



OPEN ACCESS

EDITED BY
Zhi-Han Zhu,
Harbin University of Science and
Technology, China

REVIEWED BY
Aili Wang,
Harbin University of Science and
Technology, China
Zhigang Peng,
Beijing University of Technology, China
Wang Gao,
North University of China, China

*CORRESPONDENCE
Li Li,
ll@cust.edu.cn

SPECIALTY SECTION
This article was submitted to Optics and
Photonics,
a section of the journal
Frontiers in Physics

RECEIVED 02 September 2022
ACCEPTED 27 September 2022
PUBLISHED 17 October 2022

CITATION
Chang D and Li L (2022), Edge GPU
cluster processing system for laser
interference image collection.
Front. Phys. 10:1034932.
doi: 10.3389/fphy.2022.1034932

COPYRIGHT
© 2022 Chang and Li. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Edge GPU cluster processing system for laser interference image collection

Dajun Chang^{1,2} and Li Li^{1*}

¹College of Computer Science and Technology, Changchun University of Science and Technology, Jilin, China, ²School of Electrical Information, Changchun University of Architecture and Civil Engineering, Jilin, China

In order to realize the real-time processing and communication of multi-laser interference images, a GPU cluster processing system for laser interference images was designed. The measured target was illuminated by multiple lasers, and the image information of the multi-laser interference fringes was obtained by the CCD. The data of laser interference images from CPU was transmitted by the GPU cluster, and the image feature recognition and transmission were completed through the cluster image processing module. A multi-channel laser interference image transmission algorithm was designed by the multi-core and multi-buffer (MCMB) algorithm. In the experiment, the laser interference images were collected by the CPU, and then the real-time communication of multi-channel images data was completed by the GPU cluster. The packet loss rate experiment showed that when the data traffic reached 110, the data was lost with the traditional UDP communication algorithm, and the slope of the fitting curve was 0.6,053. When the data flow reached 160, the data was lost with MCMB algorithm, and the slope of the fitting curve was 0.2,181. In contrast, the GPU occupancy of this algorithm was improved, and it was still not saturated at 210 data streams. In a word, the system has better optimization effect in real-time processing and communication of multi-laser interference images.

KEYWORDS

laser interference image, GPU cluster, MCMB, data packet drop rate, GPU resource occupancy rate

Introduction

Spectral analysis of laser is an important method to analyze target features [1]. The analysis and recognition of laser interference images are used in many fields, such as laser warning, drug identification, and so on [2]. If a large amount of information can be quickly and accurately classified and communicated, it will be very practical.

With the continuous development of computer science and technology in recent years, its application has become more and more extensive in laser spectrum analysis and laser information processing [3–6]. The data mining technology for laser interference fringes can improve the processing capability of the system. According to the characteristics of frequency and amplitude of the laser interference images, the effective identification of similar targets and the suppression of similar noises can be realized. For the classification of laser

interference images, algorithms such as machine learning [7], deep learning [8], and adaptive parameter adjustment [9] can be used. Liu Xuan et al. [10] used machine learning algorithms to classify and identify hyperspectral vegetation data, and the obtained band characteristic interval provided data support for agricultural vegetation evaluation. The spectral resolution of the system was 5.0 pm, and it took 13.6 s to complete the spectral image transmission of the entire area. The main reason for the slow speed was the large amount of image data transmission. Yuan Shuping [11] used data mining technology to classify laser spectrum data, and greatly improved the classification speed through multi-node parallel processing. It worked well for fluorescence feature data, but for image data, the parallel communication speed was still limited. Yu Xiaoya et al. [12] completed the classification and recognition of algae images by partial least squares algorithm, and its average accuracy rate was over 80%. This method has a good effect on the acquisition of laser interference images, but when there are many features of the laser interference images, the recognition accuracy will drop significantly. Reference [13] proposed a network layering method, which was dynamically divided according to the resource situation of the terminal equipment. Due to the powerful computing power of GPU (Graphics Processing Unit), it had been widely used in laser interference images, artificial intelligence and other fields. According to this design idea, the paper applied the GPU cluster to the fast processing and communication of a large number of laser interference fringe images. Reference [14] used the efficient packet processing capability of DPDK (Data Plane Development Kit [15]) to implement a packet capture system, which greatly improved the data processing capability of the system. For a large amount of laser interference images data, the use of GPU cluster computing can improve the real-time computing capability of the system [16]. GPU cluster has the advantages of strong data processing capability and fast speed. And compared with multi-channel communication, it can also realize the reasonable allocation of core processing system resources according to the data flow state. In order to improve the data processing and communication capabilities of GPU clusters and meet the fast processing requirements for a large amount of laser interference images data, the GPU cluster processing and communication transmission system was proposed based on data stream processing. The efficient transmission of laser interference images data stream had been realized by the system.

System design

The system consists of two parts, the laser interference fringe acquisition module (module A) and the GPU cluster processing module (module B). The laser interference fringes can be dynamically collected in real time by module A. When the multi-laser signals are mixed into the interferometer, the CCD sensor at the end will collect the corresponding interference fringe image. With the increase of time, the system can obtain a large number of interference fringe

images, which constitute the laser interference image dataset. These data can be used for analysis of the laser signal. Images data can be quickly classified, identified and transmitted in real time by module B. The system structure is shown in Figure 1.

As shown in Figure 1, the driver to complete the scanning of the laser intensity and wavelength is controlled by CPU, to realize the wide-band dynamic scanning of the target area. Any material, structure, *et c.* Whose spectral properties need to be obtained can be the target to be measured. The laser interference images data stream from module A is entered into the GPU cluster. When the laser irradiates the target, the diffusely reflected light will be received by the optical system. In order to obtain better laser interference images, multi-channel lasers were used to get interference fringes, so that it can reduce the impact on image processing and communication quality analysis. The optical system converges and collimates the incident light, making it more coherent. The light emitted from the optical system enters the interferometer, thereby forming an interference fringe image, which is imaged on the CCD sensor. The image is transmitted to the GPU cluster through the CPU, and the extraction of image features and data analysis are completed in the cluster operation. Finally, the signal is sent to the end user through the wireless reflection module. The innovation in the architectural design of this system is to match the processing power of the GPU cluster with a large number of laser interference fringe images. In the matching process, edge computing was used to realize the self-organization of terminal contacts, so as to optimize the data flow distribution in the process of laser interference fringe image communication.

Principles of the system

Laser interference images

The Fourier transform interferometer was used to complete the fast spectral analysis of the incident laser light. The interference module can generate two beams of coherent light, and control the optical path difference to make them interfere, thereby obtaining interference images. The Fourier transform interferometer was realized by frequency demodulation.

Assuming that the incident laser wave number is ν , it is irradiated on the beam splitter (BS) plate with reflectivity r and transmittance t . The optical path difference of the two beams is l , so the interference spectrum image can be expressed as

$$\begin{cases} I(l, \nu) = 2RTB_0(\nu)(1 + \cos \varphi) \\ \varphi = 2\pi \times \nu \times l \end{cases} \quad (1)$$

Among them, $R = r \cdot r^*$, which is the reflectivity of BS. B_0 is the incident beam intensity, and ν is the wave number. $T = t \cdot t^*$, which is the transmittance of BS. After the laser was selected, ν

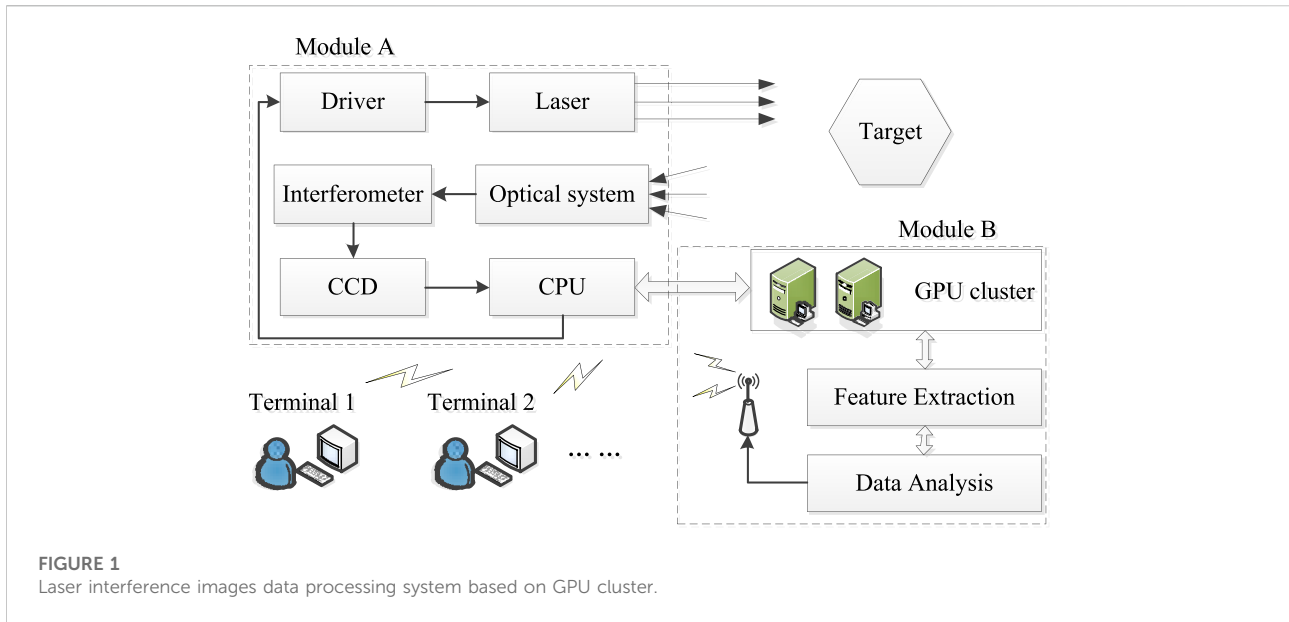


FIGURE 1
Laser interference images data processing system based on GPU cluster.

became a constant, so after substituting the constant, the independent variable was only l . The light source is usually an extended-spectrum light source in practical applications. It can be seen that in the full spectrum range, the interference images should be the integral of all wave numbers, and it is

$$I(l) = \int_0^\infty B(\nu) \cos(2\pi\nu l) d\nu \quad (2)$$

Among them, $B(\nu) = 2RTB_o(\nu)$, which is the light source intensity obtained by correcting various influences. The above formula represents an input beam, which wave number is ν after passing through the interferometer. So the image surface curve of the relationship between strong $I(l)$ and optical path difference can be calculated. When the optical path difference of the two beams of light is $\Delta = (Z1-Z2)$, $Z1$ and $Z2$ represent the optical path distances of the laser light in the two directions in the interferometer, respectively. The light intensity measured by the detector is the sum of all frequency light intensities, which is

$$I_R(\Delta) = 2TR \left[\int_0^\infty E_{0(\sigma)}^2 d\sigma + \int_0^\infty E_{0(\sigma)}^2 \cos(2\pi\sigma\Delta) d\sigma \right] \quad (3)$$

Among them, R is the reflectance and T is the transmittance.

Because the interference images are a real even function, the cosine change is changed to a Fourier change, and the final spectral intensity is

$$B(\sigma) = \int_{-\infty}^\infty I(\Delta) e^{-i2\pi\sigma\Delta} d\Delta \quad (4)$$

Then the interference images can be obtained by

$$I(\Delta) = \int_{-\infty}^\infty B(\sigma) e^{i2\pi\sigma\Delta} d\sigma \quad (5)$$

GPU cluster

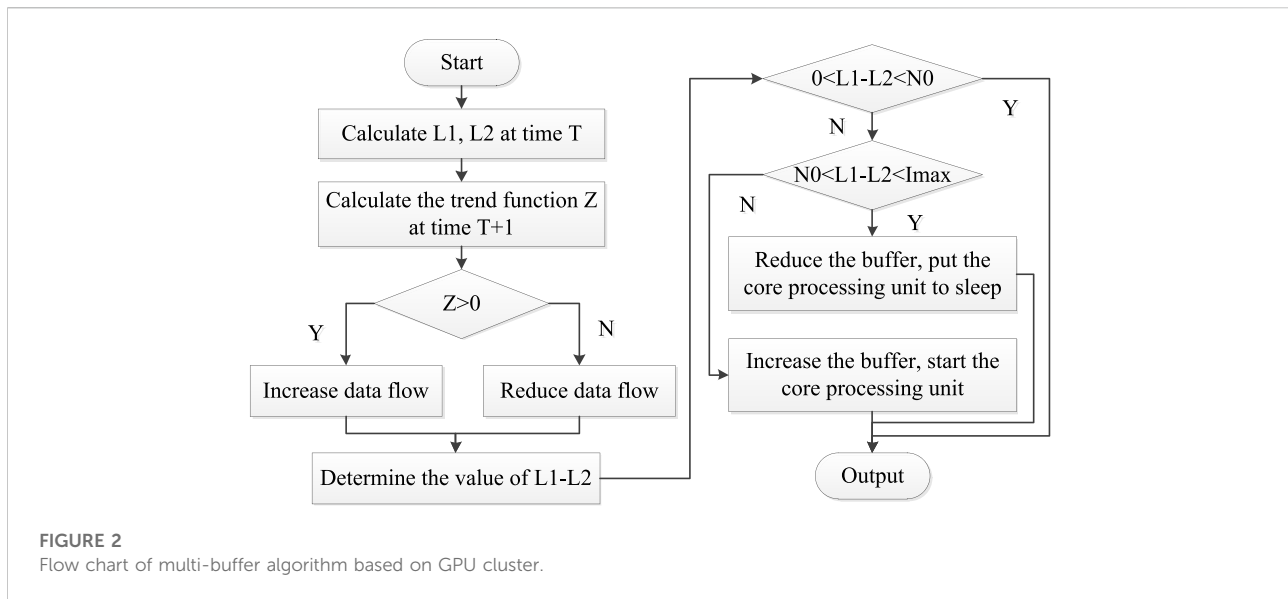
From the perspective of hardware architecture, the GPU cluster only connects the GPU as a peripheral device through the high-speed PCI bus to the inside of the node. The addition of GPU makes the cluster presentation node's internal computing resources heterogeneous. Since the GPU not only has computing resources that are heterogeneous to the CPU, the data transfer between the CPU and the GPU must be performed explicitly under the control of the CPU. As a computing resource for multi-scale data parallel computing, GPU clusters have multi-level parallel computing capabilities. GPU clusters can not only support single-program multi-data and multi-program multi-data computing capabilities of conventional granularity, but also support more fine-grained large-scale data-oriented single-program multi-data and single-instruction multi-data computing capabilities. It provides powerful large-scale data parallel processing capabilities to the cluster.

Since a large number of images data streams are generated during the real-time acquisition of laser interference images, it has significant advantages that the use of GPU clusters to classify, identify and communicate laser interference images.

Algorithm design

Data transfer protocol

The data of the laser interference images data stream was composed of a steady stream of numerical matrix data, including a large number of floating point numbers. In order to reduce the overhead of the data to be transmitted and realize the rapid classification and communication transmission of laser



interference images, a message-based data transmission protocol was designed to reduce the amount of data, and according to the content of the protocol, the receiving end can directly obtain it with only one time.

The system maintained multiple cyclical laser interference images data stream buffers, and the number of buffers was dynamically adjusted with the changes of the laser interference images data stream. The circular data stream buffer was used to cache the laser interference images in a classified manner, and the images of each frame belonged to the processing cycle. DPDK used the memory pool to manage the data. After the data received, the received core applied for the object from the memory pool. After completing the data encapsulation, it cached the data packet to the write data packet according to the data packet identifier and processing cycle. In the cache line corresponding to the pointer, each frame was processed for a data stream with the same processing cycle. After the data was buffered, the communication was in a one by one correspondence with the buffer table of the data stream.

MCMB algorithm

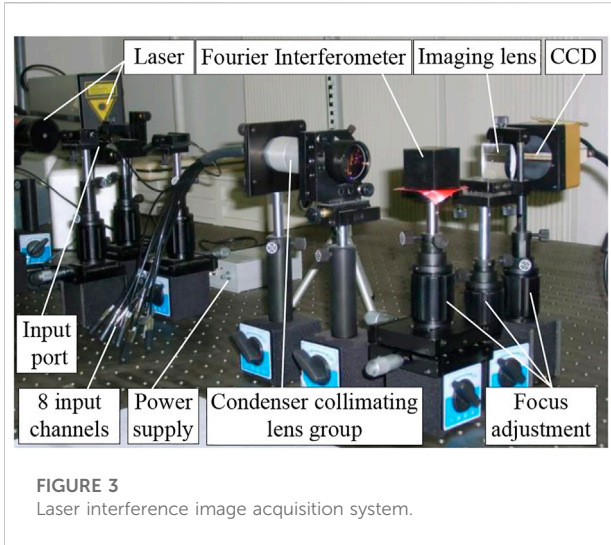
In order to reduce the occupation of shared resources and improve the efficiency of data processing, the affinity of the data flow and the core was set. The data packets belonging to the same data flow were handed over to the designated core for processing. Among them, each forwarding core managed a circular laser interference images data stream buffer, which was responsible for forwarding the laser interference images in the managed buffer, and the selection of multi-frame images was dynamically selected according to the load balance of the ports. Thus, MCMB was constructed, and its flow is shown in Figure 2.

L1 and L2 represent two adjacent laser interference images, and T represents the sampling time for collecting the laser interference images. Z represents the data flow trend line, which is quantitatively analyzed by GPU work resource occupancy. And according to its size to increase or decrease the data flow control. N0 represents the amount of image feature difference, which is used to measure the total amount of feature data in the image. Imax represents the feature data amount of the laser interference image including the most feature data. The acquisition unit completed the acquisition of the interference fringe image data stream from the CCD, and performs matching calculation on the two images whose time interval was T. When the trend function $Z > 0$, increase the data flow. Conversely, when the trend function $Z < 0$, the data flow is reduced. This enables real-time adjustments to the data flow. The difference calculation was performed on the two sets of interference fringe image data, and the interval range was judged. When it was between $(0, N0)$, it was directly reserved and output. When it was between $(N0, Imax)$, reduce the buffer and put the core processing unit to sleep.

Experiments

Acquisition of laser interference images

Under laboratory conditions, real-time interference image processing and communication were performed for the aliased spectral data of three lasers at 639.1 nm, 654.5 nm and 660.1 nm. Different lasers entered the irradiation window at the same time through different pigtailed, and then the beamforming and collimation module completed the optical path shaping. A Fourier transform interferometer



was fixed at 4 cm from the output end, and an imaging lens was installed between the output end of the interferometer and the camera to focus the interference fringe image on the CCD sensor. The experimental structure is shown in Figure 3.

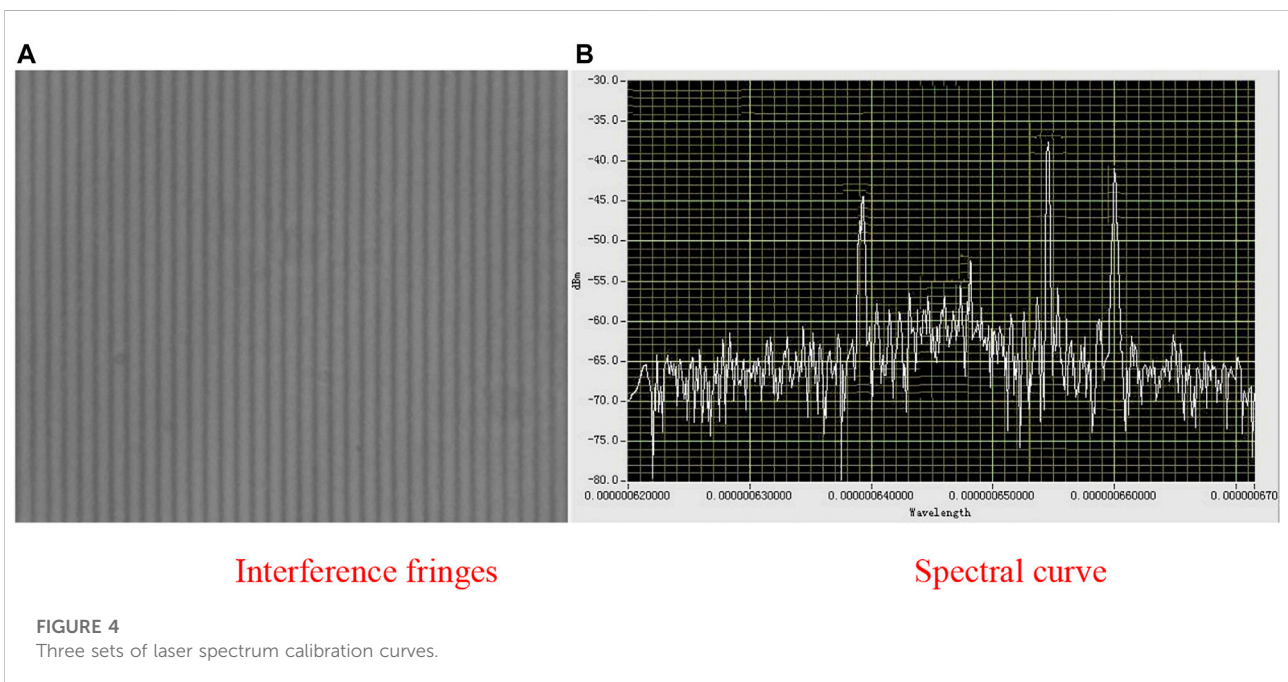
The aliasing spectrum is calibrated with a Q8344A spectrometer. The laser spectrum curve is shown in Figure 4.

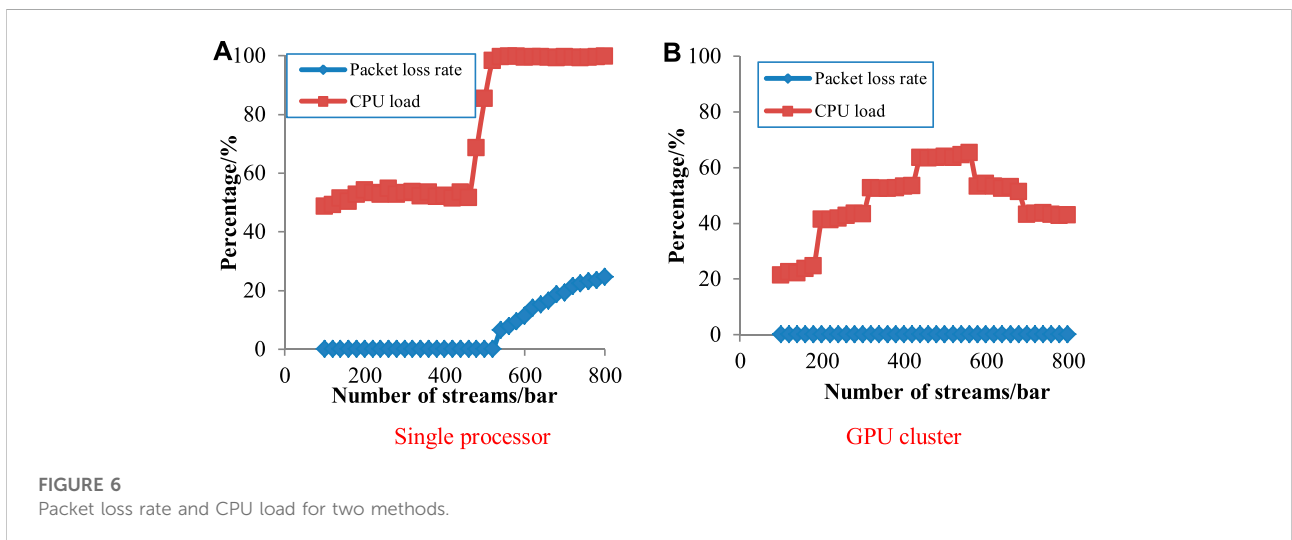
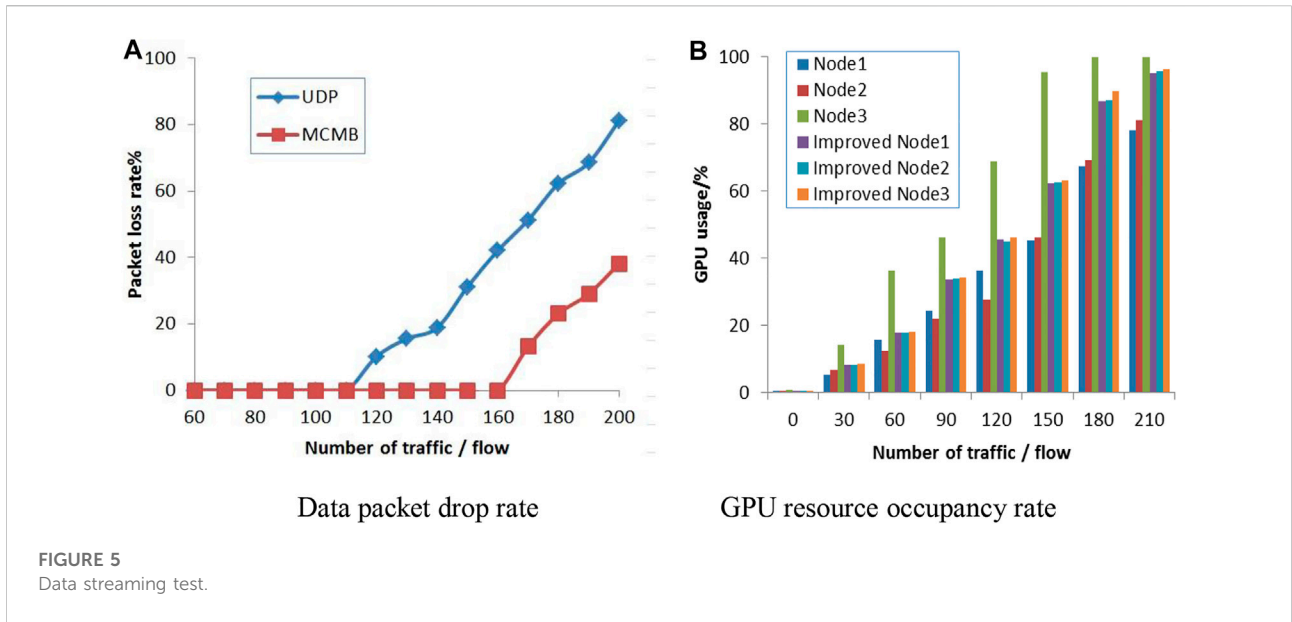
From Figure 4, although the interference fringes entered the CCD sensor at the same time, there was a certain aliasing phenomenon, but the position of the main laser peak after inversion could still be easily identified, and the optimal response value was about -37.6 dB.

Data streaming test

In order to verify the performance of the scheme, eight ordinary servers were used to simulate eight channels of laser interference images to transmit data streams in real time. Three computer nodes were used to build a GPU cluster, and each node was configured with a gigabit network card. Each server can simulate multiple terminal devices at the same time. The GPU cluster was built by computing nodes with GPUs installed. The number of computing nodes in the cluster can be adjusted according to changes in the number of terminal devices in the actual application scenario. The traditional UDP (User Datagram Protocol) algorithm was selected to compare with MCMB algorithm. Comparing the algorithms applicable to different systems can characterize the pros and cons of the algorithms from the program running effect. In the next section, we will also compare the resource usage of hardware modules. When the flow of the data stream was continuously increased, the test result of the data packet drop rate is shown in Figure 5. At the same time, the GPU resource occupancy rates of the traditional communication algorithm and the optimized communication algorithm were compared, as shown in Figure 5.

In Figure 5, when the number of data streams exceeded 110 by the traditional UDP algorithm, packets started to drop during the transmission of images data. With the further increase of data traffic, the packet loss rate also increased. The change of the test curve is close to linear, which shows that the data communication capability of the system has reached its limit. As the data flow continues to grow, the number of lost packets will also increase linearly. The calculation shows that the fitting curvature of the curve is 0.6053. When the number of data





streams exceeded 160 by the MCMB algorithm, packets started to drop during the transmission of images data. Its packet loss curve is also close to linear, and the curve fitting curvature is 0.2181. It verifies that the improved algorithm has lower packet loss rate and stronger communication ability. In Figure 5, the GPU occupancy ratios on the three data nodes using the traditional method are significantly different, and the resource occupancy ratio of the third node is relatively high. As the data traffic increases, the third node overflows when the traffic reaches between 150 and 180, resulting in significant packet loss. After the optimization of this algorithm, the GPU occupancy difference of the three nodes is very small. They evenly distribute the total amount of data flow across the three nodes. When the data traffic reaches 210, the GPU occupancy

rate of the three nodes reaches more than 95% without overflow, which avoids packet loss caused by excessive data traffic. At node 3, many data streams passed through this place, so the communication pressure was high and the resource occupancy rate was high. This also shows that when the communication network data flow is not evenly distributed, it will lead to a decrease in communication capacity.

Analysis of packet loss rate and load status

In order to verify that the GPU cluster can greatly improve the data transmission capability of laser interference images, a comparison test was carried out with the communication unit of

a single processor. The number of image data streams was increased from 100 to 800, and the packet loss rate and CPU core load status of the system were recorded for each additional 100. The test results were shown in Figure 6.

As shown in Figure 6, the test results of a single processor showed that when the data flow exceeds 520, the CPU load increased significantly. When there were more than 550 data streams, the CPU load basically remained around 99.5%. At the same time, the packet loss phenomenon occurred in the system, and the packet loss rate increases with the increase of the number of data streamed, reaching 24.6% in 800 data streams. As shown in Figure 6, the test results of the GPU cluster showed that there was no packet loss during the entire process from 100 to 800 data streams. After the load of each CPU in the system was increased by more than 60%, the data processing task was redistributed, so that the original CPU load was gradually reduced. When the offload of data processing was increased, the average CPU load was only 43.2% at 800 entries.

Conclusion

Aiming at the problem of fast processing and communication, a multi-laser interferometric image real-time processing and communication system was proposed based on GPU cluster. The multi-core multi-buffer algorithm was designed to improve the resource allocation of multi-nodes. It is a certain contribution to the resource optimization problem of simultaneous processing of multiple data streams by the GPU cluster design and improved algorithm proposed in the paper. The feature of the paper is to apply the image synchronization processing capability of GPU cluster to feature data extraction and real-time data communication of multi-laser interference images. However, it did not discuss the attenuation effect of the optical structure of the system on signals of different

References

- Palomares I, Martinez L, Herrera F. A consensus model to detect and manage non-cooperative behaviors in large scale group decision making. *IEEE Trans Fuzzy Syst* (2014) 22(3):516–30. doi:10.1109/tfuzz.2013.2262769
- Leira R, Julian-Moreno G, Gonzalez I, Gomez-Arribas FJ, Lopez de Vergara JE. Performance assessment of 40 Gbit/s off-the-shelf network cards for virtual network Probes in 5G networks. *Computer Networks* (2019) 152(4):133–43. doi:10.1016/j.comnet.2019.01.033
- Zhao Z, Bai Z, Jin D, Qi Y, Ding J, Yan B, et al. Narrow laser-linewidth measurement using short delay self-heterodyne interferometry. *Opt Express* (2022) 30(17):30600–10. doi:10.1364/OE.455028
- Cui L, Chen Q. Dynamic monitoring model based on DPDK parallel communication. *J Comp Appl* (2020) 40(2):335–41. doi:10.11772/j.issn.1001-9081.2019081405
- Gao P, Zhao D, Chen X. Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework. *IET Image Process* (2020) 14(7):1257–64. doi:10.1049/iet-ipr.2019.0588
- Ponsard R, Janvier N, Kieffer J, Houzet D, Fristot V. RDMA data transfer and GPU acceleration methods for high-throughput online processing of serial crystallography images. *J Synchrotron Radiat* (2020) 27(5):1297–306. doi:10.1107/s1600577520008140
- Gao-xing Z, Hu S-H, Zhang J-X. Transfer protocols for data center networks: A survey. *J Comp Res Development* (2020) 57(1):74–84. doi:10.1007/s11107-015-0550-y
- Raychaudhuri D, Nagaraja K, Venkataramani A. Mobilityfirst: A robust and trustworthy mobility-centric architecture for the future internet. *SIGMOBILE Mob Comput Commun Rev* (2012) 16(3):2–13. doi:10.1145/2412096.2412098

wavelengths, and the corresponding correction parameters of the GPU cluster. It will also be the main content of further in-depth research in the later stage of the thesis.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DC is responsible for experimental testing and writing papers. LL is responsible for the design of the system structure and simulation analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

9. Laoutaris N, Che H, Stavr Akakis I. The LCD interconnection of LRU caches and its analysis. *Perform Eval* (2006) 63(7):609–34. doi:10.1016/j.peva.2005.05.003
10. Liu X, Zhang Y, Teng Y. Estimation of vegetation water content based on Bi-inverted Gaussian fitting spectral feature analysis using hyperspectral data. *Remote Sensing Technology Appl* (2016) 31(06):1075–82. doi:10.11873/J.ISSN.1004-0323.2016.6.1075
11. Yuan S. Research on pattern recognition of laser fluorescence spectrum data in big data background. *Laser J* (2018) 39(5):124–7. doi:10.14016/j.cnki.jgzz.2018.05.124
12. Yu X, Zhang Y, Yin G, Xiao X, Zhao N, Duan J, et al. Feature wavelength selection of phytoplankton fluorescence spectra based on partial least squares. *Acta Optica Sinica* (2014) 34(9):0930002–7. doi:10.3788/aos201434.0930002
13. Qu-cheng C, Qing-kui C. Deep neural network layering strategy for embedded devices. *J Chin Comp Sys Tems* (2019) 40(7):1455–61. doi:10.0012/20(2019)07-1455-07
14. Rui-lin L, Jun-feng L, Dan L. Realtime capture of high-speed traffic on multi-core platform. *J Comp Res Developer* (2017) 54(6):1300–13. doi:10.7544/issn1000-1239.2017.20160823
15. Matthew Helm J, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, et al. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Curr Rev Musculoskelet Med* (2020) 13(10):69–76. doi:10.1007/s12178-020-09600-8
16. Nathali Silva B, Khan M, Kijun H. Internet of things: A comprehensive review of enabling technologies, architecture, and challenges. *IETE Tech Rev* (2018) 35(2):205–20. doi:10.1080/02564602.2016.1276416