comparison with some common 3D sensors based on speckle projection. These 3D sensors are classified into monocular 3D sensors and stereo 3D sensors in Fig. 6(a) and Table. 1. Due to the expensive overhead of dense and accurate depth sensing algorithms in mobile applications, it is worth noting that mainstream 3D sensors are generally equipped with ASIC to realize real-time data processing and transmission of 3D measurements, such as MX series chips of Orbbec and IMI series chips of HJIMI. However, these 3D sensors need to guarantee

Table 1
Parameters of 3D sensors based on speckle projection.

Parameter	The developed 3D sensor	HJIMI A200	Orbbec Dabai	RealSense D435
3D Imaging principle	Monocular	Monocular	Binocular	Binocular
Algorithm platform	GPU	ASIC	ASIC	ASIC
Baseline	35mm	55mm	40mm	50mm
Resolution	640 × 480	640 × 400	640 × 400	640 × 480
Minimum unit	0.01mm	1mm	0.1mm	0.1mm

efficient data transmission from hardware platforms to mobile devices at the cost of 3D measurement accuracy. As shown in Table. 1, HJIMI A200 stably outputs the depth map with a minimum unit of 1mm at 30 FPS, and the minimum unit of the depth map output by Orbbec Dabai and RealSense D435 is 0.1mm.

Likewise, in order to estimate the 3D imaging accuracy quantitatively, 3D measurement results for the planar target at distances between 400mm and 1500mm using different 3D sensors in Fig. 6(b)-(c). For 3D measurement results at 1000mm in Fig. 6(b), as a monocular 3D sensor with a wider baseline, it can be found that the plane measurement results provided by HJIMI A200 are relatively rough, which may be related to the cost aggregation strategy in the used stereo matching algorithm. The point cloud smoothing method used in Orbbec Dabai may lead to obvious step-like measurement errors in Fig. 6(b). For the measurement results of RealSense D435, it suffers from severe measurement errors which limits its 3D imaging quality and precludes the re-

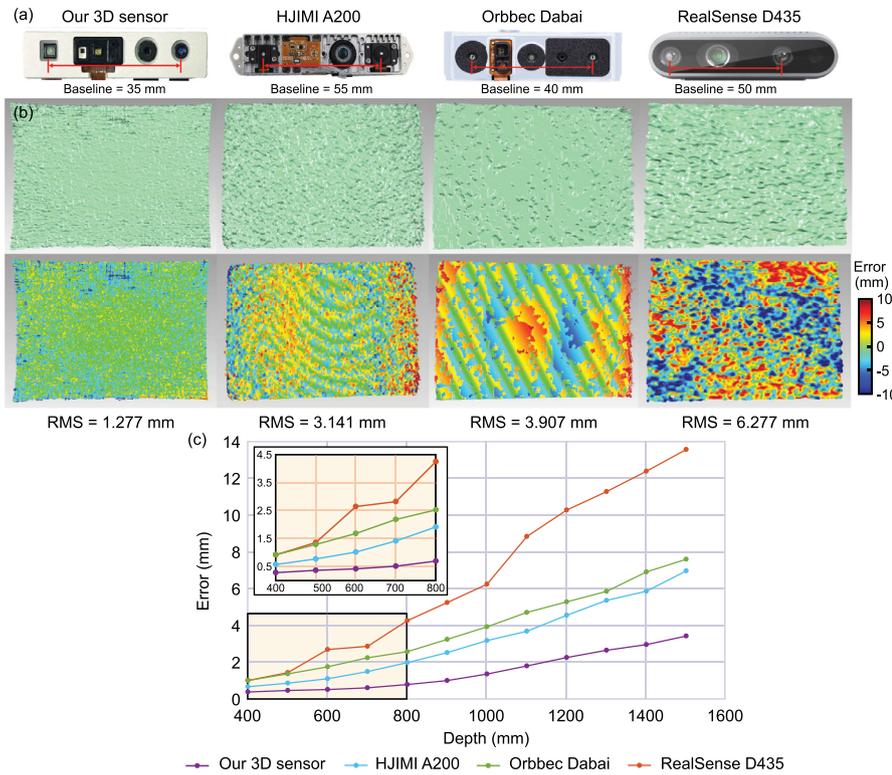


Fig. 6. Precision analysis of 3D measurement results for the planar target at distances between 400mm and 1500mm using different 3D sensors. (a) Photograph of different 3D sensors based on speckle projection. (b) 3D measurement results of the planar target at 1000mm. (c) Precision analysis results for the planar target at distances between 400mm and 1500mm.

Table 2

Runtime of the optimized SGM algorithm based on OpenCL acceleration on different GPUs.

Runtime(ms)	GTX 1060	ARM Mail G52
Stage 1 (GPU)	0.39	3.83 ms
Stage 2 (GPU)	2.88	15.67
Stage 3 (GPU)	4.51	29.36
Stage 4 (CPU)	5.48	18.55
Total	13.26	67.41
Frame rate	75.41 FPS	14.83 FPS

covery of the fine details of the measured plane. In contrast, benefited from the proposed reference plane calibration method and the optimized SGM algorithm based on OpenCL acceleration, our monocular 3D sensor can run on a desktop computer (GTX 1060) and the Android tablet PC (UNISOC T618 with ARM Mali-G52) to obtain 3D measurement results losslessly, enabling high-accuracy and real-time 3D imaging. The runtime on different GPUs is shown in Table 2, which achieves fast monocular matching at speeds of 75.41 FPS and 14.83 FPS. These experimental results verify that the proposed 3D sensor can significantly increase the measurement accuracy to 2 – 3 times that of other 3D sensors, the measurement accuracy of our monocular 3D sensor (within 800mm, 1000mm, and 1500mm) are about 0.7mm, 1.5mm and 3.5mm in Fig. 6(c).

Measuring the target scenes with ridged, complex, or discontinuous surfaces is a challenging task for 3D sensors based on speckle projection. To verify the reliability of these 3D sensors for scanning the challenging surfaces, a complex indoor scene is measured within a large 3D space volume of 0.8m(width) × 0.5m(height) × 1m(depth) as shown in Fig. 7(a), the corresponding 3D reconstruction results output by different 3D sensors are shown in Fig. 7(b)–(e). In terms of the accuracy of the measurement results, the depth map output by RealSenseD435 has the worst quality and edge fattening in Fig. 7(e). There have been over-smoothing and obvious distorted measurement errors in measuring the teacup at 500mm, and severe corrugated measurement errors in measuring the wall

at 900mm, making it impossible to measure the power line on the wall in Fig. 7(i). In the measurement results of Orbbec Dabai, due to the common occlusion problem in stereo vision systems, it cannot accurately measure the power line on the wall, but the measurement results of the soldier dolls and a teacup have the best quality in Figs. 7(d) and (h). Based on different 3D imaging principles, the stereo 3D sensor matches the left and right images of the same scene, while the monocular 3D sensor matches the target scene image with the reference plane image. Therefore, stereo 3D sensors are good at measuring complex scenes in Fig. 7, while monocular 3D sensors are good at measuring planar targets in Figs. 6(b)–(c). HJIMI A200 with a wider baseline (55mm) outputs the depth map with the lowest completeness and the coarse measurement results of the teacup, but reconstructs the 3D point clouds with high quality of the power line in Figs. 7(c) and (g). In contrast, our 3D sensor provides a more complete and high-quality reconstruction of the power line and the teacup in Figs. 7(b) and (f), which demonstrates its reliability of high-accuracy 3D imaging for measuring complex scenes.

3.3. Dynamic scene measurement

Next, our 3D sensor is applied to record a dynamic scene to further validate its capability of real-time 3D shape measurement in Fig. 8. The scene consisted of two face masks and a moving hand mostly. During this experiment, the exposure time of cameras is set at 33ms to capture speckle images for achieving 3D reconstruction at 30 FPS. Figs. 8(a)–(b) show representative 2D camera images and corresponding color-coded 2.5D depth reconstructions at different time points. For the corresponding enlarged local details shown in Fig. 8(c), the 3D point clouds of face masks and the hand are presented to illustrate that our method can achieve robust 3D shape measurement for objects with complex surfaces and geometric discontinuities. The whole 3D measurement results can be referred to in Visualization1, and it can be found that complex surfaces of the dynamic scene are correctly reconstructed with high quality, verifying the reliability of the proposed monocular 3D sensor to perform accurate 3D measurement with high completeness dynamically.

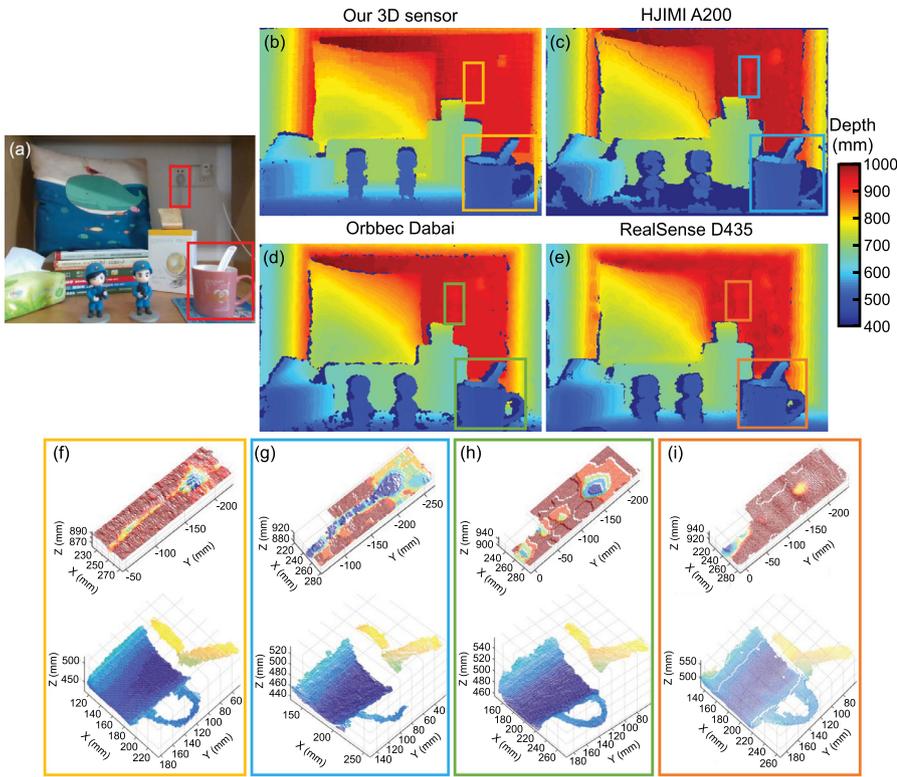


Fig. 7. Comparison of 3D measurement results for complex scenes. (a) Photograph of a complex indoor scene. (b)-(e) The depth maps output by our 3D sensor, HJIMI A200, Orbbec Dabai, and RealSense D435. (f)-(i) The 3D point clouds of the enlarged local details in (b)-(e).

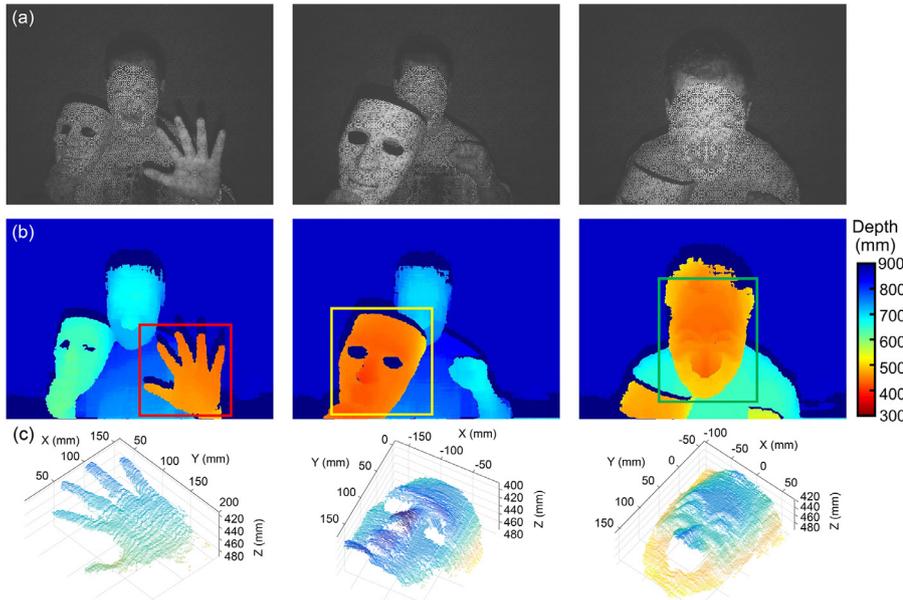


Fig. 8. The 3D reconstruction results for a dynamic scene (*Visualization1*). (a)-(b) 2D camera images and corresponding color-coded depth reconstructions at different times. (c) The 3D point clouds of the enlarged local details in (b).

4. Conclusions and discussion

In this work, we proposed a real-time and accurate monocular 3D sensor using the reference plane calibration and an optimized SGM based on OpenCL acceleration. In terms of hardware, benefited from the customized VCSEL-based speckle projection scheme, the IR dot projector can project a total of 35 thousand clear and uniform speckle spots to the measured scene, enabling low-cost and high-quality speckle projection. According to the 3D imaging principle of monocular 3D sensors above, it revealed an important relationship between the quality of the reference speckle image and the measurement accuracy. Then, a calibration

method is proposed to estimate the extrinsic parameters between the reference plane and the IR camera for acquiring the high-quality reference speckle image at Z_{ref} . In addition, considering that 3D sensors are generally mounted on mobile devices, an optimized semi-global matching (SGM) algorithm using OpenCL acceleration is proposed to obtain efficient, dense, and accurate matching results, enabling high-quality 3D reconstruction dynamically. Concretely, in the OpenCL environment, the local memory mechanism is widely used to improve computational efficiency in the census transform, cost calculation, cost aggregation, disparity calculation, sub-pixel disparity estimation, and median filtering. Since OpenCL can execute multiple queues for different operations

simultaneously, the Census-based cost calculation and the penalty P_2 for the target image will first be computed simultaneously in two different queues, respectively. For cost aggregation, considering that the same aggregation cost is used in the disparity calculation of the reference image and the target image, an efficient program is implemented to first launch three queues to calculate the aggregated cost of three paths simultaneously, and then invoke one queue to aggregate another path while averaging the total cost from all paths. Finally, some post-processing operations are implemented to identify and correct the mismatch points and the occlusion regions, guaranteeing the high completeness of 3D measurement results. The experimental results of precision analysis proved that the proposed 3D sensor can significantly increase the measurement accuracy to 2 – 3 times that of other 3D sensors, our measurement accuracy (within 800mm, 1000mm, and 1500mm) is less than 0.7mm, 1.5mm and 3.5mm, which confirms that the proposed calibration method for the reference speckle image is practical and feasible. For complex scene measurement, the experimental comparison of different 3D sensors illustrated that our 3D sensor can provide a more complete and high-quality reconstruction of the power line and the teacup at different distances. The dynamic measurement experiment has verified the success of the proposed method in its ability to effectively achieve fast and accurate 3D imaging within the large measurement range of 0.8m (length) \times 0.5m (width) \times 1m (depth) at 75 FPS on GTX 1060 and 15 FPS on ARM Mail G52(mobile platform).

It should be discussed here that the 3D sensor developed in this paper can achieve more accurate 3D measurement compared with others, but it cannot meet the requirements of 3D measurement with the sub-millimeter accuracy in some applications. How to improve the matching accuracy while maintaining the miniaturization of 3D sensors is still a problem to be solved. Recently, some 3D measurement systems based on multi-frame speckle projection have been proposed, which can greatly improve the measurement accuracy by utilizing the spatial-temporal stereo matching algorithm, but it is no longer miniaturized and portable. Based on the above analysis, we will explore other methods to design a single-shot SPP system with higher performance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Wei Yin: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Lu Cao:** Software, Formal analysis. **Hang Zhao:** Data curation, Validation. **Yan Hu:** Writing – review & editing. **Shijie Feng:** Writing – review & editing. **Xiaolei Zhang:** Writing – review & editing. **Detong Shen:** Software, Formal analysis. **Huai Wang:** Writing – review & editing. **Qian Chen:** Formal analysis, Writing – review & editing, Funding acquisition. **Chao Zuo:** Formal analysis, Writing – review & editing, Funding acquisition.

Acknowledgments

This work was supported by National Key Research and Development Program of China (2022YFB2804603, 2022YFB2804604), National Natural Science Foundation of China (62075096, 62205147, U21B2033), China Postdoctoral Science Foundation (2022M711630, 2022M721619), Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB254), The Leading Technology of Jiangsu Basic Research Plan (BK20192003), The “333 Engineering” Research Project of Jiangsu Province (BRA2016407), The Jiangsu Provincial “One belt and one road” innovation cooperation project (BZ2020007), Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense

(JSGP202105), Fundamental Research Funds for the Central Universities (30921011208, 30919011222, 30920032101, 30922010405), and National Major Scientific Instrument Development Project (62227818).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2023.107536.

References

- [1] Ganapathi V, Plagemann C, Koller D, Thrun S. Real time motion capture using a single time-of-flight camera. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010. p. 755–62.
- [2] Kolb A, Barth E, Koch R, Larsen R. Time-of-flight cameras in computer graphics. In: Computer Graphics Forum. Wiley Online Library; 2010. p. 141–59.
- [3] Hansard M, Lee S, Choi O, Horaud RP. Time-of-flight cameras: principles, methods and applications. Springer Science & Business Media; 2012.
- [4] Sun J, Zheng N-N, Shum H-Y. Stereo matching using belief propagation. IEEE Trans Pattern Anal Mach Intell 2003;25(7):787–800.
- [5] Hirschmüller H. Stereo processing by semiglobal matching and mutual information. IEEE Trans Pattern Anal Mach Intell 2007;30(2):328–41.
- [6] Lazaros N, Sirakoulis GC, Gasteratos A. Review of stereo vision algorithms: from software to hardware. Int J Optomechatronics 2008;2(4):435–62.
- [7] Jin S, Cho J, Dai Pham X, Lee KM, Park S-K, Kim M, et al. Fpga design and implementation of a real-time stereo vision system. IEEE Trans Circuits Syst Video Technol 2009;20(1):15–26.
- [8] Geiger A, Roser M, Urtasun R. Efficient large-scale stereo matching. In: Asian Conference on Computer Vision. Springer; 2010. p. 25–38.
- [9] Bleyer M, Rhemann C, Rother C. Patchmatch stereo-stereo matching with slanted support windows. In: Bmvc; 2011. p. 1–11.
- [10] Seitz SM, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2006. p. 519–28.
- [11] Goesele M, Snavely N, Curless B, Hoppe H, Seitz SM. Multi-view stereo for community photo collections. In: 2007 IEEE 11th International Conference on Computer Vision. IEEE; 2007. p. 1–8.
- [12] Zhang Q, Su X. High-speed optical measurement for the drumhead vibration. Opt Express 2005;13(8):3110–16.
- [13] Gorthi SS, Rastogi P. Fringe projection techniques: whither we are? Opt Laser Eng 2010;48:133–40.
- [14] Feng S, Zhang L, Zuo C, Tao T, Chen Q, Gu G. High dynamic range 3d measurements with fringe projection profilometry: a review. Mea Sci Technol 2018;29(12):122001.
- [15] Zhang S. Absolute phase retrieval methods for digital fringe projection profilometry: a review. Opt Laser Eng 2018;107:28–37.
- [16] Yin W, Feng S, Tao T, Huang L, Zhang S, Chen Q, Zuo C. Calibration method for panoramic 3d shape measurement with plane mirrors. Opt Express 2019;27(25):36538–50.
- [17] Guo W, Wu Z, Li Y, Liu Y, Zhang Q. Real-time 3d shape measurement with dual-frequency composite grating and motion-induced error reduction. Opt Express 2020;28(18):26882–97.
- [18] Wu Z, Guo W, Zhang Q. High-speed three-dimensional shape measurement based on shifting gray-code light. Opt Express 2019;27(16):22631–44.
- [19] Wolcott RW, Eustice RM. Visual localization within lidar maps for automated urban driving. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE; 2014. p. 176–83.
- [20] Hirschmüller H, Innocent PR, Garibaldi J. Real-time correlation-based stereo vision with reduced border errors. Int J Comput Vis 2002;47(1–3):229–46.
- [21] Hirschmüller H, Scharstein D. Evaluation of cost functions for stereo matching. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2007. p. 1–8.
- [22] Hirschmüller H, Scharstein D. Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans Pattern Anal Mach Intell 2008;31(9):1582–99.
- [23] Scharstein D, Hirschmüller H, Kitajima Y, Krathwohl G, Nešić N, Wang X, Westling P. High-resolution stereo datasets with subpixel-accurate ground truth. In: German Conference on Pattern Recognition. Springer; 2014. p. 31–42.
- [24] Salvi J, Pages J, Batlle J. Pattern codification strategies in structured light systems. Pattern Recognit 2004;37(4):827–49.
- [25] Zhang Z. Review of single-shot 3d shape measurement by phase calculation-based fringe projection techniques. Opt Laser Eng 2012;50:1097–106.
- [26] Zuo C, Tao T, Feng S, Huang L, Asundi A, Chen Q. Micro fourier transform profilometry (μ ftp): 3d shape measurement at 10,000 frames per second. Opt Laser Eng 2018;102:70–91.
- [27] Zuo C, Qian J, Feng S, Yin W, Li Y, Fan P, Han J, Qian K, Chen Q. Deep learning in optical metrology: a review. Light: Sci Appl 2022;11(1):1–54.
- [28] Zhang S. High-speed 3d shape measurement with structured light methods: a review. Opt Laser Eng 2018;106:119–31.
- [29] Heist S, Dietrich P, Landmann M, Kühmstedt P, Notni G, Tünnemann A. Gobo projection for 3d measurements at highest frame rates: a performance analysis. Light: Sci Appl 2018;7(1):1–13.
- [30] Yin W, Zuo C, Feng S, Tao T, Hu Y, Huang L, Ma J, Chen Q. High-speed three-dimensional shape measurement using geometry-constraint-based number-theoretical phase unwrapping. Opt Laser Eng 2019;115:21–31.

- [31] Wu Z, Guo W, Zhang Q. Two-frequency phase-shifting method vs. gray-coded-based method in dynamic fringe projection profilometry: a comparative review. *Opt Laser Eng* 2022;153:106995.
- [32] Schaffer M, Grosse M, Kowarschik R. High-speed pattern projection for three-dimensional shape measurement using laser speckles. *Appl Opt* 2010;49(18):3622–9.
- [33] Schaffer M, Grosse M, Harendt B, Kowarschik R. High-speed three-dimensional shape measurements of objects with laser speckles and acousto-optical deflection. *Opt Lett* 2011;36(16):3097–9.
- [34] Zhou P, Zhu J, Jing H. Optical 3-d surface reconstruction with color binary speckle pattern encoding. *Opt Express* 2018;26(3):3452–65.
- [35] Yin W, Hu Y, Feng S, Huang L, Kemao Q, Chen Q, Zuo C. Single-shot 3d shape measurement using an end-to-end stereo matching network for speckle projection profilometry. *Opt Express* 2021;29(9):13388–407.
- [36] Su X, Chen W. Fourier transform profilometry: a review. *Opt Laser Eng* 2001;35:263–84.
- [37] Feng S, Chen Q, Gu G, Tao T, Zhang L, Hu Y, Yin W, Zuo C. Fringe pattern analysis using deep learning. *Adv Photonics* 2019;1(2):025001.
- [38] Yin W, Zhong J, Feng S, Tao T, Han J, Huang L, Chen Q, Zuo C. Composite deep learning framework for absolute 3d shape measurement based on single fringe phase retrieval and speckle correlation. *J Phys: Photon* 2020;2(4):045009.
- [39] Zuo C, Feng S, Huang L, Tao T, Yin W, Chen Q. Phase shifting algorithms for fringe projection profilometry: a review. *Opt Laser Eng* 2018;109:23–59.
- [40] Feng S, Zuo C, Zhang L, Yin W, Chen Q. Generalized framework for non-sinusoidal fringe analysis using deep learning. *Photonics Res* 2021;9(6):1084–98.
- [41] Su X, Chen W. Reliability-guided phase unwrapping algorithm: a review. *Opt Laser Eng* 2004;42:245–61.
- [42] Zhao M, Huang L, Zhang Q, Su X, Asundi A, Kemao Q. Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies. *Appl Opt* 2011;50(33):6214–24.
- [43] Wu Z, Guo W, Zhang Q, Wang H, Li X, Chen Z. Time-overlapping structured-light projection: high performance on 3d shape measurement for complex dynamic scenes. *Opt Express* 2022;30(13):22467–86.
- [44] Wang Y, Zhang S. Novel phase-coding method for absolute phase retrieval. *Opt Lett* 2012;37(11):2067–9.
- [45] Zhong K, Li Z, Shi Y, Wang C, Lei Y. Fast phase measurement profilometry for arbitrary shape objects without phase unwrapping. *Opt Laser Eng* 2013;51:1213–1222.
- [46] Zuo C, Huang L, Zhang M, Chen Q, Asundi A. Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review. *Opt Laser Eng* 2016;85:84–103.
- [47] Tao T, Chen Q, Feng S, Hu Y, Zhang M, Zuo C. High-precision real-time 3d shape measurement based on a quad-camera system. *J Opt* 2017;20(1):014009.
- [48] Cai Z, Liu X, Chen Z, Tang Q, Gao BZ, Pedrini G, Osten W, Peng X. Light-field-based absolute phase unwrapping. *Opt Lett* 2018;43(23):5717–20.
- [49] Yin W, Chen Q, Feng S, Tao T, Huang L, Trusiak M, Asundi A, Zuo C. Temporal phase unwrapping using deep learning. *Sci Rep* 2019;9(1):1–12.
- [50] Yin W, Feng S, Tao T, Huang L, Trusiak M, Chen Q, Zuo C. High-speed 3d shape measurement using the optimized composite fringe patterns and stereo-assisted structured light system. *Opt Express* 2019;27(3):2411–31.
- [51] Su X, Zhang Q. Dynamic 3-d shape measurement method: a review. *Opt Laser Eng* 2010;48:191–204.
- [52] Feng S, Zuo C, Tao T, Hu Y, Zhang M, Chen Q, Gu G. Robust dynamic 3-d measurements with motion-compensated phase-shifting profilometry. *Opt Laser Eng* 2018;103:127–38.
- [53] Ito M, Ishii A. A three-level checkerboard pattern (tcp) projection method for curved surface measurement. *Pattern Recognit* 1995;28(1):27–40.
- [54] Maruyama M, Abe S. Range sensing by projecting multiple slits with random cuts. *IEEE Trans Pattern Anal Mach Intell* 1993;15(6):647–51.
- [55] Boyer KL, Kak AC. Color-encoded structured light for rapid active ranging. *IEEE Trans Pattern Anal Mach Intell* 1987(1):14–28.
- [56] Zhang L, Curless B, Seitz SM. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: *First International Symposium on 3D Data Processing Visualization and Transmission*. IEEE; 2002. p. 24–36.
- [57] Pagès J, Salvi J, Collewet C, Forest J. Optimised de bruijn patterns for one-shot shape acquisition. *Image Vis Comput* 2005;23(8):707–20.
- [58] Morita H, Yajima K, Sakata S. Reconstruction of surfaces of 3-d objects by m-array pattern projection method. In: *Second International Conference on Computer Vision*. IEEE; 1988. p. 468–73.
- [59] Gu F, Song Z, Zhao Z. Single-shot structured light sensor for 3d dense and dynamic reconstruction. *Sensors* 2020;20(4):1094.
- [60] Li Y, Zheng S, Wang X, Ma H. An efficient photogrammetric stereo matching method for high-resolution images. *Comput Geosci* 2016;97:58–66.
- [61] Wang L, Gong M, Gong M, Yang R. How far can we go with local optimization in real-time stereo matching. In: *Third International Symposium on 3D Data Processing, Visualization, and Transmission*. IEEE; 2006. p. 129–36.
- [62] Gallup D, Frahm J-M, Mordohai P, Yang Q, Pollefeys M. Real-time plane-sweeping stereo with multiple sweeping directions. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2007. p. 1–8.
- [63] Woetzel J, Koch R. Real-time multi-stereo depth estimation on gpu with approximate discontinuity handling. In: *1st European Conference on Visual Media Production*; 2004.
- [64] Ernst I, Hirschmüller H. Mutual information based semi-global stereo matching on the gpu. In: *International Symposium on Visual Computing*. Springer; 2008. p. 228–39.
- [65] Banz C, Blume H, Pirsch P. Real-time semi-global matching disparity estimation on the gpu. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE; 2011. p. 514–21.
- [66] Banz C, Hesselbarth S, Flatt H, Blume H, Pirsch P. Real-time stereo vision system using semi-global matching disparity estimation: Architecture and fpga-implementation. In: *2010 International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation*. IEEE; 2010. p. 93–101.
- [67] Hernandez-Juarez D, Chacón A, Espinosa A, Vázquez D, Moure JC, López AM. Embedded real-time stereo estimation via semi-global matching on the gpu. *Procedia Comput Sci* 2016;80:143–53.
- [68] Stone JE, Gohara D, Shi G. Opencl: a parallel programming standard for heterogeneous computing systems. *Comput Sci Eng* 2010;12(3):66–73.
- [69] Morinaga M, Gu X, Shimura K, Nakahama M, Matsutani A, Koyama F. Vcsel amplifier dot projector with folded-path slow-light waveguide for 3d depth sensing. In: *2018 IEEE International Semiconductor Laser Conference (ISLC)*. IEEE; 2018. p. 1–2.
- [70] Morinaga M, Gu X, Shimura K, Matsutani A, Koyama F. Compact dot projector based on folded path vcsel amplifier for structured light sensing. In: *2019 Conference on Lasers and Electro-Optics (CLEO)*. IEEE; 2019. p. 1–2.
- [71] Hartley R, Zisserman A. *Multiple view geometry in computer vision*. Cambridge University; 2003.
- [72] Dal Mutto C, Zanuttigh P, Cortelazzo GM. *Time-of-flight cameras and microsoft kinecttm*. Springer Science & Business Media; 2012.
- [73] Fanello SR, Rhemann C, Tankovich V, Kowdle A, Escolano SO, Kim D, Izadi S. Hyperdepth: Learning depth from structured light without matching. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*; 2016. p. 5441–50.
- [74] Nguyen CV, Izadi S, Lovell D. Modeling kinect sensor noise for improved 3d reconstruction and tracking. In: *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*. IEEE; 2012. p. 524–30.