2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing

A Structured Light 3D Measurement System Based On Heterogeneous Parallel Computation Model

Xiaoyu Liu[1], Hao Sheng [1,2], Yang Zhang[1], Zhang Xiong[1,2]
[1]*State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing 100191, P.R.China*
[2]*Research Institute of Beihang University in Shenzhen, Shenzhen 518057, P.R.China*
xiaoyuliu@buaa.edu.cn,shenghao@buaa.edu.cn,zhangyang1991@buaa.edu.cn,xiongz@buaa.edu.cn

***Abstract*—We present a structured light measurement system to collect high accuracy surface information of the measured object with a good real-time performance. Utilizing phase-shifting method in conjunction with a matching method proposed in this paper which can significantly reduce the noisy points, we can achieve high accuracy and noiseless point cloud in a complex industrial environment. Due to the use of the heterogeneous parallel computation model, the parallelism of the algorithm is developed in a deep way. The OpenMP+CUDA hybrid computing model is then used in the system to get a better real-time performance.**

***Keywords*-structured light, phase matching, parallel computation model, real-time**

I. INTRODUCTION

In recent years, three-dimensional measurement technology has been used for various applications in many areas, including industrial inspection, product quality control, reverse engineering, medicine, etc. Due to the advantages of fast, high accuracy and non-contact, structured light technology has been developed rapidly in the field of optical three-dimensional measurement. With the development of cameras and projectors, the phase measuring profilometry of structured light technology has been widely used[1].

After years of development, the method of structured light three-dimensional measurement has made great progress, especially in the accuracy and real-time performance. A kind of method[2] using a high-speed camera and DLP projector have been proposed, which has a good real-time performance. However, it cannot satisfy the requirement of many industrial applications for its lack of accuracy. A kind of method[3] using a multi-layer neural network to establish the mapping relationship between height distribution and phase map has a high accuracy result but a poor real-time performance. A kind of accurate three-dimensional measurement system[4] with good real-time performance has been implemented based on GPU, but the measured scene of these systems is too small to meet most industrial applications. Additionally, cameras of these systems rarely have a high resolution.

Currently, a lot of industrial applications require the measurement system of both good real-time performance and high accuracy, such as online measurement system. In such an industrial application, measurement should be completed with a high accuracy on the premise of not affecting the production efficiency in a complex environment. Moreover, the measurement result would be affected by many factors, such as the ambient light, the photography background and inter-reflections in the scene, etc. In order to solve these problems, we proposed a structured light 3D measurement system based on the heterogeneous parallel computation model. To gain both good real-time performance and high accuracy, works have been done in the following aspects. Primarily, a new matching method has been adopted to make the system obtain a stabilized measurement result for different measured object and complex measuring environment. Secondly, algorithms in the system have been designed based on heterogeneous parallel computing model. Finally, OpenMP+CUDA parallel programming model has been used in the system.

The remainder of this paper is organized as follows. Section 2 reviews the related work and fundamental theories of the structured light and heterogeneous parallel computation model. Section 3 illustrates the framework of the proposed system. The theory of phase calculation method, the phase matching method, and heterogeneous parallel computation model is elaborated in Section 4. The use of heterogeneous parallel programming model is presented in Section 5. Section 6 shows the experiment and analyses the experimental results. The conclusion is summarized in Section 7.

II. RELATED WORK

A. Structured Light 3D Measurement

Typical structured light 3D measurement system is made up of CCD cameras used for image capturing and a DLP projector used for digital fringe pattern image projection[5]. Measuring process is reproduced below. We use a DLP projector to project a set of fringe pattern images to the measured object. At the same time, we capture the images modulated by the object using CCD cameras. Then, we use the depth information contained in the captured images to obtain the three-dimensional surface information of the object. In conventional structured light measurement methods,

978-1-4799-8006-2/15 $31.00 © 2015 IEEE
DOI 10.1109/CCGrid.2015.69

two kinds of structured light are widely used: binary Gray-code patterns and series of sine waves[6]. Generally, using binary Gray-code patterns can avoid the influence caused by objects with different reflectivity in the scene, but more patterns need to be projected. Utilizing the method of sine waves needs fewer patterns and gets higher resolution, but it is sensitive to the gamma nonlinear of the projector.

With the development of CCD cameras and digital projection technology, the cost of PMP (phase measuring profilometry) technology has been greatly reduced which promotes its extensive use. Phase calculate is an important part of the PMP, a variety of phase-shifting algorithms have been proposed, such as standard N step phase-shifting algorithm, N frames average algorithm, N + 1 step phase-shifting algorithm and any equal step length phase-shifting algorithm, etc. Phase unwrapping is a hot issue in the field of phase measurement. So far, a lot of phase unwrapping algorithms have been proposed. Temporal phase unwrapping algorithm has been widely used in industrial measurement because of its good performance in measuring objects with discontinuous Gray-code with phase-shift algorithm and multi-frequency heterodyne principle are two of the most popular kinds of temporal phase unwrapping algorithms[7]. Gray-code with phase-shift method is easy to implement, but the unwrapping process depends on the accuracy of image binaryzation, and it is sensitive to the color of the surface of the measured object. When measuring dark objects, a thin layer of white powder is necessary to spray on the surface. Because Gray-code pattern can only be used for phase unwrapping, it is helpless to improve the accuracy of phase calculation. Wrapped phase of pattern images are in different frequency, so the phase unwrapping algorithm based on multi-frequency heterodyne principle is insensitive to the color of the surface of the measured object. It can measure objects with dark surface without spraying powder that can make the calculation process more stable.

Ishii[8] presented a spatio-temporal selection type coded structured light projection method and image acquisition at a high-speed frame rate by using a high-speed camera. Karpinsky[9] proposed a technique, which implements the entire process on a GPU to achieve real-time measurement. Li[10] proposed a 3D measurement method based on phase-shifting and three-layer BP neural network. Neither of these methods solves the high accuracy and real-time problems in complex industrial situation.

B. Parallel Computation Model

Parallel computation model is an abstraction of the parallel computer. It evolves with the development of parallel computer. The first generation of parallel computation models, such as PRAM[11] and QSM[12] describes the shared memory parallel computer. PRAM model is commonly used in the field of parallel computing theory. This model is an abstraction of the parallel computer with shared storage structure and assumes that all processors operated in a synchronously way, so it is very easy to be used.

The second generation of parallel computation models, such as BSP[13] and LogP[14] model is an abstraction of distributed storage parallel computer. The purpose of BSP model is to build a bridge between software and hardware for parallel computations, which can improve the efficiency of software and hardware designing. So far, many scholars are still using BSP model for related research work. LogP model was proposed by Culler, mainly based on the following background. On the one hand, the MPP is the mainstream of high-performance computing systems during that time, so it is an urgent mission to establish a new parallel computation model to describe the characteristics of hardware. On the other hand, the existing model did not consider the actual situation such as the distribution storage and asynchronous communication, which lead the model cannot accurately reflect the program running status on a parallel machine. Nowadays LogP class model still has an important means of designing parallel algorithms and simulating network performance.

As the problem of the storage wall was getting more and more serious, the third generation of parallel computation models pays more attention to describing the complex memory level in parallel computers, such as Log_nP[15] and HPM[16]. Kirk W. Cameron proposed the log_nP model, which combined the throughput-driven copy-transfer model with the memory communication cost estimated by an earlier model (memory LogP[17]). The $lognP$ model considered the influence of middleware on performance and incorporated its cost into a model of point-to-point communication on emergent systems at the expense of complexity.

III. Framework Overview

To build a high accuracy and real-time structured light measurement system, the following problems need to be solved. Phase calculation method should be considered primarily because the result of phase calculation will affect the final accuracy of the measurement directly. The matching method should be improved to get a point cloud with higher quality. The system should be implemented in a high-performance way to obtain a good real-time performance.

The framework of the system proposed in this paper was illustrated in Fig. 1. Firstly, the standard four-step phase shift algorithm is selected as the calculation method of the wrapped phase on the basis of comparing various phase shift algorithms. Secondly, the phase unwrapping algorithm based on multi-frequency heterodyne principle is selected as the phase unwrapping method on the basis of analyzing the advantages and disadvantages of the existing phase unwrapping algorithms. Then we propose a new matching algorithm after analyzing the traditional algorithms to reduce the noise. Fully developed the parallelism of the methods above. Algorithms in the system are designed by using a

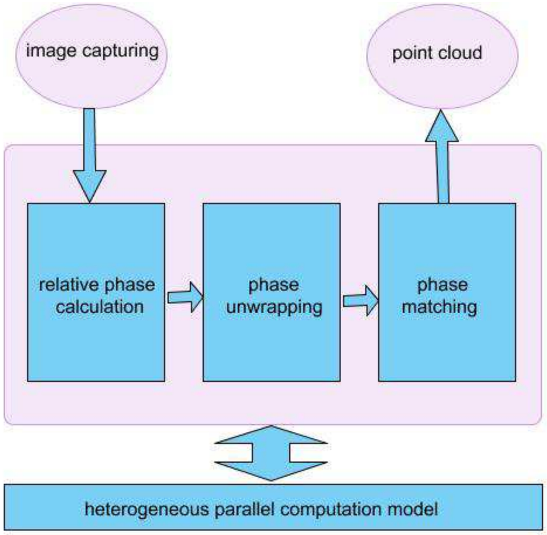


Figure 1. The framework of the process pipeline in this paper

parallel computation model. Finally, the system is realized in a high-performance way of using OpenMP+CUDA hybrid parallel programming model.

IV. Theory

A. Phase Calculation

Phase calculation, including wrapped phase calculation and phase unwrapping, is the basis of the phase measurement technology. Its basic idea is to collect several frames of images of phase shift to calculate the three-dimensional surface information of the measured object. Assuming the light intensity in the fringe pattern image is sine distribution. The intensity distribution function could be shown as:

$$I_i(x,y) = a(x,y) + b(x,y)\cos[\phi(x,y) + \delta_i] \tag{1}$$

In the equation above, to each pixel (x,y) in the fringe pattern image, $a(x,y)$ is the intensity of the background; $b(x,y)$ is the amplitude of the modulation; $\phi(x,y)$ is the phase to be calculated. δ_i is the time-varying phase shift of the ith image. Varieties of phase-shifting algorithms have been proposed, but the stability and error response is different. So the selection of phase-shifting algorithms has an important impact on phase calculation and point cloud reconstruction subsequently. Standard N frame phase-shifting algorithm with optimal inhibitory has an effect to reduce the random noise, which has become one of the most widely used algorithms in structured light measurement technology. Four-step phase-shifting method is used to calculate $\phi(x,y)$ in our system. Therefore, each pattern should be projected four times, and the phase shift should be 0, $\pi/2$, π, $3\pi/2$. Expressions are as follows:

$$\begin{aligned}
I_1(x,y) &= a(x,y) + b(x,y)\cos[\phi(x,y)] \\
I_2(x,y) &= a(x,y) + b(x,y)\cos[\phi(x,y) + \pi/2] \\
I_3(x,y) &= a(x,y) + b(x,y)\cos[\phi(x,y) + \pi] \\
I_4(x,y) &= a(x,y) + b(x,y)\cos[\phi(x,y) + 3\pi/2]
\end{aligned} \tag{2}$$

Wrapped phase can be calculated according to 2.

$$\phi(x,y) = \arctan\frac{I_4 - I_2}{I_1 - I_3} \tag{3}$$

Because of the arctangent function, this phase map is wrapped by 2π. Phase unwrapping method should be used to obtain a continuous unwrapped phase map $\Phi(x,y)$. Phase unwrapping is a hot issue in the field of phase measurement. After years of development, there are much phase unwrapping algorithms. We choose a phase unwrapping algorithm based on multi-frequency heterodyne principle, which has a good stability and high accuracy. This method can be divided into two steps. First step can be named as phase resultant. The phase function $\Phi_1(x,y)$ superimposed on another phase function $\Phi_2(x,y)$ with different frequency to obtain phase function $\Phi_3(x,y)$ with a lower frequency. Assuming the frequency of $\Phi_1(x,y)$ is λ_1, the frequency of $\Phi_2(x,y)$ is λ_2 and the frequency of $\Phi_3(x,y)$ is λ_3, the relationship of these parameters can be expressed as 4:

$$\lambda_3 = \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2} \tag{4}$$

According to 4, continuous phase map can be obtained when the frequency of $\Phi(x,y)$ is 1. At the same time, the continuous phase also called absolute phase. Frequency of $\Phi(x,y)$ only related to the frequency of the initial phase map. So, in order to get the absolute phase which has a frequency is 1, appropriate initial frequency should be selected. The second step can be called heterodyne unwrapping. With the continuous phase map, heterodyne principle can be used to unwrap the wrapped phase value of each point. Expression of unwrapping phase is shown as 5:

$$\Phi_1(x,y) = \phi_1(x,y) + N(x,y) \times 2\pi \tag{5}$$

In this equation, $N(x,y)$ is

$$N(x,y) = INT\left(\frac{\Phi(x,y) \times r - \phi_1(x,y)}{2\pi}\right)$$

In $N(x,y)$, r is the ratio of the two frequency and $\Phi(x,y)$ is the absolute phase in the process of heterodyne principle.

B. Phase Matching

In binocular measuring methods, the depth information of a point in the scene is calculated by the disparity of its location in the left and right images. The disparity is estimated by phase matching. We can assume that the phase map of left image is $\Phi_l(x_l,y_l)$ and the right is $\Phi_r(x_r,y_r)$ and the fringe pattern images have been stereo corrected before matching. The processing of phase matching is that for a point (x_l, y_l) in the left image find a correspond point (x, y_r) in the right image to satisfy the 6.

$$\begin{cases} \Phi_l(x_l, y_l) = \Phi_r(x_r, y_r) \\ y_l = y_r \end{cases} \tag{6}$$

The disparity of this point can be expressed as

$$d(x, y) = x_l - x_r$$

We use three different frequencies, four phases of each frequency, a total of 12 pictures. Theoretically, phase matching can be done with just one unwrapped phase map of any frequency. Then, point cloud could be calculated. However, in actual industrial applications, due to the ambient light, the background of the scene and other factors, the raw point cloud calculated in this way would contain a lot of noise. It will bring many troubles to further operation. Almost all the further operation depends on the effect of noise reduction. In our system, we were unwrap all these three frequency fringe patterns, and get three unwrapped phase map: $\Phi_1(x, y)$, $\Phi_2(x, y)$ and $\Phi_3(x, y)$. Then three disparity maps will be got by match the unwrapped phase maps respectively. The three disparity maps can be expressed as $d_1(x, y)$, $d_2(x, y)$ and $d_3(x, y)$. The final disparity will be calculated by the three disparity maps, it will depend on three conditions as below:

- $d(x, y) = (d_1(x, y) + d_2(x, y) + d_3(x, y))/3$. For each point, if its three disparity values are close to each other, the final disparity should be the algebra average of these three values. For example, if three disparity values of point (u, v) are $d_1(u, v) = 100.01$, $d_2(u, v) = 100.02$ and $d_3(u, v) = 100.03$, three values are close enough to each other, then $d(u, v) = 100.02$.
- $d(x, y) = (d_i(x, y)+d_j(x, y))/2$. For each point, if two of its three disparity values are close to each other, then the final disparity will be the algebra average of these two values. For example, if three disparity values of point (u, v) are $d_1(u, v) = 100$, $d_2(u, v) = 100.02$ and $d_3(u, v) = 101.01$, $d_1(u, v)$ and $d_2(u, v)$ are closed to each other but the distance of $d_3(u, v)$ between these two values is too far, than $d(u, v) = 100.01$.
- 3. For each point, if neither of its three disparity values is close to each other, then the final disparity will be null. For example, if three disparity values of point (u, v) are $d_1(u, v) = 100$, $d_2(u, v) = 101$ and $d_3(u, v) = 102$, , the distance between any two points is too far, then $d(u, v) = null$.

Using this method can significantly reduce the number of noisy point in the raw point cloud data. However, the downside of this method is that its computational complexity is higher than the original method.

C. Parallel Computation Model

Nowadays, the design of processor is increasingly constrained by energy consumption. In view of GPU has very high floating-point computation ability and a capacious memory access bandwidth, scholars have applied it into many research fields. In the field of green high performance computing, the CPU and GPU heterogeneous system have great advantages in performance, energy consumption and has a huge development potential. On the one hand, multi-core CPU can handle scalar computing and provide general computing ability, making the heterogeneous system can be applied to a variety of applications. On the other hand, GPU can provide powerful computing performance for applications in some specific areas, so that the heterogeneous system has good performance and low energy consumption. The algorithm with superior performance in heterogeneous systems, however, must be designed depends on the reasonable parallel computation model.

$Hlog_nGPM$ model[18] combines the advantages of $HlogGP$[19]model which describing heterogeneous cluster and log_nP[15] model which describing point-to-point communication. $LogP$ model mainly describes various communication operations in parallel computing system, including storage communication and network communication. Storage communication refers to data communication by the operation of memory copy between processors while network communication refers to data transmission through high-speed interconnection network between different processors. $HLogGP$ model mainly considers communication operations in loosely-coupled heterogeneous cluster, and it describes the heterogeneity of a distributed parallel system composed of high-speed interconnection network. The heterogeneity involves both computing nodes and network. $HLogGP$ model follows the basic parameters definition of $LogGP$[20] model and extend each parameter in view of the compute nodes and interconnection network heterogeneity. Log_nP model has more consideration in the storage communication, in which the parameter n is the number of storage and network communication times in the message transmission. So $Hlog_nGPM$ is more suitable for the current high-performance heterogeneous parallel computing systems. In this model, names of these six parameters:L, o, g, n, G, and P are still used. The meaning of these parameters is explained as follows:

- Latency, L: It is the maximum latency when a single (short) message transmitted from the source to the destination. Define vector $L = \{l_1, l_2, \ldots, l_n\}$, in which l_i is the latency of ith atomic communication. This parameter is affected by the performance of the transmission medium and transmission direction.
- Overhead, o: It is the time that a processor needs to do an atomic communication. This parameter depend on the performance of the processor. Define vector $o =$

$\{o_1, o_2, \ldots, o_n\}$, in which o_i represent the overhead cost when processor does ith atomic communication.

- Gap between massages, g: It is the processor performance for sending consecutive messages, and the minimum interval of the processor sending two consecutive messages. It mainly depends on the network interface. A gap vector $g = \{g_1, g_2, \ldots, g_n\}$ is defined because of the different kinds of atomic communications.
- Atomic communication number, n: The total number of atomic communications in the model.
- Gap per byte, G: It describes the ability of the transmission medium to transmit long messages. The reciprocal of this parameter is the bandwidth of the transmission medium. A message can cross different transmission medium with different bandwidths in heterogeneous systems, so gap vector $G = \{G_1, G_2, \ldots, G_n\}$ can be defined. G_i depends on the data size.
- Computational power, P: The performance of the processor in the system. For heterogeneous systems, the performance of CPU and GPU should be paid attention to. So it can be defined as $P = \{P_C, P_G\}$.

Assuming that point-to-point communication operations in the heterogeneous systems (message length is k) contains n atomic communications, the corresponding minimum communication bandwidth is $\{1/G_{1l}, 1/G_{2l}, \ldots, 1/G_{nl}\}$, and maximum communication bandwidth is $\{1/G_{1u}, 1/G_{2u}, \ldots, 1/G_{nu}\}$, then communication time can be predicted based on $Hlog_nGPM$ model. The prediction time is $t_l < t < t_u$, in which $t_l = \sum_{i=1}^{n}\{2o + (k-1)G_{iu} + L\}$, $t_u = \sum_{i=1}^{n}\{2o + (k-1)G_{il} + L\}$.

V. System Implementation Using Parallel Programming Model

Parallel programming model is an abstraction of parallel algorithms execution. A good programming model should be easy to understand, easy to use, cross-platform, and high-performance. However, the existing programming models are unable to balance all of the requirements above.

In order to improve the performance of the structured light measurement system, OpenMP+CUDA hybrid programming model has been chosen. CUDA is a programming language developed by NVIDIA, providing a lot of efficient instructions, which enables developers to build a high-performance computing solutions based on the powerful computing capabilities of GPU[21]. Moreover, CUDA supports C, C++, FORTRAN and other high-level languages, which makes it easy to use. In academia and industry, scientists have been using CUDA successfully accelerate many applications and made dozens of times performance improvement. The main disadvantage of many hybrid models using GPU is the computing capability of the CPU cannot be effectively utilized. To solve this problem, OpenMP model has been adopted to improve the parallelism of multi-core CPU. In fact, OpenMP has become the de facto standards for parallel programming on shared storage architectures. It provides program interfaces to share storage applications on numerous platforms through a set of commands, function libraries and environment variables. OpenMP program is derived and merged. Program starts with a single thread, and then derives a number of additional threads. These threads will merge together after the execution of the parallel codes. OpenMP program has two features of equivalence series and increasing parallel, which significantly reduce the difficulty of the parallel program development and maintenance. In the current field of high-performance computing, completely shared storage architecture parallel system has been rarely seen. Pure OpenMP program could not be effectively extended to large systems. OpenMP is usually utilized in conjunction with other models (like CUDA) to composite hybrid programming model.

The phase calculation algorithm with no correlation between each pixel is designed in this paper. The processing of a single pixel runs on the GPU as a fine-grained thread, and each thread only takes up a very small GPU local shared memory, which can take full advantage of GPU's high-performance parallel computing ability. Therefore, on the premise of ensuring speed, we do more complex calculations in the algorithm to ensure the accuracy and stability of phase calculation. The parallelism of data has been developed in a deep way in the matching algorithm to make multi-core CPU and GPU devices hide all kinds of delays by scheduling a large number of active threads. A lot of shared memory storage has been used to save memory bandwidth. Coalesced memory access model has been used as much as possible, to make good use of the complicated GPU memory hierarchy. Hardware occupancy rate has a large influence on application performance. Therefore, in order to obtain a higher hardware occupancy, resource usage, hardware features and configuration factors need to be considered.

Although GPU has powerful computing ability, for a lot of preliminary implementation of GPU programs, its performance is only a few times and even less than one time of the CPU implementation. Therefore, the performance optimization is very important for parallel programming. The performance of parallel applications is related to many factors, such as memory access delay and hardware occupancy, so the characteristics of the underlying hardware structure must be considered while optimizing parallel programs. Our system has been optimized based on CUDA performance optimization guidelines and related programming practice experience[22], from the memory hierarchy, maximize occupancy rate and reduced the influence of the branch instruction. Thus, our structured light measurement system gains a very high performance through the design, implementation, and optimization method in this paper.

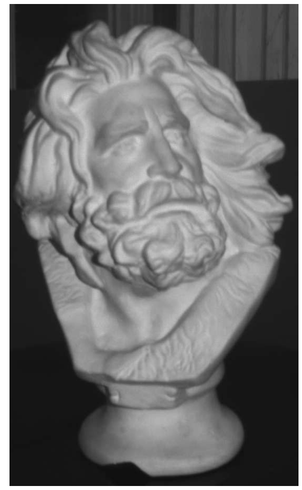
VI. Experiment And Analysis

First of all, the experiment equipment is composed of two cameras (KSJ MU3C500M with a resolution of 2549x1920) and a projector (Texas Instruments DLP Light Crafter Pro 4500) as shown in Fig. 2. The TI Light Crafter Pro 4500, which is designed for 3D measurement using 450nm blue LED light source to reduce the influence of environmental light, can greatly reduce the non-sinusoidal distortion. In this system, the calibration approach is based on Zhang's method[23], and the measuring distance is 700 mm. The calibration needs to be done before the measurement.

Figure 2. The equipment in this paper

Three-frequency heterodyne principle can accomplish the phase calculation of complex objects. In order to test the performance of this algorithm in measuring complex objects and objects with discontinuous surface, a plaster statue with a complex surface and a building block is measured in this experiment. The results are shown in Fig. 3. Fig. 3(a) is the fringe pattern image captured by the camera, and Fig. 3(b) and Fig. 3(c) is the corresponding wrapped phase and unwrapped phase calculated by our system. Fig. 3(d) is the numerical value in line 400 of the unwrapped phase. The color changes from deep to shallow smoothly in Fig. 3(c), and the data strictly increase monotonically in Fig. 3(d). It means that the unwrapped phase continuously and uniquely in the entire scene. Through the experiment, the algorithm is able to complete phase unwrapping of complex and surface discontinuous objects robustly.

The measuring result of the system directly depends on the phase matching results. In industrial applications, the conditions like strong ambient light and darker color of the object surface are more common, which will lead to an unsatisfactory result of phase calculation. If using the traditional method of phase matching, it will generate a lot of noisy points. Using the method we proposed in this paper can significantly avoid it. A plaster statue and a shoe-last with brown surface were chosen as measured objects in this section. The picture of the original objects are shown in Fig. 4. And the measured results are shown in Fig. 5. Fig. 5(a) and Fig. 5(c) are the point cloud measured in a traditional way, and Fig. 5(b) and Fig. 5(d) are measured in the same conditions but using the matching method in this paper. In order to facilitate the observation, the point clouds in these figures have been rendered. Obviously, there are a lot of noisy point in Fig 5(a) and Fiig. 5(c). Especially in Fig. 5(c), quantities of noisy points greatly affected the reconstruction result. In fact, it is caused by the dark color of the surface of the object. However, there is rarely noisy point in the other two figures. In view of the results, using the method in this paper can ensure the quality of point cloud.

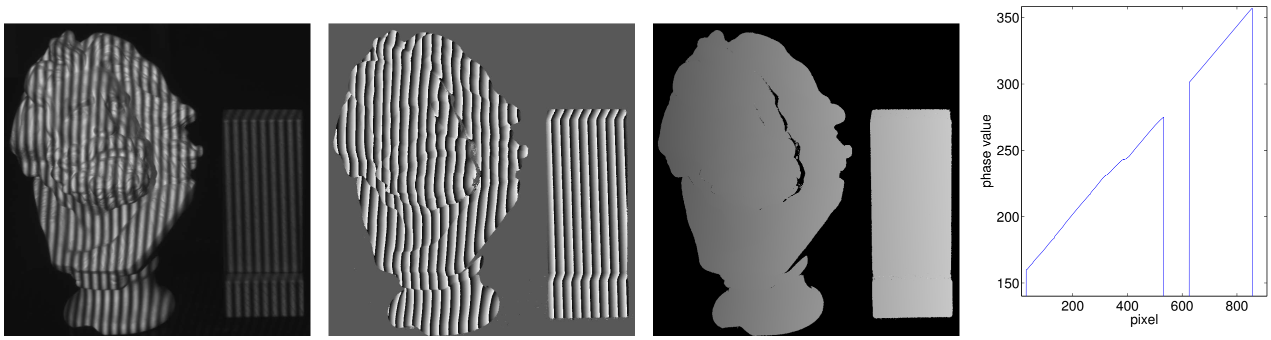


(a)

(b)

Figure 4. Original picture of measured objects. 4(a) Plaster status; 4(b) shoe-last.

Statistics for the number of points of the three point clouds and the amount of noisy point in the corresponding point cloud are shown in Table I. The point clouds are reconstructed from a plaster statue and a shoe-last. The noisy points were calculated by k-neighborhood noise reduction algorithm. By using the traditional matching method, point cloud of the plaster statue has 328256 points, and the number of noisy points is 16888. Thus, the proportion of noisy points is about 5.1%. Using the same method on the shoe-last, 333369 points were got with 30526 noisy points, and the proportion of noisy points is 9%. In fact, for this shoe-last point cloud, it is hard to clear all noisy points simply with k-neighborhood noise reduction algorithm. The reduction result is not good enough even using professional point cloud processing software. Compared with these two point clouds, the dark color of the surface of the object has a big impact on the traditional method. Using the matching method in this



(a) (b) (c) (d)

Figure 3. Visualization of phase calculation process. 3(a) Image captured; 3(b) wrapped phase; 3(c) unwrapped phase; 3(d) value in line 400 of unwrapped phase.

(a) (b) (c) (d)

Figure 5. Visualization of point cloud reconstructed with different matching methods. 5(a) Plaster status reconstructed with the traditional method; 5(b) plaster status reconstructed with our method; 5(c) shoe-last reconstructed with the traditional method; 5(d) shoe-last reconstructed with our method.

Table I. POINTS NUMBER INFORMATION OF OBJECT USING DIFFERENT PHASE MATCHING METHOD

matching method	measured object	number of points	number of noise	proportion of noise
traditional method	plaster statue	328256	16888	5.1%
	shoe-last	333369	30526	9.1%
this paper method	plaster statue	307221	3305	1%
	shoe-lat	267241	2801	1%

paper, 307221 points are captured of the plaster statue, in which the proportion of noisy points is 1%, and its number is 3305. For the shoe-last, 267241 points are captured, and the number of noisy points is 2801 accounted for about 1%. For a same measured object, although the points number of cloud point generated by the method in this paper less than the traditional method, it also saved enough topology and geometry information. At the same time, the point clouds calculated by the method in this paper have a high quality with few noisy points. It can provide great convenience for further processing. Through the data above we can see, the matching algorithm in this paper is more robust, even the measured object and the measuring environment is not ideal.

Processing speed can be promoted by using $Hlong_nGPM$ model designed algorithm and implemented with OPENMP+CUDA programming model. In this section, we will compare the time consumption with same algorithms of different programming models. Our experiment environment is HP Z820(Intel Xeon E5-2603 CPU, nVIDIA Quadro 4000 GPU). The algorithms include phase calculation and phase matching. For each algorithm, we implemented it in three models – CPU, CPU+CUDA, and OpenMP+CUDA. The time consumption is shown in Table II. For phase calculation algorithm, its most frequent operation is unrelated with each pixel, so this algorithm is suitable for fine-grain parallelism implemented on GPU, while implementing it with CPU requires a lot of time to traverse. As it is shown in table, for each implementation, the time consumptions of reconstruction the plaster statue is 2.83s, 1.09s, and 0.21s. The time consumption of OpenMP+CUDA implemented is one-fifth of the OpenMP implemented and one-thirteenth of CPU implemented. The table shows that using OpenMP+CUDA has over ten times increasing than using CPU in speed. Several groups of images in the algorithm have a logical order, which limits the performance of the hardware. Another measured object, shoe-last, almost has the same time consumption as the plaster statue with the same implementation. In the view of time consumption, phase calculation algorithm we proposed would not be affected by different measured objects. Experiment results show that the phase calculation algorithm based on OpenMP+CUDA has a great improvement in time consumption.

There is no logical relationship between any two rows in the unwrapped phase map. So this algorithm has a good parallelism. The amount of calculation in this method is about three times than that of the conventional method. Two methods based on CPU consumed more time is the reaction of this factor. For the plaster statue, the time consumption of the conventional method is 4s, and it is 10s by using the method in this paper. Compared to the time consumption of the plaster statue and shoe-last in the conventional method, it is obvious that the consumption of the shoe-last is higher than that of the plaster statue. It stems from the fact that the matching is affected by different measured objects. The phase calculation result of the shoe-last is not ideal due to the dark surface. Thanks to the huge computing power of hardware, most operations could be executed in parallel. So it is easy to find that the time consumption of different objects and matching methods with OpenMP+CUDA implemented is almost the same. The time consumptions of the shoe-last using the method in this paper implemented by three programming model is 15s, 9s, and 0.76s. Compared with the CPU implementation, the OpenMP+CUDA implementation has twenty times promotions in speed.

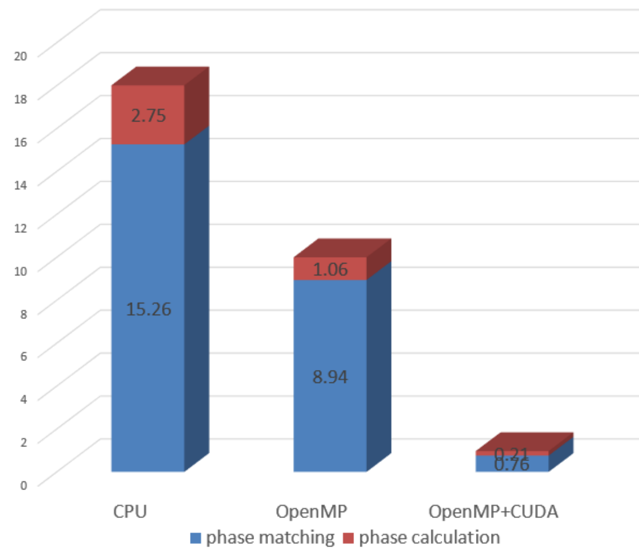


Figure 6. Time consumptions with different implementation when measuring shoe-last

The time consumptions of the entire process using the matching method in this paper with different implementations when measuring shoe-last are shown in Fig. 6. It can be seen that the speed is greatly improved by using OpenMP+CUDA programming model. Although the measured objects and matching method in this paper have effect to the processing speed, using OpenMP+CUDA programming model, the matching algorithm in this paper has nearly the same time consumption with the conventional method, but we get a better result with fewer noisy points.

VII. Conclusion

In this paper, we present a structured light measurement system with high accuracy and good real-time performance. We present the techniques involved in the system including phase calculation, phase matching, and parallel computation. We propose a phase matching method which can significantly reduce the noisy points in the measured point cloud on the promise of saving enough topology and geometry

Table II. TIME CONSUMPTION OF ALGORITHMS WITH THE THREE IMPLEMENTATIONS

measured object	Phase Calculation			matching method	Phase Matching		
	CPU	OpenMP	OpenMP+CUDA		CPU	OpenMP	OpenMP+CUDA
plaster statue	2.83s	1.09s	0.21s	traditional method	4.34s	2.29s	0.75s
				this paper method	10.01s	4.62s	0.76s
shoe-last	2.75s	1.06s	0.21s	traditional method	6.45s	3.4s	0.76s
				this paper method	15.26s	8.94s	0.76s

information. It makes sure the system achieves an accuracy and noiseless point cloud in a complex industrial environment. This study uses the idea of parallel computation model and the concept of parallel programming model for structured light measurement. The system we implemented with OpenMP+CUDA programming model has a good real-time performance. In the future work, with the development of industrial application demand, we will pay attention to further operators of point cloud, such as point cloud registration, polygonization, and surface reconstruction, etc.

ACKNOWLEDGMENT

This work was partially supported by the Natural Science Foundation of China (NSFC) under grant 61272350, the National High Technology Research and Development Program of China (No.2013AA01A603), and the Aeronautical Science Foundation of China (No.2013ZC51), and the National Science & Technology Pillar Program under grant 2015BAF14B00.

REFERENCES

[1] Jason Geng. Dlp-based structured light 3d imaging technologies and applications. In *SPIE MOEMS-MEMS*, pages 79320B–79320B. International Society for Optics and Photonics, 2011.

[2] Patrick Wissmann, Robert Schmitt, and Frank Forster. Fast and accurate 3d scanning using coded phase shifting and high speed pattern projection. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2011 International Conference on*, pages 108–115. IEEE, 2011.

[3] Tang Yan, Chen Wen-Jing, Zhang Qiang, Su Xian-Yu, and Xiang Li-Qun. Bp neural network applied to 3d object measurement based on fringe pattern projection. *Optik-International Journal for Light and Electron Optics*, 120(7):347–350, 2009.

[4] Song Zhang and Peisen S Huang. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering*, 45(12):123601–123601, 2006.

[5] Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: whither we are? *Optics and Lasers in Engineering*, 48(IMAC-REVIEW-2009-001):133–140, 2010.

[6] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–195. IEEE, 2003.

[7] Henrik O Saldner and Jonathan M Huntley. Temporal phase unwrapping: application to surface profiling of discontinuous objects. *Applied optics*, 36(13):2770–2775, 1997.

[8] Idaku Ishii, Kenkichi Yamamoto, T Tsuji, et al. High-speed 3d image acquisition using coded structured light projection. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 925–930. IEEE, 2007.

[9] Nikolaus Karpinsky, Morgan Hoke, Vincent Chen, and Song Zhang. High-resolution, real-time three-dimensional shape measurement on graphics processing unit. *Optical Engineering*, 53(2):024105–024105, 2014.

[10] Zhong-wei Li, Yu-sheng Shi, Cong-jun Wang, Da-hui Qin, and Kui Huang. Complex object 3d measurement based on phase-shifting and a neural network. *Optics Communications*, 282(14):2699–2706, 2009.

[11] Steven Fortune and James Wyllie. Parallelism in random access machines. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 114–118. ACM, 1978.

[12] Phillip B Gibbons, Yossi Matias, and Vijaya Ramachandran. Can a shared-memory model serve as a bridging model for parallel computation? *Theory of Computing Systems*, 32(3):327–359, 1999.

[13] Leslie G Valiant. A bridging model for parallel computation. *Communications of the ACM*, 33(8):103–111, 1990.

[14] David E Culler, Richard M Karp, David Patterson, Abhijit Sahay, Eunice E Santos, Klaus Erik Schauser, Ramesh Subramonian, and Thorsten von Eicken. Logp: A practical model of parallel computation. *Communications of the ACM*, 39(11):78–85, 1996.

[15] Kirk W Cameron, Rong Ge, and Xian-He Sun. $log_n p$ and $log_3 p$: Accurate analytical models of point-to-point communication in distributed systems. *Computers, IEEE Transactions on*, 56(3):314–327, 2007.

[16] Xiangzhen Qiao, Shuqing Chen, and Laurence Tianruo Yang. Hpm: a hierarchical model for parallel computations. *International Journal of High Performance Computing and Networking*, 1(1):117–127, 2004.

[17] Kirk W Cameron and Xian-He Sun. Quantifying locality effect in data access delay: memory logp. In *Parallel and Distributed Processing Symposium, 2003. Proceedings. International*, pages 8–pp. IEEE, 2003.

[18] Lu Fengshun. *Research of Parallel Computing on CPU/GPU Heterogeneous Architecture*. PhD thesis, National University of Defense Technology,Changsha,Hunan,P.R.China, 2012.

[19] Jose Luis Bosque and Luis Pastor. A parallel computational model for heterogeneous clusters. *IEEE Transactions on Parallel and Distributed Systems*, 17(12):1390, 2006.

[20] Albert Alexandrov, Mihai F Ionescu, Klaus E Schauser, and Chris Scheiman. Loggp: incorporating long messages into the logp modelłone step closer towards a realistic model for parallel computation. In *Proceedings of the seventh annual ACM symposium on Parallel algorithms and architectures*, pages 95–105. ACM, 1995.

[21] Shane Cook. *CUDA programming: a developer's guide to parallel computing with GPUs*. Newnes, 2013.

[22] C Nvidia. Nvidia cuda programming guide (version 5.5). *NVIDIA Corporation*, 2013.

[23] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.