Article

Spatiotemporal Correlation-Based Accurate 3D Face Imaging Using Speckle Projection and Real-Time Improvement

Wei Xiong ⓘ**, Hongyu Yang, Pei Zhou** *ⓘ**, Keren Fu and Jiangping Zhu**

College of Computer Science, Sichuan University, Chengdu 610064, China; xiongweib@scu.edu.cn (W.X.); yanghongyu@scu.edu.cn (H.Y.); fkrsuper@scu.edu.cn (K.F.); zjp16@scu.edu.cn (J.Z.)
* Correspondence: michille78@163.com

Abstract: The reconstruction of 3D face data is widely used in the fields of biometric recognition and virtual reality. However, the rapid acquisition of 3D data is plagued by reconstruction accuracy, slow speed, excessive scenes and contemporary reconstruction-technology. To solve this problem, an accurate 3D face-imaging implementation framework based on coarse-to-fine spatiotemporal correlation is designed, improving the spatiotemporal correlation stereo matching process and accelerating the processing using a spatiotemporal box filter. The reliability of the reconstruction parameters is further verified in order to resolve the contention between the measurement accuracy and time cost. A binocular 3D data acquisition device with a rotary speckle projector is used to continuously and synchronously acquire an infrared speckle stereo image sequence for reconstructing an accurate 3D face model. Based on the face mask data obtained by the high-precision industrial 3D scanner, the relationship between the number of projected speckle patterns, the matching window size, the reconstruction accuracy and the time cost is quantitatively analysed. An optimal combination of parameters is used to achieve a balance between reconstruction speed and accuracy. Thus, to overcome the problem of a long acquisition time caused by the switching of the rotary speckle pattern, a compact 3D face acquisition device using a fixed three-speckle projector is designed. Using the optimal combination parameters of the three speckles, the parallel pipeline strategy is adopted in each core processing unit to maximise system resource utilisation and data throughput. The most time-consuming spatiotemporal correlation stereo matching activity was accelerated by the graphical processing unit. The results show that the system achieves real-time image acquisition, as well as 3D face reconstruction, while maintaining acceptable systematic precision.

Keywords: three-dimensional face reconstruction; spatiotemporal correlation; speckle projection; stereo matching

1. Introduction

In the field of computer vision and computer graphics, acquiring, modelling and synthesising three-dimensional (3D) human faces has become an active research topic. In particular, 3D face models have been used in various situations, such as medical plastic surgery [1], 3D face recognition [2], entertainment [3] and artistic rendering [4]. Hence, 3D face reconstruction has attracted widespread attention.

Existing image-based 3D face reconstruction methods have two main development directions. One emerges from the perspective of object measurement, which usually requires special image acquisition equipment. The technology includes multiview stereo vision [5], structured light [6] and time-of-flight [7]. Predicting a 3D face model from a single image is another important research area. This method is driven by the prior data of the face, which constructs a 3D model based on a learning method. The 3D deformable model [8], the "shape-from-shading" method [9,10], and the convolutional neural network (CNN) regression [11–14] are the most common techniques.

Shape-measurement technology based on structured light encoding [6] has been widely used in the field of 3D face reconstruction. There are two mainstream structured

light-encoding methods: fringe projection profilometry (FPP) [15] and speckle projection profilometry (SPP) [16]. In the FPP, the projector projects a series of fringe-coded patterns onto the measurement target. The fringe image modulated by the measured object is synchronously captured by a stereo camera and is processed using various phase recovery techniques, such as phase-shift profilometry [17], to obtain absolute phase information. Using these methods, it is necessary to perform phase unwrapping to eliminate phase ambiguity and convert the wrapped phase to the absolute phase.

In SPP, a variety of structured light-pattern encoding strategies (e.g., non-formal codification [16] and De Bruijn sequences [18]) have been proposed to increase or enrich the surface texture information of the measured object, which can avoid the matching ambiguity problem caused by binocular stereo vision [19,20]. Therefore, the reliability and accuracy of 3D object reconstruction results can be improved. Among these, research on speckle encoding structured light has been very extensive.

Compared with sinusoidal FPP in measurement technology, the 3D measurement system based on SPP does not require the harsh conditions that are required by the precise phase shift, and only a low-cost speckle-pattern imaging system is required to obtain excellent performance of 3D reconstruction results [21]. Many companies have introduced 3D sensors based on diffractive optical element (DOE) [22] laser illumination, such as RealSense [23] and Kinect V1 [24], which have the characteristics of small size, low power consumption, and high integration. Gu et al. [25] used a semi-global matching (SGM) algorithm with DOE projection to improve the accuracy of stereo matching and achieved dense dynamic scene reconstruction. However, owing to the limited number of projection speckles, the spatial resolution and measurement accuracy of the 3D reconstruction data are relatively low, making them unsuitable for special applications, such as 3D face recognition. Therefore, there is a developing trend towards projecting high-density speckle encoding patterns in order to improve the precision of 3D reconstruction.

In many applications, to reduce the motion blur caused by the movement of the measured object, only one speckle-coded image is projected onto the target to reduce the acquisition time. Khan et al. [26] designed a system that adaptively changed the speckle size to the optimal size required by the algorithm, according to the distance of the object. Various pre-processing methods were used to reduce the spread of speckles in order to increase the density of the 3D point cloud. Guo et al. [27] presented an automatic and rapid whole-body 3D measurement system based on multi-node 3D sensors using speckle projection. Yin et al. [28] proposed a single-shot 3D shape-measurement method using an end-to-end stereo-matching network with speckle projection. A high-precision absolute phase map was obtained by combining a phase-shift profile measurement and a time-phase unwrapping technology in order to construct a high-quality disparity map for training the network model, and the disparity image was inferred and predicted by the speckle matching network. Zhou et al. [29] proposed a high-precision 3D surface profile measurement scheme that projected only a single-shot colour binary speckle pattern and a spatiotemporal correlation matching algorithm, which was applied to the measurement of dynamic and static objects.

Existing speckle-encoding structured light 3D measurement methods can be divided into two types [30]: temporal- and spatial-domain. Temporal-domain encoding structured light encodes the entire image along the time axis by successively projecting multiple pre-set patterns. The measurement accuracy of the system depends heavily on the number of projection patterns. The spatial-domain method only collects a set of stereo image pairs and makes full use of the correlation information of adjacent pixels where the accuracy is susceptible to the steep surface of the measured object.

The 3D measurement of the single-shot speckle method mentioned earlier mostly adopts spatial domain correlation methods, and its calculation accuracy is limited to a certain extent. To improve the measurement accuracy of speckle-structured light, speckle-coded image-based spatiotemporal correlation computation has become a research hotspot. Ishii et al. [30] proposed a pattern system based on temporal and spatial encoding, which

can measure slow-moving objects. Meanwhile, temporal and spatial encoding schemes are designed according to different speed and accuracy requirements. Große et al. [31] comprehensively utilised the characteristics of two methods to find a pair of homologous points through the temporal correlation of the grey value sequence, where spatial correlation was used to verify the correctness of the homologous points. Under equivalent accuracy, the model can reduce the spatial correlation area and make it suitable for the surface reconstruction of objects with large fluctuations. Mainly, it reduces the number of images required for 3D reconstruction, which makes it possible to reconstruct dynamic scenes. Harendt et al. [32] proposed a weighted spatiotemporal correlation method to reconstruct static or moving objects, and the weight value was used to adjust the spatial and temporal parameters of the matching region. Tang et al. [33] evaluated the relationship between the size of the correlation area and the number of projection patterns. Experiments showed that selecting an appropriate window and a reasonable number of speckle patterns produced more accurate results. However, they only analysed the dumbbell gauges, and their results were not universally generalisable for face modelling. The 3D measurement system proposed by Zhou et al. [34] consists of an expensive industrial camera and a motor-driven projection device with high-density speckles to ensure the integrity and accuracy of the 3D face. However, the device is too large and expensive for routine use, and the matching procedure is not optimised. Its computation speed is slow, and it cannot meet the application requirements of a portable 3D face recognition system. Fu et al. [35] projected a visible-light speckle pattern on the human face and utilised a spatiotemporal stereo matching algorithm to achieve fast 3D face reconstruction. However, it only projected three frames of speckle patterns, and the selected parameters were not optimally evaluated, which limits reconstruction accuracy.

To quickly obtain a high-precision real 3D face model, the authors conducted research on measurement methods and parallel acceleration processing strategies. The contribution of this study is to the following three aspects:

1. A framework for implementing a high-precision 3D face imaging technique with a coarse-to-fine spatiotemporal correlation is designed. A spatiotemporal box filter is used to improve the spatiotemporal correlation stereo matching process and accelerate the calculation.

2. The relationship between the number of projected speckle patterns, matching window size, reconstruction accuracy and time cost in the spatiotemporal correlation 3D face imaging method is illuminated through experiments. The optimal combination of parameters to meet the requirements of a balanced strategy of reconstruction speed and accuracy is obtained. A reference for real-time high-precision 3D reconstruction is provided in the next step.

3. Based on previous research, a compact 3D face acquisition system using a fixed three-speckle projection method is designed. Using the optimal combination parameters for three-speckle stereo images, a parallel pipeline strategy is adopted between each core functional unit to maximise system resource utilisation and data throughput. The most time-consuming spatiotemporal correlation stereo-matching process is accelerated by the graphical processing unit (GPU). Experiments verify that the entire system achieves the goals of real-time data acquisition and real-time 3D reconstruction while maintaining high-precision modelling.

2. Methodology

2.1. Stereo Matching

Stereo matching [36] is the core procedure of the 3D reconstruction method for binocular vision. The goal of stereo matching is to find the corresponding homologous point pairs in the two views and to acquire the disparity. The disparity of the corresponding pixels is estimated by establishing a similarity measurement function to minimise the correlation value [37] in order to obtain the depth. According to epipolar geometry [36], stereo matching constrains and simplifies the complex 2-dimensional (2D) search problem

into a simple 1-dimensional search. The matching process uses the "winner takes all" (WTA) criterion to estimate the disparity of each pixel. General similarity measurement functions in stereo-matching algorithms include the sum of absolute difference, sum of squared difference and zero normalised cross correlation (ZNCC).

2.2. Stereo-Matching Method Based on Spatiotemporal Correlation

A diagram of stereo matching based on the spatiotemporal correlation is shown in Figure 1. A series of time-varying speckle patterns are projected onto the surface of the measured object, and two cameras synchronously acquire a set of speckle-coded stereo-image pairs. To implement the speckle spatiotemporal correlation calculation, the projector must project multiple different speckle patterns continuously. Most current solutions use commercial projectors [30–33], but they are too large to meet the requirements of 3D face imaging equipment production. Given these considerations, our group designed a 3D face acquisition device based on a rotating speckle projector, which consists of a light-emitting diode (LED) light source, multiple sets of lenses and speckle diaphragms, which can project any number of speckle patterns. A schematic of a 3D face acquisition device with a rotary speckle projector is shown in Figure 2. After the binocular camera acquires a pair of stereo images, the speckle mask rotates at a certain angle and stops. The camera then acquires the next pair of images. In the experimental part of this study, the equipment is employed to acquire multiple speckle stereo image pairs to analyse the relationship between the number of projection patterns, 3D reconstruction accuracy and calculation speed.
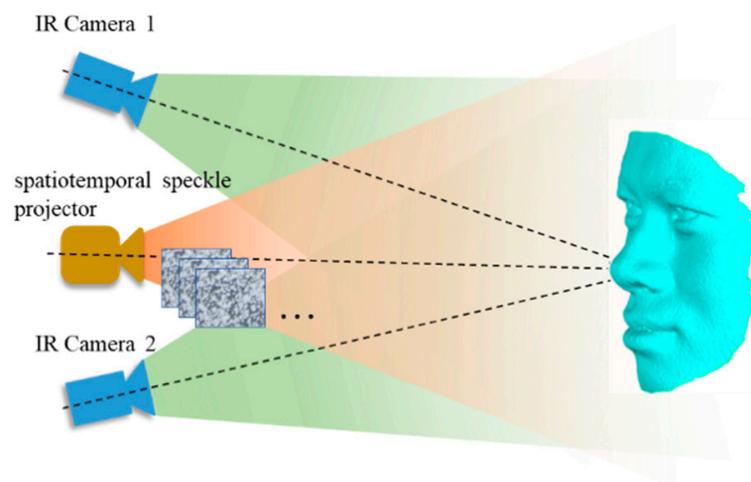

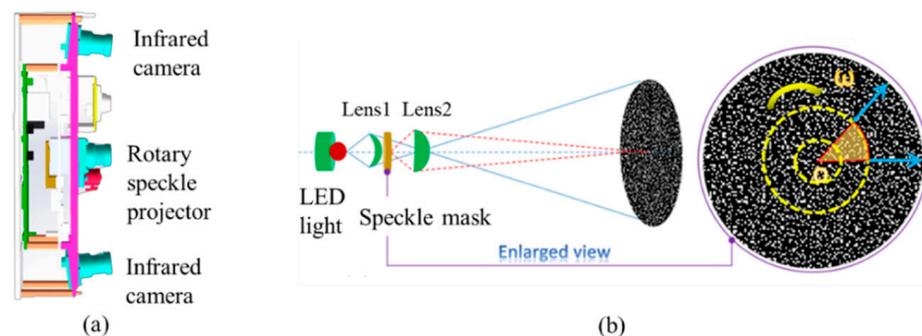**Figure 1.** Spatiotemporal stereo scheme for fast three-dimensional face reconstruction.

Figure 2. Schematic diagram of 3-dimensional face acquisition device with rotary speckle projector: (**a**) structure map of the device; (**b**) internal structure of the rotary speckle projector, where the speckle mask rotates at angular velocity ω.

When performing spatiotemporal stereo matching, by calculating the correlation coefficients in two cubes composed of correlation windows in the time domain and the spatial domain, the matching positions of homologous points can be obtained. The stereo-matching method based on the spatiotemporal correlation is shown in Figure 3. Taking p, one matching point in the left image, as the centre, a rectangular region, Ω, having a width of w_x and a length of w_y on the temporal domain along the sampling time, the axis is formed. The same-size cube from the right camera is chosen to perform the spatiotemporal correlation operation with the volume around p.
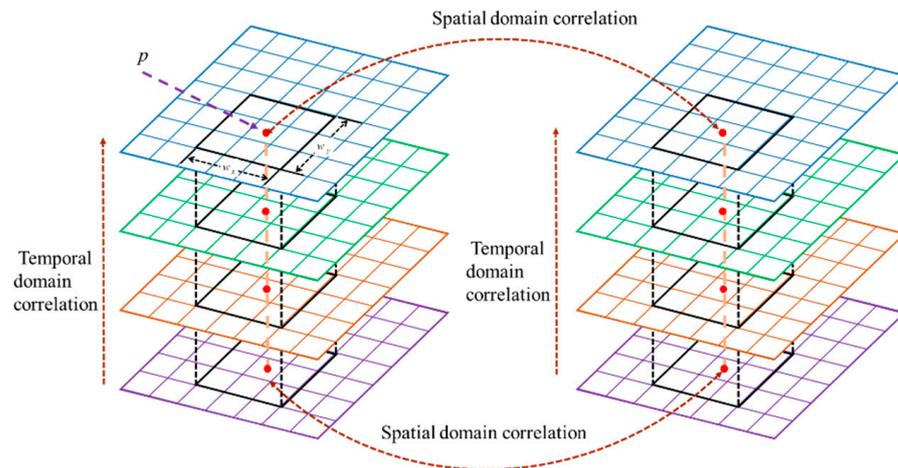

Figure 3. Stereo matching based on spatiotemporal correlation. Correlation coefficients are calculated between two cubes composed of correlation windows in the temporal and spatial domains.

According to the epipolar constraint characteristics of binocular vision, the corresponding points in the left and right views are searched on a line parallel to the baseline. A similarity measure function is considered for spatiotemporal correlation: the spatiotemporal zero mean normalised cross-correlation (STZNCC) [34]. This function can better adapt to the characteristics of speckle images, which effectively reduce the impact of inconsistent brightness of the left and right images. STZNCC expression is defined as follows:

$$C_{STZNCC}(x,y,d) = \frac{\sum\limits_{t=1}^{N} \sum\limits_{h=-w_y/2}^{w_y/2} \sum\limits_{l=-w_x/2}^{w_x/2} \left[L_{(x,y)}(l,h,t) - \overline{L_{(x,y)}} \right] \cdot \left[R_{(x-d,y)}(l,h,t) - \overline{R_{(x-d,y)}} \right]}{\sqrt{\sum\limits_{t=1}^{N} \sum\limits_{h=-w_y/2}^{w_y/2} \sum\limits_{l=-w_x/2}^{w_x/2} \left[L_{(x,y)}(l,h,t) - \overline{L_{(x,y)}} \right]^2} \sqrt{\sum\limits_{t=1}^{N} \sum\limits_{h=-w_y/2}^{w_y/2} \sum\limits_{l=-w_x/2}^{w_x/2} \left[R_{(x-d,y)}(l,h,t) - \overline{R_{(x-d,y)}} \right]^2}}$$

(1)

where (x,y) denotes the pixel coordinates, d is the disparity and $C_{STZNCC}(x,y,d)$ is the correlation coefficient of the disparity; d, in the right image. $L_{(x,y)}(l,h,t)$ and $R_{(x-d,y)}(l,h,t)$ denote the pixel intensity in the matching window of the t-th left image and right image, respectively; $\overline{L_{(x,y)}}$ and $\overline{R_{(x-d,y)}}$ denote the average intensity of the left and right in a $w_x \times w_y$ window, respectively; and where N is the total number of speckle stereo image pairs by ranging l, l over interval $[-w_x/2, w_x/2]$, h over interval $[-w_y/2, w_y/2]$ and t over $[1, N]$.

$$\overline{L_{(x,y)}} = \frac{\sum\limits_{t=1}^{N} \sum\limits_{h=-w_y/2}^{w_y/2} \sum\limits_{l=-w_x/2}^{w_x/2} L_{(x,y)}(l,h,t)}{Nw_xw_y},$$

(2)

$$\overline{R_{(x-d,y)}} = \frac{\sum\limits_{t=1}^{N} \sum\limits_{h=-w_y/2}^{w_y/2} \sum\limits_{l=-w_x/2}^{w_x/2} R_{(x-d,y)}(l,h,t)}{N w_x w_y}. \tag{3}$$

After rectification, the matching-cost computation of homonymous points can be performed on the same line. The WTA criterion is used to obtain the final disparity, where the best cumulative matching cost is calculated.

2.3. Coarse-to-Fine Spatiotemporal Correlation Computation Scheme

As shown in Figure 4, the proposed scheme uses binocular stereo [36] as the basic architecture. The infrared speckle projector continuously projects a set of binary random speckle patterns to the measured face, and the binocular infrared camera simultaneously collects stereo image pairs. Because the camera parameters are obtained through Zhang's [38] pre-calibration method, the image pairs are rectified according to its parameters to ensure that homologous points are searched along the epipolar line and matched. During the disparity computation procedure, the coarse-to-fine spatiotemporal correlation algorithm is adopted. Meanwhile, a spatiotemporal box filter (STBF) is used to accelerate correlation computation. After the subpixel disparity computation is completed, a fine-grained 3D face model can be obtained.
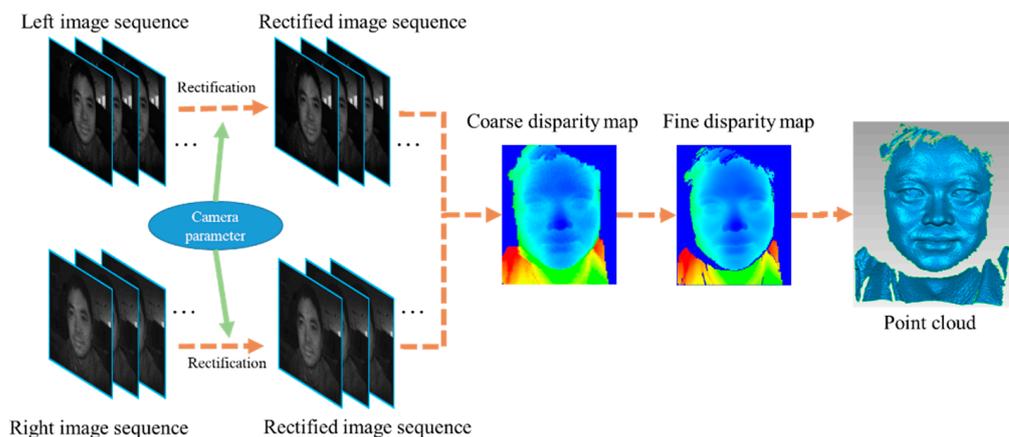

Figure 4. Block diagram of the fast stereo-matching method for 3D face reconstruction.

During the process of disparity computation in this study, a coarse-to-fine two-level spatiotemporal correlation matching strategy based on a grid search is adopted.

2.3.1. Coarse Disparity Estimation

The left images were used as reference images and the right images were used as targets. The reference images are divided into small horizontal and vertical grids by a fixed step interval, and the disparity at the cross point in the reference image is obtained by searching for the matching point in the target image. After the correlation threshold is set, the pixel with the maximum correlation value is identified as the homologous point. A smaller matching window can improve the computational speed of the matching-point search, but a larger window can greatly reduce the error rate of matching because matching tends to be easier. At this stage, the coarse disparity of each grid point is only used as a guide value for the subsequent fine disparity calculation; hence, a relatively large correlation window was chosen to reduce the matching error rate. The grid-point interval is consistent with the size of the coarse matching window in order to ensure full coverage of the initial disparity. The search range $[d_c^{\min}, d_c^{\max}]$ of the disparity is determined by the depth region of the measured object. Disparity consistency and sequential constraints were employed to remove outliers. To further reduce the computation of disparity, a disparity propagation strategy [39] is introduced to narrow the search range to a smaller

radius, r_c^p. Because our measurement target is a human face, the disparity calculation satisfies the continuous and slowly changing constraint. Therefore, after the previous effective matching disparity was obtained, the disparity search range was also reduced to a certain interval when searching for adjacent matching points. Note that the correlation computation at the current stage may still cause some missing disparity, thereby creating holes. It is thus necessary to interpolate the disparity map to fill-in the missing disparity. According to the aforementioned continuity assumption, because the disparity of a reliable grid point, p_g, has been acquired, the disparity of its next adjacent point can be initialised using the disparity of p_g by up-sampling.

2.3.2. Fine Disparity Estimation

Following the coarse disparity, d_c^p, computation, the square area centred on the grid point in the disparity map is filled with the disparity of the grid point. Because the coarse disparity map has given the initial position of the fine-matching search, it is only necessary to calculate the fine disparity, d_f^p, within the narrow search range of $[d_f^{min}, d_f^{max}]$, which significantly reduces the matching search range. This means that the final disparity, d, fits the condition, $d = d_c^p + d_f^p$. Because the final disparity is determined by the result of fine matching, the selection of the fine matching window size directly determines the accuracy of 3D face reconstruction. To ensure measurement accuracy, a smaller correlation window is introduced in order to effectively reduce the computation cost.

2.3.3. Disparity Selection and Sub-Pixel Disparity Refinement

The disparity selection is performed using the WTA strategy, wherein each disparity is selected corresponding to the maximum correlation coefficient, as follows:

$$D(\mathrm{p}) = \underset{d \in S_d}{\mathrm{argmax}} C_{stzncc}(\mathrm{p},d). \tag{4}$$

Now that the integer pixel position, $(x - d_{int}, y, N)$, having the largest correlation value, is found in the target image as the best candidate point, a quadratic curve is fitted with five points centred on the matching point. The matched sub-pixel disparity, d_{sub}, is computed according to the coordinates corresponding to the extreme points of the quadratic curve:

$$C_{stzncc}(x, y, x - d_{sub}, y, N) = a(x - x_{sub})^2 + c, x \in [x_{\max} - 2, x_{\max} + 2], x_{\max} = x - d_{int}, \tag{5}$$

where a and c are the fitting coefficients, $(x - d_{sub}, y, N)$, representing the coordinate of the maximum value of the fitting curve and $C_{stzncc}(x, y, x - d_{sub}, y, N)$ is the new matching position point in place of $(x - d_{int}, y, N)$. Finally, the desired disparity map is acquired through post-processing (e.g., filtering-out outliers and smoothing the surface).

2.4. Spatiotemporal Box Filter

A box filter is a fast-filtering algorithm performed recursively in a 2D space. The basic idea is to make full use of the previous computation result during the loop computation, which ensures that the next computation result can be acquired with only a small computational effort. The typical implementation is a sliding window. The summing process is accelerated by adding move-in pixels to the sliding window and subtracting the move-out pixels. The computation is independent of window size. Therefore, a box filter that eliminates redundant information is more prominent than an integral image method [35]. The box filter can reduce the complexity of the summation operation from $O(w_x w_y)$ to $O(1)$. The traditional box filter is used only in the spatial domain. This method is now extended to the temporal domain. This STBF is schematically shown in Figure 5.

$$S(x + 1, y) = S(x, y) + \sum_{t=1}^{N} \sum_{h=-w_y/2}^{w_y/2} \left[I_{(x,y)}(w_x/2 + 1, h, t) - I_{(x,y)}(-w_x/2, h, t) \right], \tag{6}$$

$$S(x, y+1) = S(x, y) + \sum_{t=1}^{N} \sum_{l=-w_x/2}^{w_x/2} \left[I_{(x,y)}(l, w_y/2 + 1, t) - I_{(x,y)}(l, -w_y/2, t) \right]. \quad (7)$$
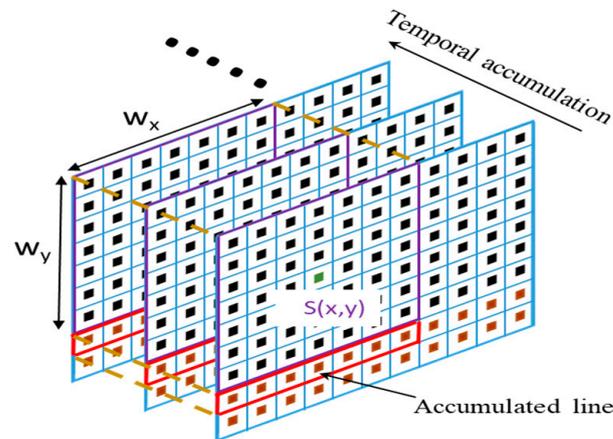

Figure 5. Spatiotemporal box filter.

The acceleration computation of STBF in the row and column directions is given in Equations (6) and (7), where $S(x, y)$ is the sum of grey values in the spatiotemporal rectangular window centred on (x, y). To perform fast-matching cost computations in the entire disparity cube, the sum and the sum of squares of grey values in the matching window in Equation (1) must be calculated in advance by STBF.

2.5. Real-Time Acquisition and Reconstruction of 3D Face

To achieve the goal of real-time acquisition and real-time reconstruction of 3D faces, the authors consider optimising the system from two aspects: shortening the image acquisition time and increasing the speed of 3D reconstruction. In the previous design method, a rotary speckle projector was used to construct a spatiotemporal correlation 3D face acquisition device. Owing to the influence of the inertial motion of the object, it was impossible to ensure that the rotary speckle mask could be switched quickly in a short time. Simultaneously, owing to the particularity of the face target, the image acquisition time should be shortened as much as possible to reduce the problem of decreased measurement accuracy caused by face movement. Therefore, the idea of using multiple fixed speckle projectors is proposed to replace the rotary speckle projector for fast acquisition of 3D faces; the feasibility of the idea is verified through subsequent experimental analysis.

To improve the computing performance of the system, the authors consider decomposing the multiple steps of 3D face reconstruction into multiple sub-function modules, and each module is executed in parallel in a pipeline. The continuous acquisition and display procedure of the 3D face based on a fixed speckle projection is shown in Figure 6. The entire pipeline is divided into four functional modules: image data acquisition, epipolar rectification and disparity computation, point-cloud generation and point-cloud display. The four modules create threads to work in a synchronous and parallel pipeline mode to ensure that each thread simultaneously obtains the maximum utilisation of each component of the hardware system. Three buffer queues are inserted between the four modules: image, disparity and point cloud queues. Among the modules, the epipolar rectification and disparity computation is the core of the entire operation which costs the most time. Therefore, the function was added to the GPU to accelerate the computation. Cooperating with the high-speed image data acquisition hardware, real-time computation and display of the 3D face is achieved.
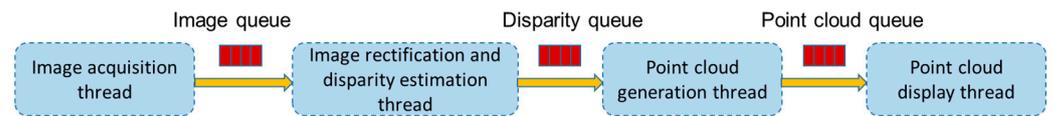
Figure 6. Flow diagram of images acquisition and 3-dimensional face generation.

3. Results

3.1. Setup

Experiments were conducted on a desktop PC equipped with an Intel I7-9700K CPU (3.6 GHz), 8-GB RAM and an NVIDIA GeForce RTX1660Ti GPU. The spatiotemporal correlation matching algorithm was implemented using C++ using Visual Studio 2015. Several function libraries (i.e., OpenMP, OpenCL and OpenGL) were used to develop our algorithm. The setup with a rotary speckle projector for capturing infrared (IR) speckle images is shown in Figure 7. The tested object was placed 500-mm directly in front of the projector. The device has two 200-Hz infrared light cameras with a baseline length of 120 mm. To avoid interference with human eyes and reduce the absorption of infrared light by the skin, the device was equipped with a 730-nm near-IR LED as speckle projection illumination, which is non-inductive to humans. IR cameras capture images simultaneously, whereas the rotary speckle projector projects speckle-encoded patterns onto a facial mask sequentially. The projected pattern number is consistent with the number of infrared stereo image pairs collected. The projector comprises an infrared light source, micromotor, pattern mask and lens group. A speckle pattern mask was generated by pseudo-random encoding. When the binocular camera acquires images synchronously, it is necessary to ensure that the speckle mask in the rotary projector remains stationary. The motor then drives the mask gear to rotate a certain angle (e.g., 9°) to ensure that there is no correlation between the continuously collected speckle encoding images, because the speckle patterns of the two pairs of continuous images are quite different. To reduce the motion blur caused by the rotating mask, the next set of image pairs is collected after the mask is completely still. The specific relationship between the rotation angle and the accuracy of 3D reconstruction is being studied by other members of our group, and this content is beyond the scope of this article. When the experimental device is operating, the acquisition time of infrared images depends heavily on the number of speckle patterns required by the spatiotemporal correlation algorithm. The exposure time of the infrared cameras was 2 ms, and the speckle pattern rotatory switching time was 11 ms for each image pair. Therefore, N × 13 ms is required to acquire N pairs of stereo images. A 3D face acquisition device with a rotating speckle projector is shown in Figure 7.
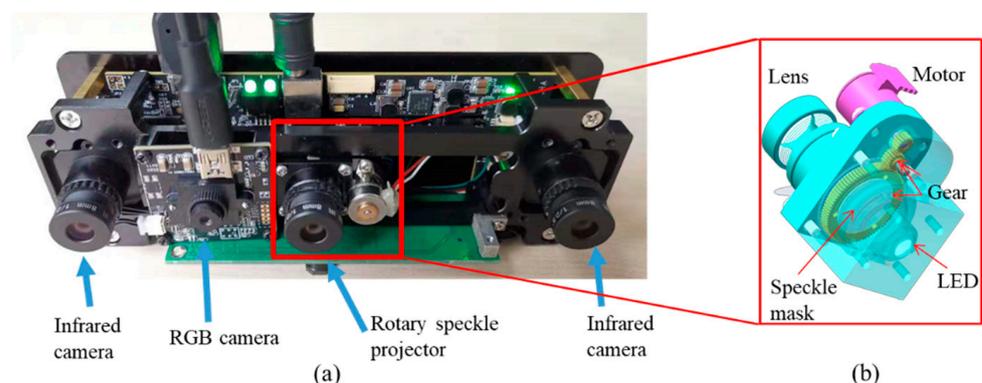

(a)　　　　　　　　　　　　　　　　　　　　　(b)

Figure 7. (**a**) Three-dimensional face acquisition equipment with rotary speckle projector. (**b**) Internal structure diagram of rotating speckle projector.

To overcome the problem of an excessively long acquisition time caused by the switching of the rotary speckle pattern, a compact 3D face acquisition device using a

fixed three-speckle projector was designed. A 3D face acquisition device with three fixed speckle projections is shown in Figure 8. The three fixed speckle projectors use different speckle masks and project the speckle patterns in sequence, and the two infrared cameras synchronously collect speckle-coded images in a time-sharing manner. The switching time of the speckle projection was reduced significantly. In the system, the interval for taking a single pair of speckle images was 5 ms, and three pairs of stereo-image acquisitions cost 15 ms total. The characteristics of short-time exposure and fast acquisition in the system showed good robustness and anti-interference characteristics for the 3D reconstruction of slight movements of the human face. The reason for adopting the structural design of three fixed speckle projections was deduced from the subsequent experimental analysis.
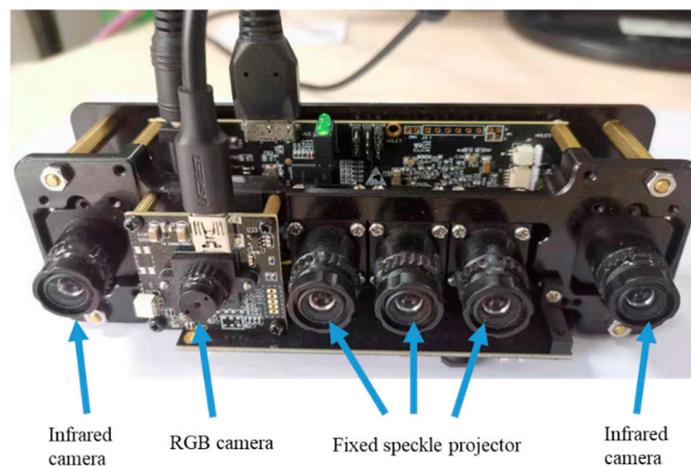

Figure 8. Three-dimensional face acquisition device via three fixed-speckle projections.

Meanwhile, a high-precision industrial 3D scanner (ATOS Core 300 with measurement accuracy ± 0.02 mm) was used to collect 3D face mask data as ground truth, certified by VDI/VDE 2634 Part2 [40]. The reconstructed 3D face model data acquired by our device were compared with Core 300 to obtain the measurement accuracy of our algorithms and equipment. Figure 9 shows a pair of speckle stereo images taken by our experimental device. Figure 10 shows the 3D scanned data gathered using the ATOS Core 300.
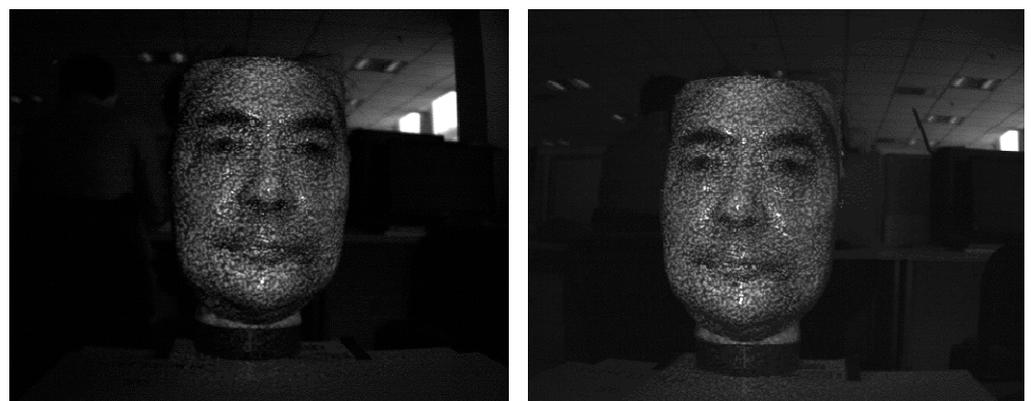

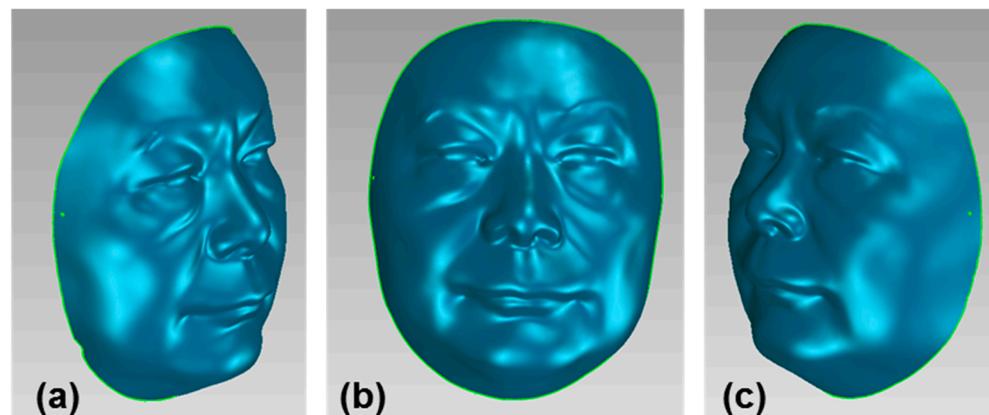**Figure 9.** Pairs of infrared image acquired by our device.

Figure 10. Three-dimensional scanned data with ATOS Core 300; (**a**–**c**) are different perspectives of the 3D model.

3.2. Evaluation on 3D Reconstruction Precision and Performance

From the computation theory of spatiotemporal correlation, it can be deduced that capturing more image pairs leads to higher accuracy. On the contrary, fewer image pairs reduce the capturing time and have a weak impact on the measurement accuracy of moving objects. Consequently, this study focuses on improving the measurement efficiency and optimising the compromise between the accuracy and speed of the algorithm.

Several IR image pairs were projected with a rotary speckle projector using the experimental device shown in Figure 7. The number of speckle stereo image pairs, N, ranged from 1 to 12, and the resolution was 1280 × 1024. The binocular infrared camera was calibrated using Zhang's [38] method to obtain the camera parameters, with which the acquired IR images were rectified to facilitate correlation computation. To obtain a higher measurement accuracy, different combination parameters were used for different numbers of stereo image pairs. Different matching window sizes were adopted for image pairs, N, during coarse and fine matching. During coarse disparity calculation, usually, coarse window sizes, w_c, ranged from 5 to 15, the grid interval was equal to the window sizes, coarse disparity ranged $d_c^{\min} = -200$, $d_c^{\max} = 200$ and the shrunken search radius was $r_c^p = w_x^c + 2$ for disparity propagation. For fine disparity calculation, the disparity range, $d_f^{\min} = -w_x^f - 1$, $d_f^{\max} = w_x^f + 1$ and the search radius, $d_f^p = (d_f^{\max} - d_f^{\min})/2 + 1$, were applied for the coarse disparities, and the fine matching window size, w_f, which ranged from three to seven. Normally, the coarse matching window-size setting conditions are $w_c = w_f + \Delta w$, $\Delta w = 2$ *or* 4. In our experiments, the matching threshold, $\tau = 0.3$, was set for the spatiotemporal correlation coefficients in order to discard unreliable matching.

According to the coarse-to-fine spatiotemporal correlation method, the coarse matching window, w_c, is only used to calculate the initial value of the disparity, and the final accuracy of the measured objects are determined by the fine matching window, w_f. The cost of calculating the coarse disparity is small. Therefore, in the following description, the parameter, w_f, which has the greater impact on reconstruction accuracy and time cost, is considered.

When the matching window size needs to be selected, the larger window plays the role of a low-pass filter, which slows the steep changes of the surface of the measured objects, resulting in reduced spatial resolution or loss of high-frequency details. A smaller window can obtain richer detailed information on the surface of the object, but it also leads to an increase in the matching error rate. In this sense, the choice of matching window size is not absolute, and an appropriate window size must be investigated from a statistical perspective to balance the reliability of object reconstruction and spatial resolution. The 3D reconstruction results are shown in Figure 11, where the number of stereo image pairs, N, is combined with a fine window size, w_f.
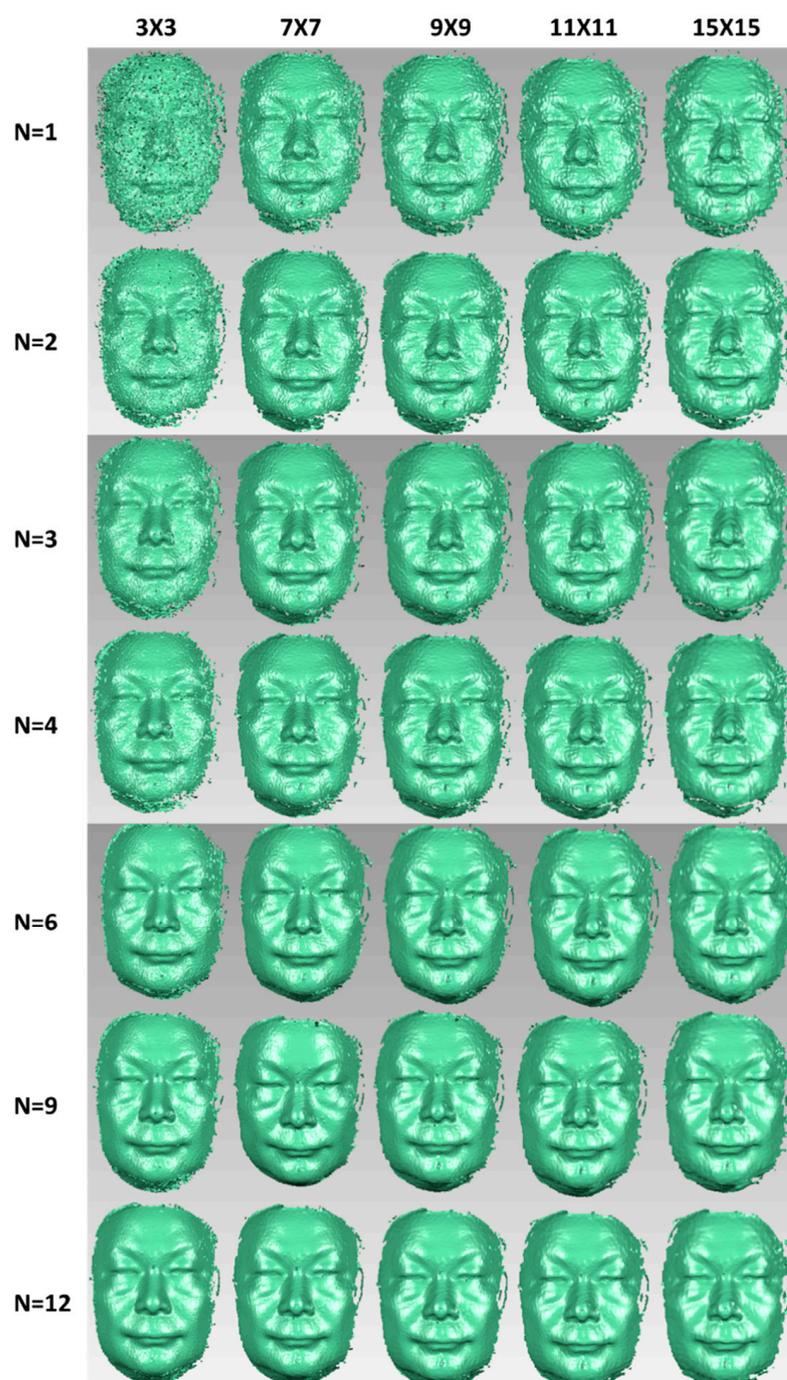
Figure 11. Three-dimensional reconstruction results with a number of stereo image pairs, N, and a fine matching window, w_f. The reconstruction accuracy continues to improve as N increases. When N is small (e.g., N ≤ 3), a larger window creates a more precise and complete model.

The impact of changes in the projected pattern number, N, and fine matching window, w_f, was analysed on the accuracy of 3D reconstruction through a specific quantitative analysis. The average error and standard deviation were used as error evaluation criteria. The variation curves of the average error and standard deviation along with the change in N, w_f, are shown in Figure 12. As shown, when N continues to increase, the measurement error decreases accordingly. Meanwhile, the optimal correlation window does not remain static but decreases as N increases. Figure 13 shows the curves of the disparity computation time corresponding to different N values with the enlargement of the fine matching window, w_f.
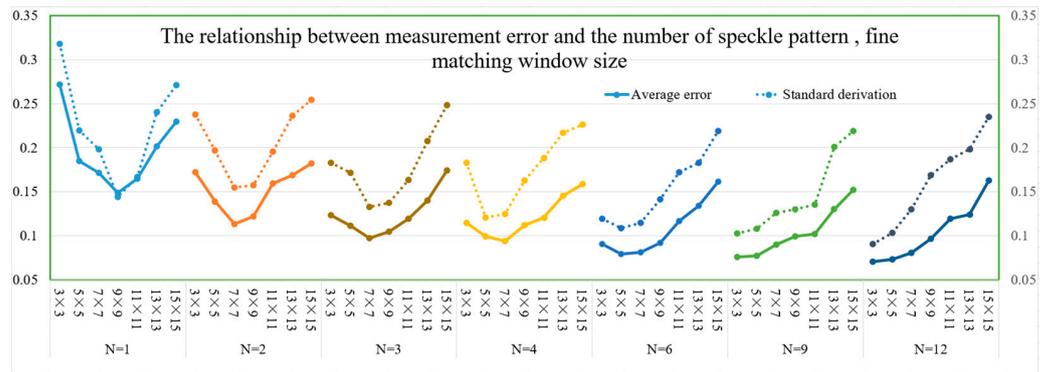
Figure 12. Comparison of reconstruction error obtained between the projected pattern numbers, N, and fine matching window, w_f. When N continues to increase, the reconstruction error decreases accordingly. Meanwhile, the optimal window size will also drop.
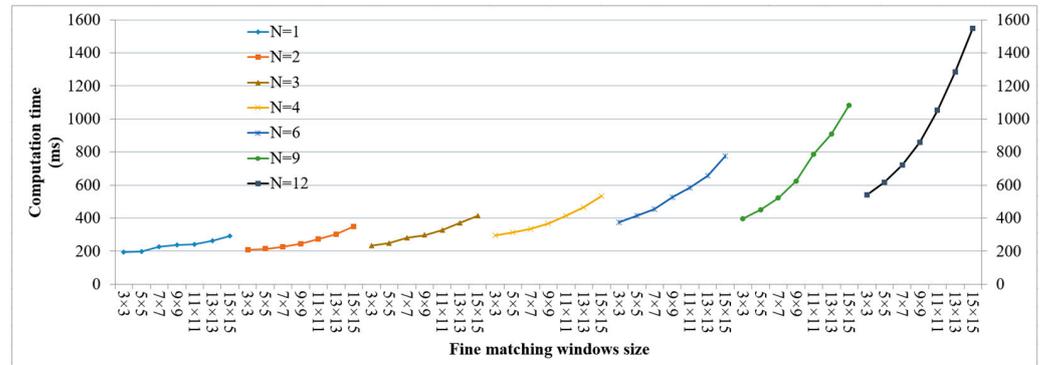

Figure 13. Curves of computation time corresponding to the change of fine window size for different N.

Regarding the choice of matching window size, a larger window image in the measurement of shape will function as a low-pass filter and suppress rapid variations in the measured field, leading to a reduction in spatial resolution or loss of high-frequency details, including an increase in computation cost. A comparison between the reconstruction results obtained by our proposed method and the scanned model using Core 300 is shown in Figure 14.

Figure 15a–c shows the three-dimensional reconstruction results of different real faces when the speckle stereo image pair N = 12, and each group of results contains a comparison of untextured and textured maps. The 3D reconstruction results of the author's various expressions when N = 12 speckle stereo image pairing are shown in Figure 16. The contours of the lips, eyebrows, eye sockets and sides of the nose of the person are clearly displayed in the results.

Based on the 3D face acquisition experimental platform and reconstruction method of rotary speckle projection, combined with Figures 10–13 and Table 1, it is found that the 3D reconstruction results have the following characteristics:

(1) In the current configuration environment, the reconstruction accuracy of the 3D model continues to improve as N increases the number of speckle stereo image pairs, but when the projected speckle patterns exceed a certain range (N ≥ 6), the trend of accuracy improvement gradually weakens.

(2) To obtain a higher 3D reconstruction accuracy, the optimal fine matching window is not as small as possible. For a given number of patterns, the optimal fine window size is determined for a given number of patterns. For example, when N = 1, 9 × 9 is the optimal fine window size and the measurement error is the smallest; when N = 3, the optimal fine window size is 7 × 7. When N = 12, the optimal fine window size

becomes 3×3, the average error is 0.071 mm, and the standard deviation is 0.091 mm. From the overall trend, as the number of speckle stereo image pairs increase, the optimal matching window continues to shrink.

(3) The computation time increases as the matching window increases. The greater the number of projected patterns involved in the calculation, the faster the computation time. As shown in Figure 10, in the case of a fixed number of speckle patterns, the computation time increases proportionally with the increase in the matching window.

(4) There is a trade-off between measurement accuracy and calculation cost. When the control average error is less than 0.15 mm, the overall reconstruction error decreases with the increase in the number of projection patterns, and the optimal window size continues to shrink. Combining the comprehensive analysis of Figures 9 and 10, it is found that to obtain the best measurement accuracy for each number, N, of stereo image pairs, the computation time also increases slightly as the number of speckle patterns increases.
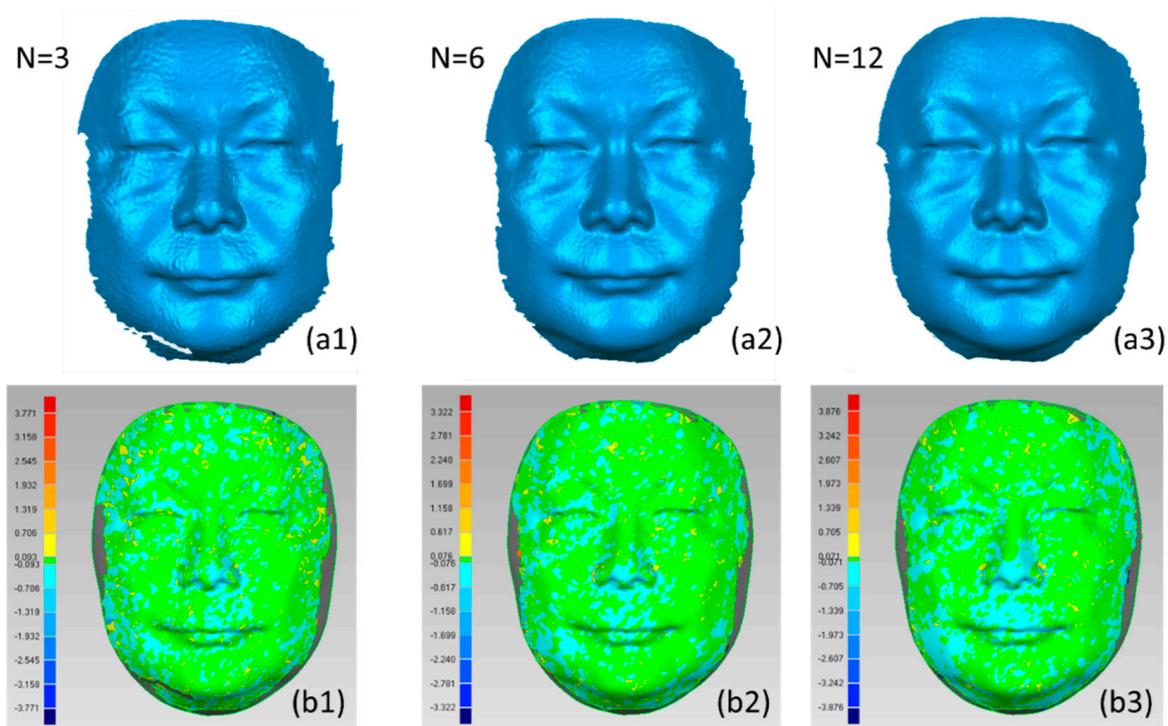

Figure 14. (**a1–a3**) Three-dimensional (3D) reconstruction results of (N = 3, 6, 12) different speckle stereo image pairs. (**b1–b3**) Corresponding comparison maps with 3D scanned data by ATOS Core 300.
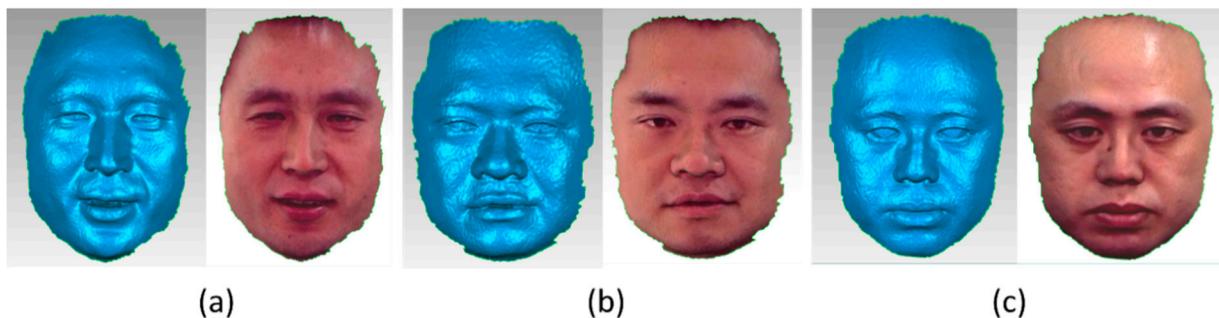

Figure 15. (**a–c**) Three-dimensional reconstruction results of different real faces with N = 12 the speckle stereo image pairs: a contrast display of textured and non-textured maps.
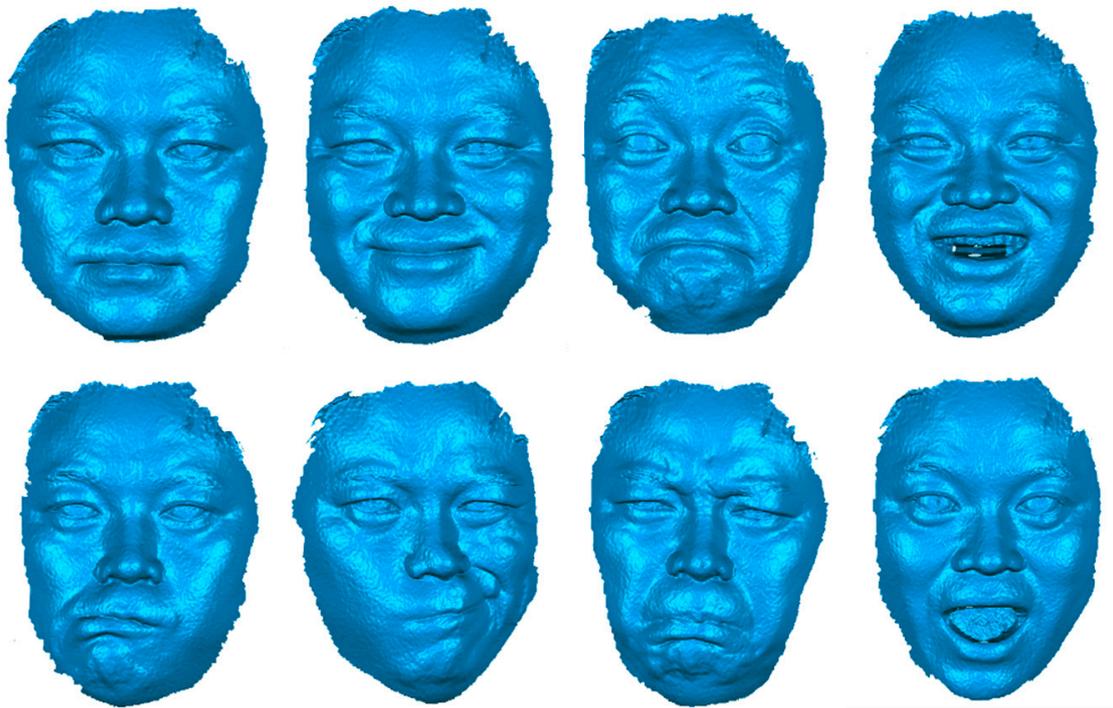
Figure 16. Three-dimensional reconstruction results of the author's various expressions when N = 12 the speckle stereo image pairs.

Table 1. Minimal error statistics and computation time of measuring a face mask (Unit: mm).

Pattern Number	Optimal Window [1] (pixel)	Min. Avg. Err [2] (mm)	Min. Std. [3] (mm)	Computation Time on CPU [4] (ms)	Computation Time on GPU [5] (ms)
N = 1	9 × 9	0.149	0.144	218	30
N = 3	7 × 7	0.097	0.133	280	35
N = 6	5 × 5	0.079	0.109	338	65
N = 9	3 × 3	0.076	0.102	390	97
N = 12	3 × 3	0.071	0.091	554	121

[1] Optimal matching window corresponding to the pattern number. [2] Minimum average error of fitting the face mask between ours and Core 300. [3] Minimum standard derivation of fitting the face mask between the proposed method and Core 300. [4]: The computation time on the CPU when the optimal matching window corresponds to N. [5]: Computation time on the GPU when the optimal matching window corresponding to N.

3.3. Real-Time Improvement for Three Speckle Patterns Projection

Following the previous optimal parameter analysis, comprehensively considering the measurement accuracy, image acquisition time, cost, volume, power consumption, use environment and other factors, the design scheme of a compact 3D face acquisition device was adopted with three fixed speckle projectors. Based on the optimum parameter combination of the three-speckle pattern projections obtained from the previous analyses, the coarse matching window with a size of 11 × 11, grid interval of 11 and fine matching size of 7 × 7 was selected. The parameters were applied to the GPU parallel acceleration process of the spatiotemporal correlation computations. The multi-core parallel function library, openMP, was employed for dense point cloud generation and triangulation of 3D surfaces. The point-cloud dynamic real-time display was achieved using the OpenGL pipeline strategy. The image acquisition, epipolar rectification and disparity estimation, point-cloud generation and 3D display were implemented in parallel through a multi-threaded pipeline. Based on the three-speckle pattern projection equipment, high-speed image acquisition and real-time 3D face reconstruction were achieved with an average

error of 0.098 mm, a standard deviation of 0.133 mm, a 15-ms image acquisition time and a 35-ms disparity computational time.

4. Discussion

Regarding the implementation of the proposed method and the experimental process, several aspects need to be further explained and discussed.

(1) Measurement accuracy. The STBF-accelerated spatiotemporal correlation matching strategy proposed in this paper differs from the test results in the literature [23]. First, the literature [23] did not consider the human face as the analysis object and did not use a coarse-to-fine matching strategy, and its calculation speed was not as good as the method in this paper. Furthermore, based on the comparison of the calculation results, it was found that the combination trend between the selected matching window size and the number of speckle patterns was different. In [23], when the number of projected speckle patterns exceeded three frames, the change in the size of the matching window had only a small effect on the reconstruction accuracy. The test results in this study show that, as the number of projection patterns increases, the 3D reconstruction accuracy continues to improve. When the number of patterns is greater than six, the trend of accuracy improvement gradually slows down. This phenomenon may be caused by differences in the measurement object.

(2) Balance of measurement accuracy and time cost. A 3D face acquisition device with a rotating speckle projector is more suitable for use in scenes where the accuracy requirements are strict, and there is no clear limitation on the acquisition time. To obtain better reconstruction accuracy, more stereo image pairs were selected to participate in the spatiotemporal stereo-matching process, which is more suitable for scenes where objects remain stationary. According to the research results of this study, six sets of image pairs met the requirements for high-precision modelling. When reconstructing fast-moving objects, the authors attempted to reduce the number of stereo image pairs and employed 3D reconstruction equipment with fixed speckle projectors for 3D face image acquisition.

(3) Real-time acquisition, reconstruction and display. Using the single-shot speckle structure [17–19], the acquisition time of the stereo image pair was short, but the real-time reconstruction effect was usually not achieved. Although the literature [27] achieved a real-time reconstruction frequency of 30 fps, the accuracy was only 0.55 mm, which is far from the accuracy of the proposed method. Aiming at 3D face acquisition equipment using a fixed speckle projector, the method in this study implements real-time image acquisition and real-time 3D face reconstruction. However, during the 3D data display process, the point-cloud structure can be used to display 3D data in real time through the openGL interface. If the triangular facet structure of texture mapping is adopted, it is limited by openGL's utilisation of the low-level image cache and does not display 3D data in real time. In our next step, the authors will study the openGL parallel display strategy, which has achieved a more realistic display effect of 3D data.

5. Conclusions

To improve the accuracy and performance of 3D face imaging, an effective coarse-to-fine spatiotemporal stereo matching scheme using speckle pattern projection was proposed, which was accelerated by STBF. Comparing our 3D face reconstruction results with the ground-truth data collected by high-precision 3D industrial scanning equipment for global errors, the impact of the projected speckle-pattern number and matching windows on measurement accuracy and reconstruction speed were further researched. Through quantitative and qualitative analyses of the experimental data, the optimum combination of parameters needed to provide a balanced strategy of reconstruction speed and accuracy was made. It was demonstrated that the system can achieve accuracy with an average measurement error of 0.071 mm and standard deviation of 0.091 mm with N = 12 speckle stereo-image pairs. Experimental results showed that increasing the number of projected

speckle patterns yields better measurement accuracy, while the optimal matching window will continue to shrink and the computation time will increase slightly. To meet the requirement of rapid 3D face imaging under slight motion, this study provides a compact 3D face acquisition device with three fixed speckle projectors. Using the optimal combination parameters for three-speckle stereo images, the spatiotemporal correlation stereo-matching process, which takes the longest time among the components, is accelerated by the GPU through the use of a parallel pipeline strategy between each core processing unit; the purpose of real-time data acquisition and real-time 3D reconstruction. In the future, the authors intend to apply the proposed scheme to 3D face verification and recognition in practical scenarios.

Author Contributions: Conceptualisation, W.X., K.F. and P.Z.; Data curation, J.Z. and P.Z.; Investigation, H.Y., J.Z. and P.Z.; Software, W.X. and P.Z.; Writing—original draft, W.X. and K.F.; Writing—review and editing, W.X. and K.F. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Khan, D.; Shirazi, M.A.; Kim, M.Y. Single-shot laser speckle-based 3D acquisition system for medical applications. *Opt. Lasers Eng.* **2018**, *105*, 43–53. [CrossRef]
2. Gilani, S.Z.; Mian, A. Learning from millions of 3D scans for large-scale 3D face recognition. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; p. 1896.
3. Hassner, T. Viewing real-world faces in 3D. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; p. 3607.
4. Sturm, J.; Bylow, E.; Kahl, F.; Cremers, D. CopyMe3D: Scanning and Printing Persons in 3D. *Pattern Recogn.* **2013**, *8142*, 405–414.
5. Fyffe, G.; Nagano, K.; Huynh, L.; Saito, S.; Busch, J.; Jones, A.; Debevec, P. Multi-view stereo on consistent face topology. *Comput. Graph. Forum* **2017**, *36*, 295–309. [CrossRef]
6. Zhang, S. High-speed 3D shape measurement with structured light methods: A review. *Opt. Lasers Eng.* **2018**, *106*, 119–131. [CrossRef]
7. Cester, L.; Lyons, A.; Braidotti, M.; Faccio, D. Time-of-Flight imaging at 10-ps resolution with an ICCD camera. *Sensors* **2019**, *19*, 180. [CrossRef] [PubMed]
8. Booth, J.; Roussos, A.; Ponniah, A.; Dunaway, D.; Zafeiriou, S. Large scale 3D morphable models. *Int. J. Comput. Vis.* **2018**, *126*, 233–254. [CrossRef] [PubMed]
9. Bouaziz, S.; Wang, Y.G.; Pauly, M. Online modelling for real-time facial animation. *ACM Trans. Graph.* **2013**, *32*, 40. [CrossRef]
10. Garrido, P.; Zollhofer, M.; Casas, D.; Valgaerts, L.; Varanasi, K.; Perez, P.; Theobalt, C. Reconstruction of personalized 3D face rigs from monocular video. *ACM Trans. Graph.* **2016**, *35*, 28. [CrossRef]
11. Jackson, A.S.; Bulat, A.; Argyriou, V.; Tzimiropoulos, G. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression. In Proceedings of the 16th IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; p. 1031.
12. Feng, Y.; Wu, F.; Shao, X.; Wang, Y.; Zhou, X. Joint 3D face reconstruction and dense alignment with position map regression network. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
13. Kwon, H.; Kim, Y.; Yoon, H.; Choi, D. Classification score approach for detecting adversarial example in deep neural network. *Multimed. Tools Appl.* **2021**, *80*, 10339–10360. [CrossRef]
14. Kwon, H.; Lee, J. AdvGuard: Fortifying Deep Neural Networks against Optimized Adversarial Example Attack. *IEEE Access* **2020**. [CrossRef]

15. Xue, J.; Zhang, Q.; Li, C.; Lang, W.; Wang, M.; Hu, Y. 3D face profilometry based on Galvanometer scanner with infrared fringe projection in high speed. *Appl. Sci.* **2019**, *9*, 1458. [CrossRef]

16. Ito, M.; Ishii, A. A three-level checkerboard pattern (tcp) projection method for curved surface measurement. *Pattern Recogn.* **1995**, *28*, 27–40. [CrossRef]

17. Zuo, C.; Feng, S.; Huang, L.; Tao, T.; Yin, W.; Chen, Q. Phase Shifting Algorithms for Fringe Projection Profilometry: A Review. *Opt. Lasers Eng.* **2018**, *109*, 23–59. [CrossRef]

18. Boyer, K.L.; Kak, A.C. Colour-encoded structured light for rapid active ranging. *IEEE Trans. Anal. Mach. Intell.* **1987**, *PAMI-9*, 14–28. [CrossRef] [PubMed]

19. Baek, S.-H.; Kim, M.H. Stereo fusion: Combining refractive and binocular disparity. *Comput. Vis. Image Underst.* **2016**, *146*, 52–66. [CrossRef]

20. Shi, H.; Zhu, H.; Wang, J.; Yu, S.-Y.; Fu, Z.-F. Segment-based adaptive window and multi-feature fusion for stereo matching. *J. Algorithm Comput. Technol.* **2016**, *10*, 3–11. [CrossRef]

21. Liu, Y.; Zhang, Q.; Liu, Y.; Yu, X.; Hou, Y.; Chen, W. High-speed 3D shape measurement using rotary mechanical projector. *Opt. Express* **2021**, *29*, 7885. [CrossRef]

22. Song, Z.; Tang, S.; Gu, F.; Shi, C.; Feng, J. DOE-based structured-light method for accurate 3D sensing. *Opt. Lasers Eng.* **2019**, *120*, 21–30. [CrossRef]

23. Keselman, L.; Woodfill, J.I.; Grunnet-Jepsen, A. Intel®RealSense™ Stereoscopic Depth Cameras. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; p. 1267.

24. Pathirana, P.N.; Li, S.Y.; Trinh, H.M.; Seneviratne, A. Robust real-time bio-kinematic movement tracking using multiple Kinects for tele-rehabilitation. *IEEE Trans. Ind. Electr.* **2016**, *63*, 1822–1833. [CrossRef]

25. Gu, F.; Song, Z.; Zhao, Z. Single-Shot Structured Light Sensor for 3D Dense and Dynamic Reconstruction. *Sensors* **2020**, *20*, 1094. [CrossRef]

26. Khan, D.; Kim, M.Y. High-density single shot 3D sensing using adaptable speckle projection system with varying pre-processing. *Opt. Lasers Eng.* **2021**, *136*, 106312. [CrossRef]

27. Guo, J.P.; Peng, X.; Li, A.; Liu, X.; Yu, J. Automatic and rapid whole-body 3D shape measurement based on multi-node 3D sensing and speckle projection. *Appl. Opt.* **2017**, *56*, 8759–8768. [CrossRef]

28. 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*, 13388. [CrossRef]

29. Zhou, P.; Zhu, J.; Jing, H. Optical 3-D surface reconstruction with colour binary speckle pattern encoding. *Opt. Express* **2018**, *26*, 3452–3465. [CrossRef]

30. Ishii, I.; Yamamoto, K.; Doi, K.; Tsuji, T. High-speed 3D image acquisition using coded structured light projection. In Proceedings of the International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October 2007; p. 925.

31. Große, M.; Kowarschik, R. *Space-Time Multiplexing in a Stereo Photogrammetry Setup*; Osten, W., Kujawinska, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 755–759.

32. Harendt, B.; Grosse, M.; Schaffer, M.; Kowarschik, R. 3D shape measurement of static and moving objects with adaptive spatiotemporal correlation. *Appl. Opt.* **2014**, *53*, 7507–7515. [CrossRef]

33. Tang, Q.J.; Liu, C.; Cai, Z.W.; Zhao, H.W.; Liu, X.L.; Peng, X. An improved spatiotemporal correlation method for high-accuracy random speckle 3D reconstruction. *Opt. Lasers Eng.* **2018**, *110*, 54–62. [CrossRef]

34. Zhou, P.; Zhu, J.P.; You, Z.S. 3-D face registration solution with speckle encoding based spatial-temporal logical correlation algorithm. *Opt. Express* **2019**, *27*, 21004–21019. [CrossRef]

35. Fu, K.; Xie, Y.; Jing, H. Fast spatial-temporal stereo matching for 3D face reconstruction under speckle pattern projection. *Image Vis. Comput.* **2019**, *85*, 36–45. [CrossRef]

36. Fu, L.; Peng, G.; Song, W. Histogram-based cost aggregation strategy with joint bilateral filtering for stereo matching. *Int. J. Comput. Vis.* **2016**, *10*, 173–181. [CrossRef]

37. Xue, Y.; Cheng, T.; Xu, X.; Gao, Z.; Li, Q.; Liu, X.; Wang, X.; Song, R.; Ju, X.; Zhang, Q. High-accuracy and real-time 3D positioning, tracking system for medical imaging applications based on 3D digital image correlation. *Opt. Laser Eng.* **2017**, *88*, 82–90. [CrossRef]

38. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]

39. Barnes, C.; Shechtman, E.; Finkelstein, A.; Goldman, D.B. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* **2009**, *28*, 24. [CrossRef]

40. *Optical 3-D Measuring Systems—Optical Systems Based on Area Scanning: VDI/VDE 2634 Blatt 2-2012*; Beuth Verlag: Berlin, Germany, 2012.