Real-time Acoustic Vibration Detection and Recovery Method by Performing High-speed Image Processing on Laser Speckle Images

Nan Wu

Supervisor: Prof. Shinichiro Haruyama

September 2021

Graduate School of System Design and Management

Keio University

ABSTRACT

Vibration is ubiquitous in nature and human's daily life. The monitoring and measurement of vibration is of great significance. Among the many vibration measurement techniques, optical vibration measurement offers advantages such as non-contact, no harm to the object's structure, and high accuracy. Compared to the contact measuring method which uses an electric sensor, the optical method offers a broader application base.

Sound, as a special kind of vibration, can also be measured with optical approaches. The optical sound measurement method uses optical instruments to measure the status of the vibrating sound source or objects actively and directly, thus, it can avoid crosstalk of different sound source signals, as opposed to traditional electric microphones; Furthermore, the optical measurement method offers a longer measurement distance, thus it can extract the sound of targets with a long distance. In the past, many optical-based audio signal measurement methods have been proposed. Among them, the method based on laser speckle and image sensor is an ideal solution since it has the characteristics of high measurement accuracy, long detection distance, and ability to work in dark environment. However, since the frequency of audio signal is usually much higher than the image processing speed, the real-time restoration of audio signals has never been realized.

In this research, a real-time acoustic vibration detection and recovery method by performing high-speed image processing on laser speckle images has been proposed. The proposed method can detect sound vibration of object and recover the audio signal in real time. Compared with the previous speckle sound measurement research which mainly focused on applications of non-real-time sound recording, our research allows real-time recovering of audio signals, making it suitable for applications such as visual microphones and vibration monitoring usages.

ACKNOWLEDGEMENTS

This research has been conducted at Graduate School of System Design and Management.

First, I want to thank my supervisor Prof. Haruyama Shinichiro. During my Ph. D. research, Prof. Haruyama not only gave me guidance and help on specific technical issues, but also cultivated my ability to discover, think and solve problems independently. All these have benefited me a lot. I also want to thank him for his support and help during my study abroad in Japan. At the same time, I am also very grateful to the Professor Keiichiro Kagawa from Shizuoka University, Associate Professor Hideaki Uchiyama from Nara Institute of Science and Technology and Professor Tetsuro Ogi from Keio University for agreeing to review my theses.. Their professional and insightful comments are very crucial to my thesis.

Next, I want to offer my thanks to Keio University for granting me the Keio University Doctorate Student Grant-in-Aid Program in 2018, 2020, and 2021. I really enjoyed my time at Keio University, and it will remain one of the best experiences of my life. My sincerest wishes for this long-history university as it grows stronger and stronger.

Last, I would like to express my gratitude to my family and friends who have always supported me. Especially, the highest gratitude goes to my parents, I could not become who I am today without your support. Thank you for your support throughout my life.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background of the Research

Vibration phenomena are everywhere in nature and humans' daily live. Intuitively speaking, vibration is the reciprocating motion of an object within a certain range. From the point of the motion scale, vibration can be divided into vibration at the micro level (such as the vibration of microscopic particles) and vibration at the macro level (such as the vibration of a machine). From the speed of reciprocating motion, vibration can be divided into low-frequency vibration and high-frequency vibration. Vibration has its positive side and its harmful side [1]. Using vibration, people have created useful machines such as facial gun or microwave oven. On the other hand, vibration can be harmful in some scenes. For example, in the measurement and industrial manufacturing field, the vibration of the instrument may reduce the accuracy of the measurement, and the vibration of the machine may cause the reduction of the surface finish of the processed workpiece [2]. In the construction field, the structure of a building may be damaged under the long-term impact of vibration generated by construction machines or vehicles [3-5]. In addition, vibration can also be used as an important parameter to measure whether an object is in good condition. Some machines, such as generators, perform vibration when they are working, and the amplitude and frequency are kept with a certain value [6]. If the value of the vibration is abnormal, such as a sudden increase in amplitude or a sudden increase in frequency, it may indicate a problem with the operating state of the machine. In summary, the monitoring of vibration is of great significance in industrial production and daily life.

Vibration measurement methods can be divided into two types: contact detection and non-contact detection [7, 8]. From the perspective of the technology used, vibration measurement methods can be divided into mechanical vibration measurement methods, electrical vibration measurement methods and optical vibration measurement methods. Among them, the mechanical vibration measurement method is the earliest. The mechanical method is a method of recording vibration signals using the principle of lever transmission or inertial reception. The commonly used instruments in this method include Tastograph, Stoppani vibrograph, and Geiger Vibrograph, as shown in Figure 1.1 [9, 10]. This type of instrument can directly record the vibration waveform, which is convenient for observing and analyzing the vibration amplitude, frequency, main harmonic component frequency and other parameters. They have the advantages of simple use, convenient carrying, no power consumption, and strong anti-interference ability. However, due to their shortcomings such as low sensitivity and narrow frequency range, such instruments are used less and less in engineering. The electrical measurement method uses sensors to convert mechanical vibration (displacement, speed, acceleration, force) into electrical energy (charge, voltage, etc.) or changes in electrical parameters (resistance, capacitance, inductance, etc.), and then use electrical energy measurement and analysis equipment to analyze the signal [11-13]. In Figure 1.2 a contact type electrical vibration measurement equipment is presented [14]. Optical measurement is a method of converting mechanical vibrations into optical signals, which are amplified by an optical device for recording and measurement [15-17]. Optical measurement method usually uses laser as the light source and uses the principle of optical interference to measure the vibration of the object. Since it has the advantages of high accuracy, high sensitivity, non-contact, long-distance and full-field measurement, optical vibration measurement has become very useful in special environments and long-distance measurements. The commonly used

optical measuring instruments include laser single-point vibrometer, laser Doppler scanning vibrometers and so on, as shown in Figure 1.3 [18-20].

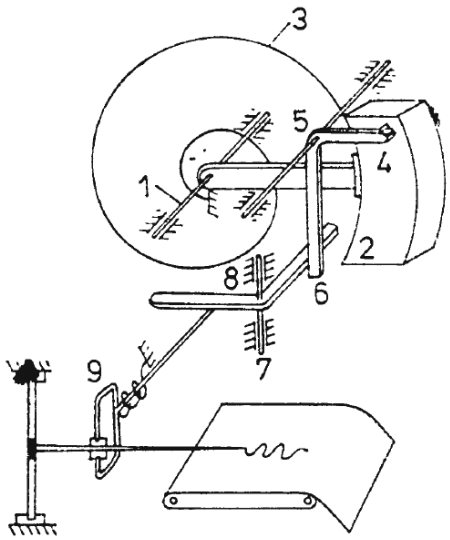

Figure 1.1 Structure of Geiger Vibrograph. 1. Axle. 2. Pendulum. 3. Coil spring. 4-9. Mechanical linkage

Figure 1.2 Contact type vibration measuring device

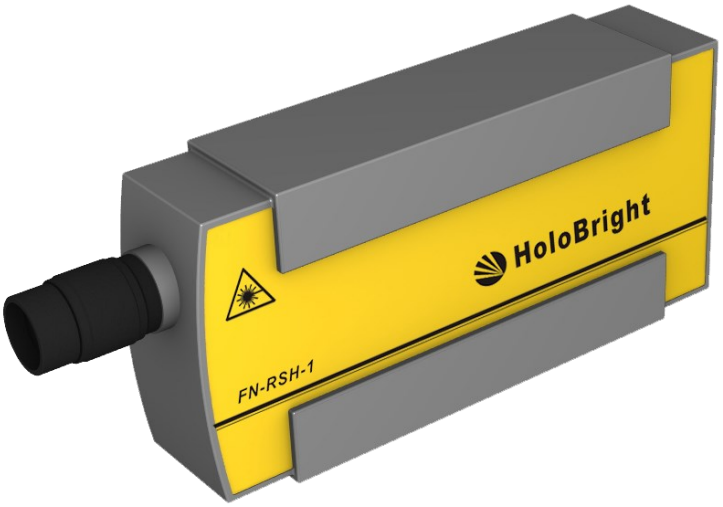

Figure 1.3 Free-space Single-point Laser Doppler Vibrometer

Generally, both the mechanical method and the electrical method belong to the contact vibration measurement method. This method requires the sensing device to be attached to the surface of the object to be measured. There is no doubt that the contact measurement method will affect the vibration characteristics of the object itself, thereby causing measurement errors. In addition, in some special cases, the measured object may be in a hard-to-reach place, such as the object is in a very far or very high position. It would be very difficult and dangerous to arrange sensors in this case. Finally, when it is necessary to measure the vibration of multiple objects or multiple positions on the surface, setting of a large number of sensors is very complex. These are the shortcomings of the contact measurement method. On the other hand, the optical measurement methods can measure object motion without touching object. The untouched measurement method will not affect the vibration characteristics of the object and the vibration information of the object can be extracted non-destructively. Second, the optical measurement method is usually based on the principle of optical interference measurement, theoretically it can achieve sub-nanometer measurement accuracy. Finally, the optical measurement method does not need to arrange sensing equipment in advance and can monitor

the vibration of distant objects efficiently and in real time.

As we all know, sound is generated by the vibration of an object, so the measurement of sound signals belongs to the vibration measurement and has the generality of vibration measurement. Of course, as a subset of vibration, sound vibration also has its particularity. Generally, the vibration amplitude of the sound signal is small, and the vibration frequency falls within a certain interval. For human, the range of audible sounds is between 20 to 20000Hz [21, 22].

The research content of this research is using optical method to measure audio signal and replay it in real time. It can not only restore audio signals in real time, but also can be applied in general vibration monitoring fields.

1.2 Review of Previous Work

Optical vibration measurement is a technology that uses optical means to sense vibration information of the interested object. In the past a significant number of studies on optical vibration sensing method have been carried out. These methods use different types of light source and sensors. Therefore, there performances are also different in terms of detection methods, detection accuracy, and response time. Here some typical optical vibration sensing studies and their features are introduced in the following part.

1.2.1 Visual microphone

First, we introduce the method of directly using the camera to capture sound vibration. This method uses passive illumination, that is, the setup does not contain a light source module, and simply uses a high-speed camera to capture object vibration [23-27]. The representative research is MIT's visual microphone [28]. In 2014, the researchers from MIT CSAIL

(Computer Science & Artificial Intelligence Lab, Massachusetts Institute of Technology) proposed an experiment setup called “Visual Microphone” that can passively recover sound from video. Figure 1.4 shows the schematic of the experiment setup.

Figure 1.4 Example of experiment setup of Visual Microphone from MIT

They use an ultra-high-speed camera (Phantom V10) to capture video of object illuminated by natural light. When the speaker plays sound, the object (a bag of chips) will vibrate, and the video could be recorded by the camera. The Phantom V10 high speed camera can achieve 2200 FPS with a spatial resolution of 700×700 pixels, which means it can sample vibration signal up to 1100Hz according to Nyquist–Shannon sampling theorem. The captured video containing the detail of the object vibration is then denoised and analyzed by their algorithm and recover the original sound.

The setup does not require active illumination, so using this approach to “capture sound” is very simple and flexible. People just use a camera to record video when they need it, and then use a computer to process the video to obtain the sound. However, this approach has some

disadvantages. First, processing the video is very time-consuming. It is impossible to simultaneously recover the sound while capturing videos. Besides, although it is claimed that this detection method does not require additional light sources, sufficient illumination is still needed to capture clear detail of the object. Otherwise, the recovered sound will be very noisy. One evidence is that although Phantom V10 camera can reach a very high frame rate, the researchers choose 2200 FPS in most experiments. Higher frame rate may bring broader sampling frequency range, but it also results in reduced exposure time and therefore more image noise. In the end the result is noisier than the results at 2200 FPS. It is obvious that the passive recovery approach will not work in a dark environment.

1.2.2 Lamphone

In 2020, the researchers from Ben-Gurion University of the Negev, Israel proposed a sound detection method called “Lamphone” that passively recovers sound from light bulb vibrations [29]. Figure 1.5 shows the concept of their proposal.

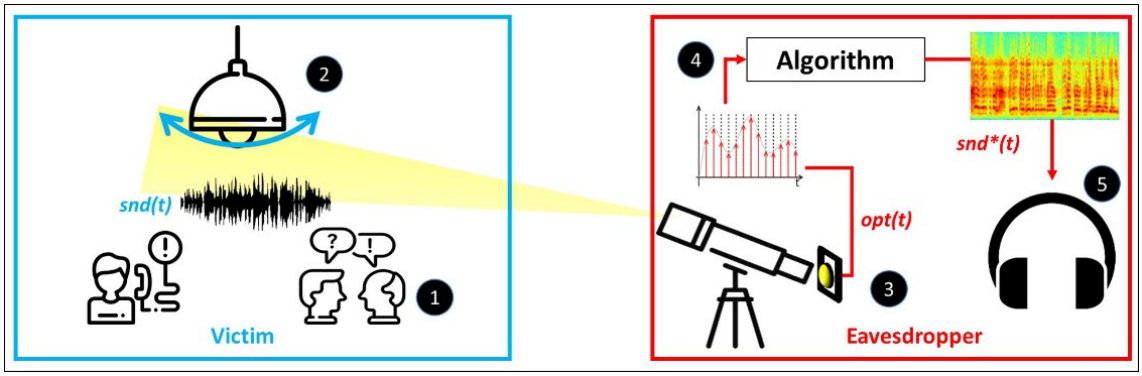

Figure 1.5 Schematic of Lamphone

They use telescope lens to focus the light from the light bulb onto an electro-optical sensor. When people speak, the fluctuations in the air pressure on the surface of the hanging bulb will cause the bulb to vibrate very slightly, which results in the variation of the received signal at

electro-optical sensor. The researchers also proposed an algorithm to recover sound from the optical measurements obtained from the captured optical signal.

Like Visual Microphone, Lamphone is also a passive detection method. But it uses electro-optical sensor for detection rather than image sensor. Processing signal of electro-optical sensor is much faster than processing image, thus the Lamphone can be achieved in real time. Actually, one usage of Lamphone is monitoring victim's talking. However, Lamphone requires a strong light source near the sound source, otherwise it cannot recover the sound. The method is suitable for surveillance, but it is not flexible and cannot be used in other scenes.

1.2.3 Laser speckle method

As mentioned above, one common point of Visual Microphone and Lamphone is that the hardware does not contain additional light source. The object illumination is done by natural light in the environment. Usually, this kind of method is called "passive illumination". Correspondingly, active illumination approach contains not only sensor module but also light source in the setup, and coherent light is usually adopted as the light source. This method utilizes coherent light interference phenomenon to measure the object vibration. One classic scheme is vibration measurement method based on laser speckle images.

The principle of laser speckle measurement method is as follows. When a laser beam hits on the object's surface, the laser beam will be scattered into multiple reflection lights since most of the surfaces in our daily life are optically rough surface. The multiple reflection lights interfere with each other and form a high-contrast pattern, which is called "speckle pattern" [30]. The speckle pattern could be observed by human eyes, or image sensor. The speckle pattern is very sensitive to object motion. When the object vibrates, the speckle pattern shows same motion. Therefore, one can extract sound from the vibrating object using observed

speckle sequence. The details, such as characteristics and the motion model of the speckle image will be introduced in Chapter 2. In the past there were many studies using laser speckle images to measure object vibration and recover audio signal. The basic principle of each study is almost same, which is object-speckle mapping. However, those methods are quite different in speckle image processing method. Here several typical laser-speckle-based measurement methods are introduced.

1.2.3.1 Speckle motion method

As stated above, when object vibrates, the scattered speckle pattern will response and shows corresponding motion. The dynamic speckle image sequence contains the information of object motion, such as frequency and amplitude. Therefore, analyzing the speckle motion could help us restore the original signal of sound. The representative research among all motion analysis methods was proposed by Zeev Zalevsky in 2008 [31]. Figure 1.6 shows the long-range talking detection using this method, and Figure 1.7 shows the experimental setups of the detection method.

The major contribution of Zeev Zalevsky's research is that they proposed the classical setup for measuring vibration using laser speckle image and imaging sensors. Moreover, they also established a motion model, analyzed the relationship between object motion and speckle image motion. They proved that if the camera is strongly defocused, the motion of the captured speckle images is only determined by the tilting motion of the object. The experimental results presented the capabilities of the approach for remote speech recording, extraction of heart beats temporal signature, cellular phone taping and hearing through wind.

Zeev Zalevsky's contributions laid the foundation for the following people's work. However, they calculate the displacement of speckle image sequence by analyzing the image cross-

correlation. This method is time-consuming and thus the image processing speed is slow. This method is not able to recover the audio signal in real time.

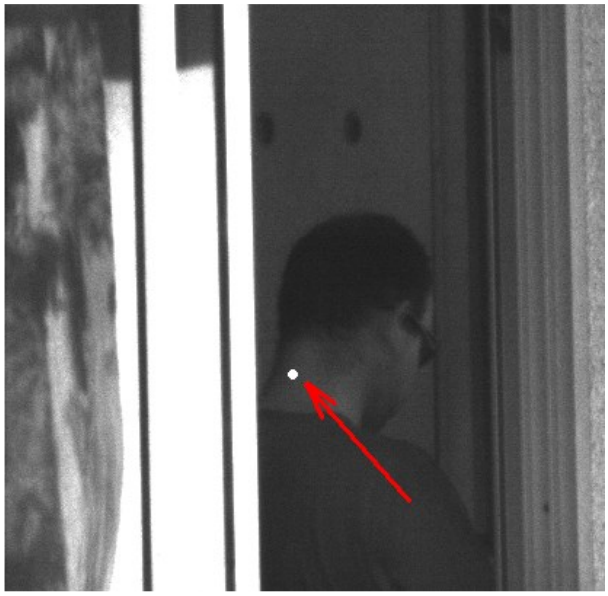

Figure 1.6 Listening to talks from the back part of the neck using laser speckle images

Figure 1.7 The experimental setups for far range detection using laser speckle images

1.2.3.2 Gray value variance method

Analyzing the displacement of dynamic speckle image sequence can restore the original

signal, but usually displacement calculation is time-consuming. In 2014, Ziyi Chen proposed an audio signal reconstruction method based on gray value variations of some selected pixels (called “seed points” in the published paper) from laser speckle images [32]. Similar research can be also found in [33]. Figure 1.8 shows the flow chart of this method. After recording the dynamic speckle images, they selected some seed points from the images. Since the speckle images is vibrating, the intensity of each seed point is temporally varying. Figure 1.9 shows the pixels’ intensity along the speckle moving direction. Basically, the points that located at the middle position of neighbor minimum (e.g., P1, P3) and maximum (e.g., P2, P4) points are suitable for being seed points. The intensity variation of those seed points could be recorded to obtain the original audio signal.

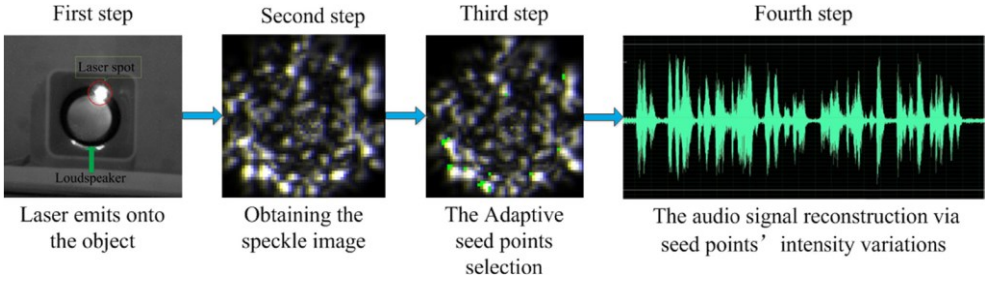

Figure 1.8 The flow chart of extracting audio signal using gray scale variation from selected seed points of laser speckle images

Since the audio signal is reconstructed based on pixel’s gray value variation, this approach is faster compared with calculating speckle displacement. However, the gray-value method requires a speckle pattern with a linear intensity distribution in the direction of motion. It is difficult to perform well when the speckle brightness is low, or the image size is small. In general, using gray-value variation to restore audio signal is an efficient way, but this method is not applicable in all situations.

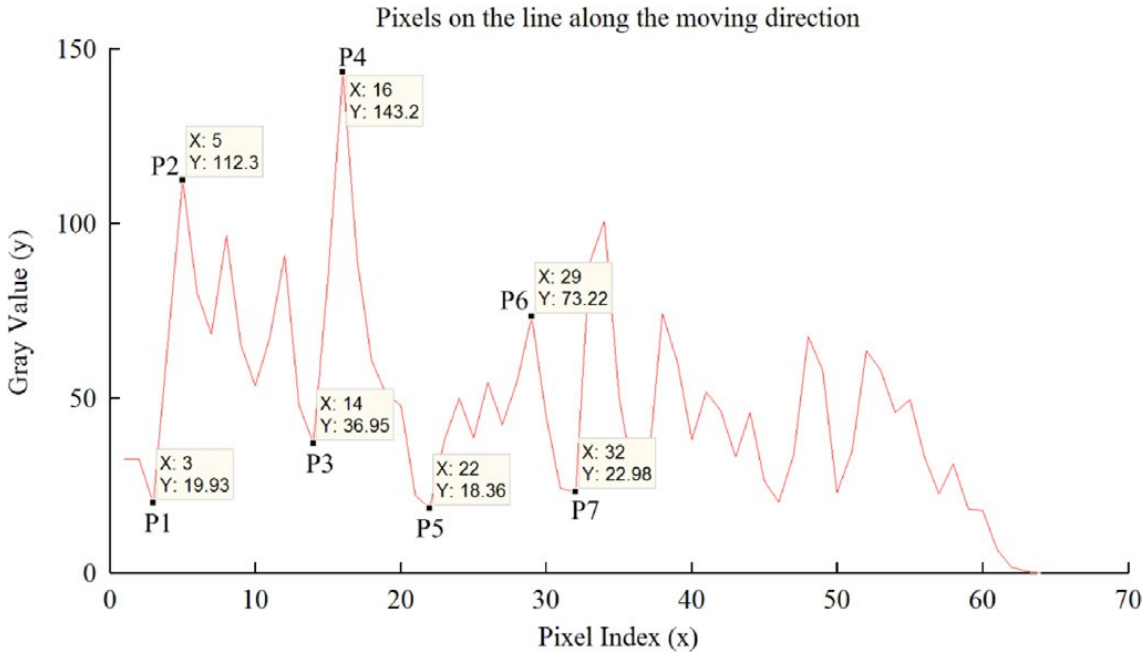

Figure 1.9 The pixels on a line along the moving direction in the speckle image

1.2.3.3 Single pixel(non-imaging) method

In 2014, Silvio Bianchi proposed a method that can sense surface vibrations by measuring signal variations from a single pixel detector that collects a small portion of the scattered light [34]. Figure 1.10 shows the schematic of the proposed method.

The single pixel detector method could be regarded as a laser-speckle-version Lamphone. It uses the received signal's variation of a single pixel to present the original signal. Different from Lamphone, the method actively adopts coherent light for illumination and uses part of the scattered laser speckle image to measure the surface vibration. Thus, the flexibility and the sensitivity of the method is superior to Lamphone. Besides, the sampling speed of single pixel is usually very fast, as well as the processing speed of electronic signal. Therefore, single pixel detector is an ideal solution for real time recovery of sound. The disadvantage of this approach is that single pixel detector has only one pixel. Unlike image sensor, it cannot capture images, which means the environment information is lost in this method.

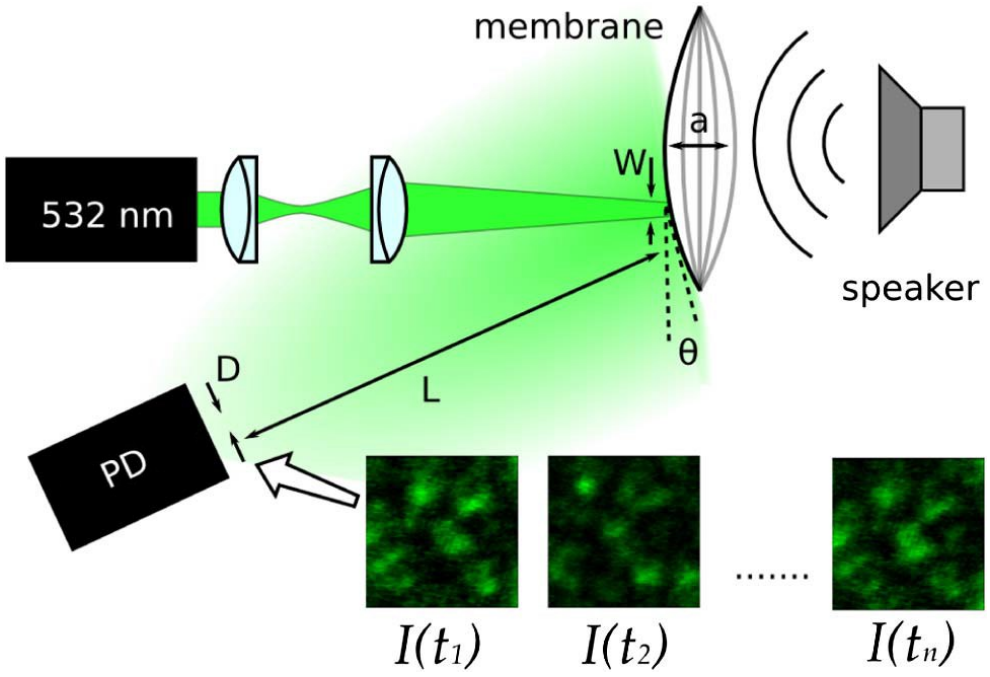

Figure 1.10 Experimental setup of detecting object vibration using photodiode and laser speckle

1.2.4 Comparison of previous work

In the previous part several typical optical vibration sensing methods have been introduced. People can see that those methods are quite different from each other, and the differences can be summarized into the following aspects. 1) The illumination. Some methods do not contain illumination modules, whereas other methods use coherent light to illuminate the object. 2) The sensor types. The sensor of the methods could be imaging sensor (e.g., digital camera) or non-imaging sensor (e.g., photodiode or single pixel detector). 3) The algorithms. Especially for the methods using imaging sensor, the processing methods of the captured speckle images could be very different. Next, the comparison of different plans will be discussed.

1.2.4.1 Passive illumination vs active illumination

One of representative passive illumination plan is the Visual Microphone proposed by MIT,

which is also an imaging-sensor-based method. Since the method does not need additional illumination, this plan has the simplest hardware structure. Nowadays, the commercial products containing imaging sensor could be found everywhere in our daily life, such as camera, smart phone, and drone. As technology advances the frame rate of those products become faster and faster. Some smart phones have already achieved the sampling rate of several thousand frames per second. Hence the passive illumination plan has a large potential for widely use. People can image that these devices can capture not only visual information but also audio information in the future.

On the other hand, active illumination method contains light source for example, a laser module, and can use it to actively emit light onto the object. Compared with passive illumination method, the additional laser module increases the complexity of the hardware. However, active illumination plan has the following advantages: 1) Although passive illumination method does not require additional light source, sufficient natural light illumination is still indispensable in order to capture the detail of the vibrating object. When the camera works at a high frame rate (i.e., the exposure time is extremely short) or in the dark environment (e.g., in the night), the camera cannot take clear image of the object and thus the method cannot recover sound. Contrarily, the active illumination method can work well at these situations because it can provide sufficient illumination using laser module. 2) With respect to measurement principle, usually active illumination method measures the micro motion of the object using coherent light interference. Theoretically the method can measure optical-wavelength-level displacement of the object. The active illumination method can provide better measurement precision than passive measurement method. 3) Generally, in an active illumination method, camera is defocused and captures the dynamic speckle patterns. While in a passive illumination method, the camera is focused on the object and captures the dynamic

images of object. From the perspective of image processing, the computational complexity of the active illumination method is lower than that of the passive illumination method. In general, active illumination method sacrifices convenience of hardware structure, but it performs better in method applicability and efficiency compared with passive illumination method.

1.2.4.2 Imaging sensor vs non-imaging sensor

Sensor is always an essential component in an optical vibration sensing method. The sensing method may not need light source, but it must need a sensor to collect signal. The sensor can be divided into two types: imaging sensor and non-imaging sensor, as shown in Figure 1.11. Imaging sensor captures images of object or speckle pattern, whereas non-imaging sensor, for example, photodiode, collect the optical signal and transfer it into electric signal.

Figure 1.11 Different sensors in laser speckle sensing method. (a) Imaging sensor (a) Photodiode

The advantage of non-imaging sensor is that it is faster than imaging sensor. Specifically, first, the signal sampling speed of a photodiode is usually faster than that of an imaging sensor. For example, the basic sampling rate of the Hamamatsu S1227-1010BR Si photodiode is 10kHz, whereas the basic sampling rate of the FLIR GS3-U3-32S4C-C camera (Sony IMX-252 imaging sensor) is only 120 FPS. Higher sampling rate means broader detectable frequency bandwidth. A normal photodiode can satisfy the needs of sampling audio signal, but one may

need a high-speed camera to capture the sound vibration. Second, processing electric signal is also faster than processing images. Apart from the speed advantage, photodiodes are usually cheaper than cameras. It also has advantages in hardware costs compared with camera.

Nevertheless, imaging sensor also has its advantages. Non-imaging sensor may be fast in sampling signal, but it cannot capture images. In other words, it is impossible to sense the visual information of the environment with a non-imaging sensor. On the other hand, imaging sensor can capture the visual information of the environment and visually locate the vibration source. The imaging sensor has advantages in information interaction with other devices compared to non-imaging sensors. With respect to the sampling speed, this problem can be solved by using cameras with high frame rate (e.g., the Phantom series high-speed camera). Some image sensors support high-frame-rate mode by reducing the image size. In this way imaging sensors can also achieve high sampling speed, for example, tens of thousands frames per second. This sampling speed can satisfy the requirement of audio signal sampling.

1.2.4.3 Motion analyzation vs intensity variation

For all those vibration sensing methods that have been introduced above, regardless of their differences in light sources and sensors, a common rule is that all these methods try to build an "object-sensor mapping" with optical means to sense the object vibration. Taking laser-speckle-and-imaging-sensor-based measurement methods as example, assuming the object vibrates with a constant frequency, the received images at the sensor must vary with same period and amplitude. This is the "object-sensor mapping". One can process the received images to restore the original signal. There are two approaches for recovering the original signal from the images. One method is motion analyzing. The principle of this method comes from the speckle sensing model, which will be introduced specifically in Chapter 2. Usually, it calculates the

displacement of the adjacent frames, then integrates the results to obtain the original signal. Recently, some researchers proposed another way of image processing. When the received speckle image vibrates, the gray scale value of each pixel also varies. They found that the intensity variation of some pixels matches with the original signal, and this could be used to present the vibration status of the object. This is the so-called intensity variation method.

It is not difficult to see that the intensity variation method is faster than the motion analyzation method with respect to image processing speed. The main processing of intensity variation method is accessing several pixels' gray value and merge them to obtain a sampling data, which does not require too much computational cost. On the other hand, analyzing image displacement is very time-consuming, especially when the image size is large, the displacement registration requires much computational cost. However, motion analyzation is more reliable compared with intensity variation method. The purpose of image processing is recovering the original signal of the vibration object from the captured images. From this point of view, one can always recover the information of the object vibration by analyzing motion of the captured images because it is based on the objective physics laws: In a laser-speckle-based measurement method, the vibration of the captured speckle images is a response to the object vibration. On the contrary, the intensity variation method is more like a "coincidence". It finds some pixels where the gray value variation matches with the object vibration among all the pixels of the image and uses it to show the original signal. From [32] people can see that there are several limitations for the selection of the ideal pixels in a speckle image: 1) Along the displacement direction, the ideal pixel should situate at the middle position of neighbor pixels with minimum gray value and maximum gray value. 2) The distance between neighbor pixels with minimum gray value and maximum gray value is twice larger than the maximum displacement during the movements. 3) The gray value variation between the neighbor pixels with minimum gray

value and maximum gray value is as linear as possible. 4) The gray difference of neighbor pixels with minimum gray value and maximum gray value is as big as possible in order to ensure a large gray variation while translation takes place.

From the above description one can see that the intensity variation method can restore the original vibration signal only when the pixels that meet those conditions exist. Moreover, when the external condition changes, for example, the received light intensity at sensor or the vibration amplitude of the object changes, the gray value variation of the selected pixels may change too and the intensity variation at those points may no longer match with the original signal. In general, usually the intensity variation method is faster than motion analyzation method, but motion analyzation method is more reliable and robust than intensity variation method.

1.3 Future Image of the Proposal Method

The major advantage of our proposed method is the real-time restoration of audio signal (or vibration) information through ultra-high-speed and high precision image processing of laser speckle images. Based on this characteristic, here several the application scenarios of the proposed method are stated.

One of the imagine usages is a laser microphone device in a large conference scene, as shown in Figure 1.12. In this situation, the area of the venue is usually very large, and also there are a large number of people in the environment. If electric microphone is used as a solution, the problem is that people need to prepare a large number of microphones; or pass the microphone to the speaker when someone speaks. The electric microphone is not an ideal solution in a large conference situation. Compared with traditional electric microphones, laser-visual microphones have the following advantages. 1) Using expanded laser beam illumination, the

camera can quickly find the sound source if someone speaks in the illumination area, then extract audio signal from the people who is speaking. 2) In the case of multiple targets talking at the same time, using the visual information from the camera can easily distinguish different sound sources, avoiding crosstalk of multiple audio signal. Since the laser and camera can covering a large number of people, only one set of equipment can complete the acquisition of audio signals within the field of view, which greatly reduces the number of required hardware devices.

Figure 1.12 Imagine usage of the proposed technique: laser microphone in large conference scene

In Figure 1.13 we show another hypothesis of the potential application of the proposed technology, which is extracting conversation from player in sports game. Currently, when broadcasting sports programs, the problem of how to obtain the content of the contestants' conversation still has not been effectively solved. Due to the football or basketball's stadiums are very large, and there is noise in the environment, it is difficult for traditional sound recording equipment to effectively extract the conversations of players in the stadium.

Although there are directional microphones on the market that can extract sound with long distance, it is still difficult for such a device to distinguish the sound signals of two near targets that are tens of meters away. However, it is possible to use our proposed method to solve this problem. We can use laser (invisible) to irradiate the skin of the target. When the target speaks, the laser speckle reflected from the throat of the target is captured by the camera, and the vibration of the laser speckle image can be used to restore the content of the target's speech in real time. Using this method, the content of conversation can be extracted from a long distance, and the active detection method will not be affected by the noise in the environment. For obtaining the content of the athlete's conversation in sports competitions, a laser-speckle-based visual microphone is an ideal solution

Figure 1.13 Imagine usage of the proposed technique: extract talking from player in sports game

1.3 Thesis Outline

The chapter contents are introduced bellow. In this thesis Chapters 2 and Chapters 3 present

the theory for our method, and Chapter 4 and Chapter 5 report the original design and the related experiment work.

Chapter 1. Introduction

This part discusses the research background of this thesis and the past research on the optical vibration measurement method. Following that, the characteristics of different studies are described, and a short overview of the advantages of our proposed sensing method is given.

Chapter 2. Basic optical principles of speckle sensing method

This chapter introduces the optical principle of the speckle sensing method. In this chapter, we begin with a discussion of the cause of the speckle and then move on to discuss the various properties of the speckle image. Afterwards, we present the image characteristics and the motion model of the speckle image in the defocus camera. Lastly, experiments have been performed to determine the validity of the theories.

Chapter 3. Image processing in speckle sensing method

The intention of this chapter is to introduce the back-end processing part of the speckle sensing method, namely, how to analyze the displacement of the speckle images. This chapter first introduces the determination method of speckle displacement in traditional speckle photography, and then introduces the analysis algorithm of image displacement for the digital image, including DIC algorithm, feature-based algorithm, and optical flow algorithm. The thesis then analyzed and compared the characteristics of different algorithms. The optical flow algorithm is found to meet the back-end processing needs of the speckle sensing algorithm, which is high-speed and high-precision processing of small displacement images.

Chapter 4. Real-time acoustic vibration sensing method using 2-D speckle image

This chapter presents the original work of the research. In this chapter a laser-speckle-based sound detection method has been proposed. The chapter first introduced the image processing method and the signal processing method of the proposed method. Then the experiment results are given to prove that our proposed method is an efficient way to regenerate audio signal under different situations. The main advantages of the proposed method are real-time audio signal regeneration with high quality and the ability of audio signal regeneration of moving sound source, which have never been realized before.

The work of this chapter is published as one conference paper (see “list of publications, 3”) and one journal paper (see “list of publications, 1”).

Chapter 5. Fast acoustic vibration sensing method using 1-D speckle image

This chapter presents another original work of the research. In this chapter, we proposed a fast motion estimation of a 1D laser speckle image and showed its application on a real time audio signal detection method. The major contribution of our work is the fast displacement estimation algorithm for 1D speckle images. Owing to this, the proposed method can achieve a 20k Hz real time sampling speed. Experimental results presented the capabilities of the proposed method for extraction of high frequency audio signals and human voice. The high speed, real time sampling method discussed in this chapter has broad application prospects, such as voice signal acquisition and high frequency vibration monitoring of industrial equipment.

The work of this chapter is published as one conference paper (see “list of publications, 4”) and one journal paper (see “list of publications, 2”).

Chapter 6. Conclusion

Summarizes the thesis work and discuss their overall impact, considers limitations, and outlines ideas for future work.

CHAPTER 2 BASIC OPTICAL PRINCIPLES OF SPECKLE SENSING METHOD

Speckle phenomenon is common in the process of optical imaging. Long time ago, Newton explained the phenomenon that stars flicker, but planets do not flicker [35]. Speckle has attracted more attention because of the invention and use of laser. After laser was widely used, researchers soon noticed a strange phenomenon: when the laser was reflected by surface, like paper or wall, a high-contrast grainy pattern could be observed by human eyes. This high-contrast noise-like pattern was after called “speckle”. By measuring the intensity of the speckle pattern, it was found that the fluctuations of the intensity distribution still exist in the space even the illumination of the laser spot is relatively uniform. People then realized that the fluctuations of the intensity in the space is generated due to the roughness of the surface. Up to now, plenty of studies on the laser speckle have been conducted. The earliest report of laser speckle can be found in Langmuir and Oliver’s publishes in 1963 [36, 37]. In the publishes, they reported the speckle phenomenon, and they hypothesized that the speckle rises due to the diffuse reflection of the coherent light. After that, Goodman conducted a systematic study of the speckle phenomenon, including the statistical properties of speckle pattern, the suppressing of speckle pattern, speckle phenomenon in various optical devices. It provides a valuable reference to the later research on speckle phenomenon [38-40].

Since the speckle appears as noise in the photography, initially people mainly studied how to reduce the influence of speckle. However, with the development of speckle research, it was

found that the speckle carries a lot of information about the beam and the object that the beam passes through, which makes speckle have many applications [41]. For example, the contrast of speckle is used to measure the roughness of the reflective surface [42, 43], the dynamic variation of the speckle is used to measure the speed of object movement [44-46], the speckle is used to measure acceleration of the object [47], and so on. In the past laser speckle can be captured by exposure on the film, but the nowadays people mainly use digital imaging sensor to capture the speckle pattern. Since this technology avoids developing and fixing, it can achieve the purpose of real-time measurement, so it is increasingly widely used in scientific research and production applications.

2.1 General Introduction

2.1.1 Characteristics of light interference

Speckle pattern is a phenomenon of laser interference. Here a short review of the basic knowledge of light waves and laser interference will be made before discussing speckle.

In the early 1960s, continuous-wave (CW) laser became commercially available. Different from other light sources, laser is a special kind of light produced by using external abilities such as light energy, heat energy, electrical energy, chemical energy, or nuclear energy to excite matter, causing it to generate stimulated radiation [48]. Laser has the characteristics of high intensity, directionality, monochromatic and coherence.

As is known to everyone, light is a kind of electronical wave. Typically, light can be represented as a sinusoidal function of time and space, and it exhibits periodicity in both time and space. The wave function of one-dimensional simple harmonic can be expressed as:

$$A(z - vt) = A_0 \cos\left[\frac{2\pi}{\lambda}(z - vt) + \varphi_0\right] \qquad (2-1)$$

In Equation (2-1), A_0 presents the amplitude of a light wave, and $[2\pi/\lambda(z - vt) + \varphi_0]$ presents the phase of a light wave, which is written as $\varphi(z, t)$ for the convenience. It could be seen that the phase of a light wave changes with both spatial variable z and time variable t, thus light wave $A(z - vt)$ is a period function of z and t. Figure 2.1 presents periodicity of a light wave. Figure 2.1(a) shows the signal periodicity in space, and Figure 2.1(b) shows the periodicity in time. Here λ is the wavelength of a light wave, which means how far light can transfer in space within a period. $T = \lambda/v$ is the time period of light wave, which means the vibration period, or the time required for the vibration cycle at any point in the space. Actually, the propagation of light waves in space is the propagation of phases.

Now we know that the amplitude A of light at a spatial point z and time point t depends on the phase φ of the light wave, and A varies with φ periodically. Therefore, when two or more propagating light waves are incident on the same point in the space, the resultant amplitude at the point is not simply determined by the sum of the individual intensities of the light waves but depends on the phase difference between the individual waves. This is the phenomenon of light wave interference. Interference can only occur between light waves that meet the following conditions. First, the light waves must be coherent. In other word, they must maintain a constant phase difference between each other. Second, the wavelength of each light waves must be same. From these conditions one can see that laser is an ideal light source for interference. Figure 2.2 shows the example of two light waves incident on the same point. The resultant amplitude could be large (called constructive interference) or small (called destructive interference), depending on the phase difference of two light waves.

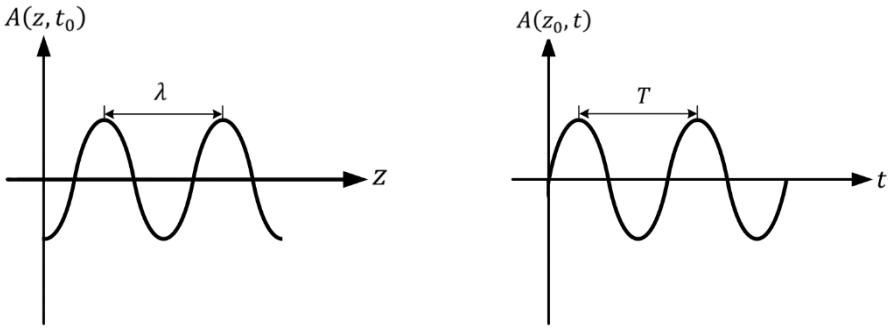

Figure 2.1 The propagation of light waves in space and time

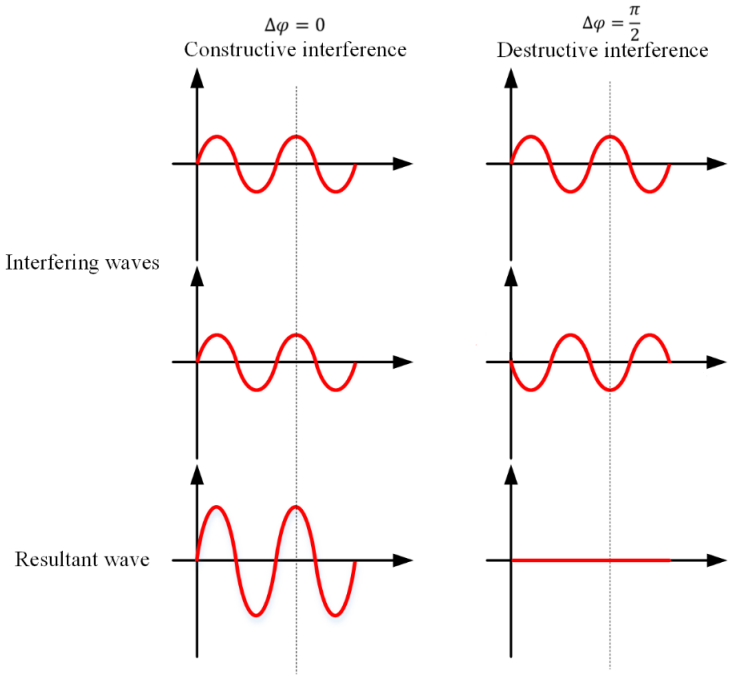

Figure 2.2 When light waves with different phase differences are superimposed, the superimposed sum is constructive or destructive

The resultant of multiple light waves can be analyzed in same way, but it is very complex if we do this work with Equation (2-1). For the convenience of expression, people introduced another way of expressing light waves: the complex amplitude form of light waves. In Equation (2-1), let $\varphi = 2\pi/\lambda(z - vt) + \varphi_0$, we got:

$$A = A_0 \cos\varphi \qquad (2-2)$$

Here we use a vector $\vec{P}(x, y)$ in complex domain to represent the state of the light wave. As shown in Figure 2.3, the amplitude A_0 of a light wave is represented by the length of the vector, and the phase φ of the light wave is represented by the angle of the vector relative to the real axis. In this way, the expression of a light wave can be written as:

$$A = x + iy = A_0 \cos\varphi + iA_0 \sin\varphi \qquad (2-3)$$

or simply as:

$$A = A_0 e^{i\varphi} \qquad (2-4)$$

As the phase changes with time or distance, the direction of the phasor will rotate in the plane of the real and imaginary axes.

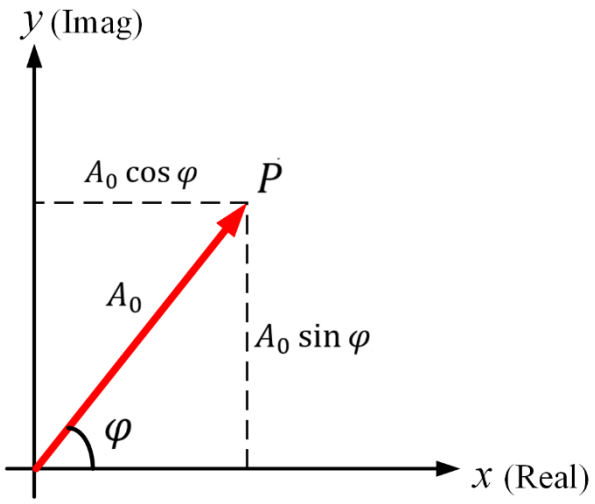

Figure 2.3 Representation of light waves in the complex plane

With help of the complex expression, the superposition of light waves becomes the vector addition on the complex plane. When we analyze the resultant of multiple lights with different

amplitudes and phases, using complex expression can make the calculation simpler and more intuitive. For example, it is not easy to obtain the resultant of three waves with different amplitudes and phases using trigonometric functions, as shown in Figure 2.4(a), but it is clear to obtain the answer with complex expression, as shown in Figure 2.4(b).

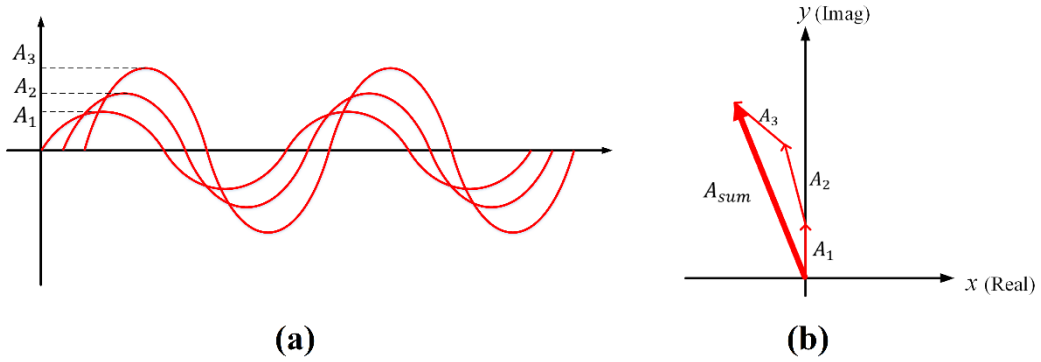

Figure 2.4 Different ways of expressing the results when multiple lights are superimposed. (a) Wave function representation (b) Complex amplitude representation

2.1.2 Intuitive explanation of cause of laser speckle

The generally accepted explanation for the cause of speckle is that the speckle rises from the interference of laser light rays. Multiple light rays could be generated when light is reflected by optical rough surface, or light transfers through a diffusive medium. In this paper we only discuss the situation that light is reflected from surface. Figure 2.5 illustrates the formation of the reflected speckle pattern.

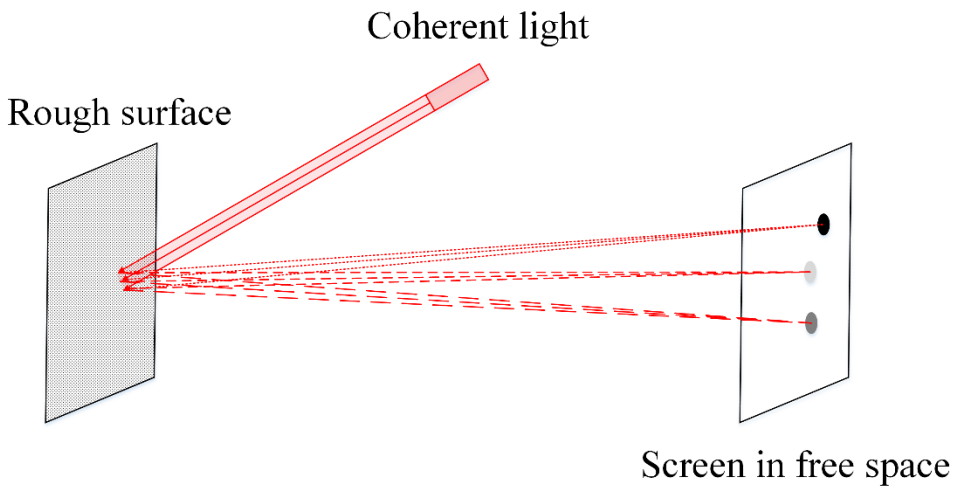

Figure 2.5 Intuitive explanation of the cause of speckle in space

When an object with rough surface is illuminated by a laser beam, laser light is scattered to all directions randomly. The illumination area could be regarded as a group of multiple virtual light sources, each light source emits light rays into space with random phase and direction. If a sensor is settled in the space, the received intensity of each pixel is determined by light rays come from all virtual light sources, and each light ray is independent phase and amplitude. The superposition of light rays with independent phases and amplitudes is called “random work”, as shown in Figure 2.6.

One can see from Figure 2.6 that the result of random work could be large or small, depending on the relative phases of the sum. When the result is large, i.e., constructive interference occurs, bright spots could be observed on the sensor. When the result is small, i.e., destructive interference occurs, dark spot is observed on the sensor. Figure 2.7 shows an example of the speckle pattern captured by a bare CMOS sensor.

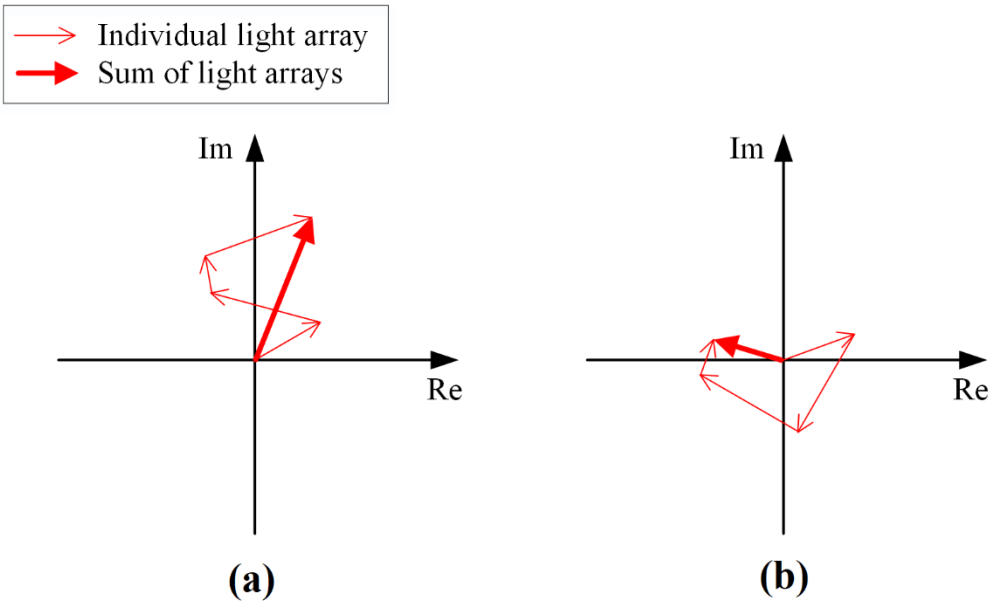

Figure 2.6 Random walk and resultant phasor with (a) largely constructive addition, and (b) largely destructive addition

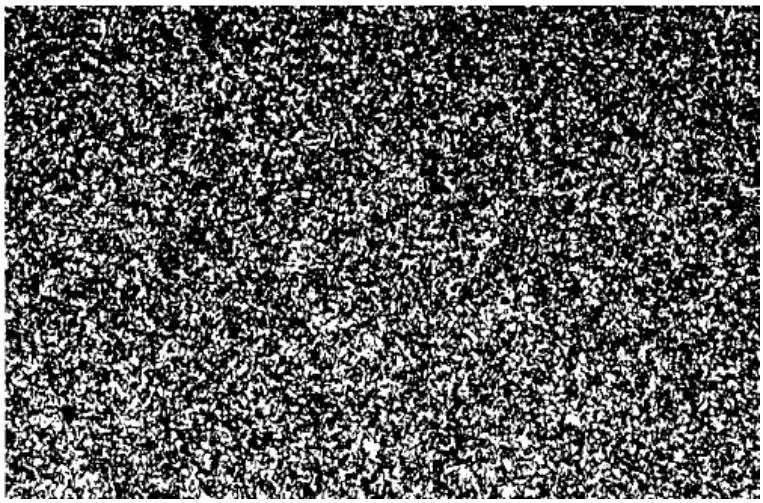

Figure 2.7 Example of the speckle pattern captured by a bare CMOS sensor

2.2 Statistic Characteristic of Speckle Pattern

It was mentioned in the former part that speckle generates from sum of a number of lights with random amplitudes and phases. The random sum results in a grainy speckle pattern in the

space. In practical situations, people are more concerned about the intensity information of the speckle pattern rather than the amplitude because the intensity information can be directly measured by experimental devices such as photodiode or imaging sensor. The intensity of a point P in the space is determined by the result of random walk. Since it is impossible to obtain the phase and amplitude information of individual component, one cannot determine the speckle intensity at a specific point. However, people can use statistical methods to calculate the probability distribution of a large number of components of random walk results, summarizing the statistical laws of speckle patterns and establishing a statistical model of speckles.

2.2.1 First-order statistics of laser speckles

The first-order statistics of laser speckles means the statistics properties of intensity and phase of a speckle pattern, and "first-order" incident these properties at s single point in space and time. Figure 2.8 shows N light rays arrive at one point in space, which is also an intuitive representation of Figure 2.4. Assuming the amplitudes of individual component are $A_1, A_2, \dots A_n$, and phases of individual components are $\varphi_1, \varphi_2, \dots \varphi_n$. Therefore, the resultant phasor $\mathbf{A}$ can be expressed as:

$$\mathbf{A} = Ae^{i\varphi} = \sum_{n=1}^{N} A_n e^{i\varphi_n} \qquad (2-5)$$

where A is the amplitude of resultant phasor, and φ is the phase of resultant phasor.

Since the laser is randomly scattered by a rough surface, the following assumptions can be made: 1) For all light rays, the amplitudes and phases are statistically independent. In other word, the amplitude A_n conveys no knowledge about another amplitude A_m when $n \neq m$,

and same for phases. 2) For individual component, the amplitude and phase are statistically independent. That is, the amplitude A_n conveys no knowledge about phase φ_n, and vice versa. 3)The phase φ_n of all light rays are uniformly distributed on 2π interval.

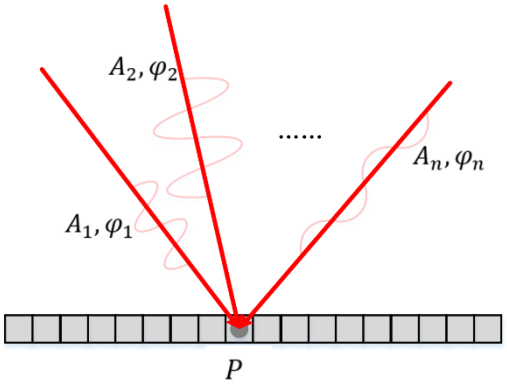

Figure 2.8 Multiple light rays arrive at one point in space

The above are the assumptions for obtaining the first-order statistical properties of laser speckles. It should be noted that the first-order statistical properties of speckle may be different when some of these assumptions are not valid. This may happen in some special physical conditions, for example, when light is reflected on a smooth surface (mirror), the amplitude and phase may no longer be statistically independent. However, in most cases the laser light is scattered on the rough surface, which means that the above assumptions are usually valid.

Back to Equation (2-5), for the resultant phasor $\mathbf{A}$, the real part and the imaginary part are:

$$\mathbf{R} = \mathrm{Re}\{\mathbf{A}\} = \sum_{n=1}^{N} A_n \cos\varphi_n \qquad (2-6)$$

$$\mathbf{I} = \mathrm{Im}\{\mathbf{A}\} = \sum_{n=1}^{N} A_n \sin\varphi_n \qquad (2-7)$$

Since φ_n of each component are uniformly distributed in $[-\pi, \pi]$, the first moment (expected value) and the second moment (variance) of $\mathbf{R}$ and $\mathbf{I}$ are simple. The expected values are:

$$E(\mathbf{R}) = E\sum_{n=1}^{N} A_n \cos\varphi_n = \sum_{n=1}^{N} E(A_n \cos\varphi_n) = \sum_{n=1}^{N} E(A_n)E(\cos\varphi_n) = 0 \qquad (2-8)$$

$$E(\mathbf{I}) = E\sum_{n=1}^{N} A_n \sin\varphi_n = \sum_{n=1}^{N} E(A_n \sin\varphi_n) = \sum_{n=1}^{N} E(A_n)E(\sin\varphi_n) = 0 \qquad (2-9)$$

The variances are:

$$\sigma_{\mathbf{R}}^2 = E(\mathbf{R^2}) = \sum_{n=1}^{N} E(A_n^2)E(\cos^2\varphi_n) = \sum_{n=1}^{N} \frac{E(A_n^2)}{2} \qquad (2-10)$$

$$\sigma_{\mathbf{I}}^2 = E(\mathbf{I^2}) = \sum_{n=1}^{N} E(A_n^2)E(\sin^2\varphi_n) = \sum_{n=1}^{N} \frac{E(A_n^2)}{2} \qquad (2-11)$$

We can see that the expected values of real part and imaginary part of the resultant phasor are identical, and same for variances. When the number of light rays is very large, i.e., $N \to \infty$, the statistics of such N independent random components is Gaussian distribution according to the Central Limit Theorem. Therefore, the joint probability density function for the real and imaginary part is:

$$p_{\mathbf{R},\mathbf{I}}(\mathbf{R},\mathbf{I}) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\mathbf{R}^2+\mathbf{I}^2}{2\sigma^2}\right) \qquad (2-12)$$

where $\sigma = \sigma_{\mathbf{R}} = \sigma_{\mathbf{I}}$.

Our interest is the intensity of the speckle pattern. The intensity I can be deduced as:

$$I = \mathbf{A}\mathbf{A}^* = (A\cos\varphi + i\,A\sin\varphi)(A\cos\varphi - i\,A\sin\varphi)$$
$$= A^2\cos^2\varphi + A^2\sin^2\varphi$$

$$= \mathbf{R^2} + \mathbf{I^2} \qquad (2-13)$$

The phase θ is:

$$\theta = \arctan\left(\frac{\mathbf{I}}{\mathbf{R}}\right) \qquad (2-14)$$

Equivalently, the real part and the imaginary part of resultant phasor can be expressed using intensity and phase, as:

$$\mathbf{R} = \sqrt{I}\cos\theta \qquad (2-15)$$

$$\mathbf{I} = \sqrt{I}\sin\theta \qquad (2-16)$$

The probability density functions of the intensity I and the phase θ are as follows:

$$p_I(I) = \frac{1}{2\sigma^2}\exp\left(-\frac{I}{2\sigma^2}\right) \qquad (2-17)$$

$$p_\varphi(\varphi) = \frac{1}{2\pi} \qquad (2-18)$$

Equation (2-17) and Equation (2-18) show that when a sufficient large number of independent light rays hit on one point in the space, the intensity of the resultant phasor follows negative exponential distribution, whereas the phase of the resultant phasor follows uniform distribution. Figure 2.9 shows the negative exponential distribution. The negative exponential distribution of the intensity means for a random point in the space, it is high probability that the intensity at this point is weak. In other word, destructive interference is more likely happened in a speckle field.

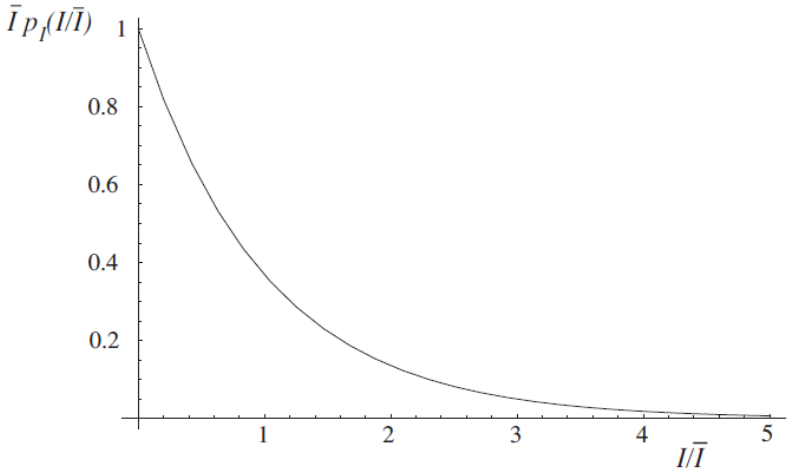

Figure 2.9 Negative exponential distribution

Besides, from Equation (2-17) we can obtain the moments of intensity distribution by direct integration:

$$\overline{I^q} = (2\sigma^2)^q q! \qquad (2-19)$$

From Equation (2-19) the mean intensity $\bar{I}$ can be easily found, which is $2\sigma^2$. Thereby, Equation (2-19) can be also expressed as:

$$\overline{I^q} = \bar{I}^q q! \qquad (2-20)$$

and the probability density function can be also expressed as

$$p_I(I) = \frac{1}{\bar{I}} \exp\left(-\frac{I}{\bar{I}}\right) \qquad (2-21)$$

Here some important statistics properties of intensity, like second moment, variance, and standard deviation can be obtained

$$\overline{I^2} = 2\bar{I}^2 \qquad (2-22)$$

$$\sigma_I^2 = \bar{I}^2 \qquad (2-23)$$

$$\sigma_I = \bar{I} \qquad (2-24)$$

Knowing these statistics properties is important because they can help us define some advanced properties of speckle pattern, for example, the speckle contrast C and the signal-to-noise ratio S/N. They are defined by

$$C = \frac{\sigma_I}{\bar{I}} \qquad (2-25)$$

$$\frac{S}{N} = \frac{1}{C} = \frac{\bar{I}}{\sigma_I} \qquad (2-26)$$

For a fully developed speckle pattern, the standard deviation $\sigma_I = \bar{I}$, therefore its contrast $C = 1$. In practical the contrast of speckle image can be calculated from its standard deviation and average image intensity.

2.2.2 High-order statistics of laser speckles

Discussing the first-order statistics helps us understand the properties of speckle at a single point in the space and time. The physical meaning is the statistical properties of intensity and phase of a speckle pattern. However, the first-order statistics does not present information about the spatial properties of speckle pattern, such as the average speckle size. In this part the high-order statistics of laser speckle will be discussed. Compared to first-order statistics, the high-order statistics reveals not a single point but the joint properties of speckles at two points. The two values could be two points in space, two points in time, or one point in two speckle patterns. Discussing the high-order statistics of laser speckle could help us find out properties like average speckle size of a speckle pattern. Those properties are significant to speckle

measurement. A speckle image contains a large number of speckles, and each speckle has different shape and size. Therefore, people usually use statistical average to characterize the speckle size.

Generally, the average speckle size can be obtained using autocorrelation of captured speckle pattern. The width of the autocorrelation result can be used to determine the speckle scale. The second-order statistical of laser speckle is its autocorrelation of speckle intensity of two points in space or time, which is:

$$R_I(x_1, y_1, x_2, y_2) = \langle I(x_1, y_1) I(x_2, y_2) \rangle \qquad (2-27)$$

This autocorrelation function presents the speed of speckle field variation in space. When $I(x_1, y_1) = I(x_2, y_2)$, which means the two points (x_1, y_1) and (x_2, y_2) are identical, the autocorrelation function becomes:

$$R_I(x_1, y_1, x_2, y_2) = \langle I^2(x_1, y_1) \rangle \qquad (2-28)$$

In this case the function value is maximum. The width of function R_I indicates at which separation the two points begin to be more or less independent, in other word, an order of average speckle size.

The speckle pattern can be observed in an observation plane in the space. Assuming the object-plane distance is z, the result can be expressed as:

$$R_I(\Delta x, \Delta y) = \langle I \rangle^2 [1 + |\mu_A(\Delta x, \Delta y)|^2] \qquad (2-29)$$

where μ_A is the normalized mutual intensity, which is given by:

$$\mu_A(\Delta x, \Delta y) = \frac{\iint I(u, v) \exp\left[\frac{2\pi i}{\lambda z}(\mu \Delta x + \mu \Delta y)\right] du dv}{\iint I(u, v) du dv} \qquad (2-30)$$

Usually, the shape of the laser spot on the object surface is circle. Assuming the diameter of the illuminated area is d, the complex coherence factor then can be assumed to be:

$$\mu_A(\Delta x, \Delta y) = \frac{2J_1(\alpha)}{\alpha} \qquad (2-31)$$

where

$$\alpha = \frac{\pi D}{\lambda z}\sqrt{\Delta x^2 + \Delta y^2} \qquad (2-32)$$

$J_1(\alpha)$ is the Bessel function of the first kind, order one. In addition, μ_A has a first zero for the following:

$$\sqrt{\Delta x^2 + \Delta y^2} = 1.22\frac{\lambda z}{D} \qquad (2-33)$$

Two points in the (x, y) plane which are separated by this distance have the same value of R_I as two statistically independent points which are far apart from each other. Therefore, Equation (2-33) gives a reasonable order of the speckle width. When laser beam is reflected by an optical rough surface, the average size of the speckles is determined by the beam diameter d, laser wavelength λ, and the distance z between reflection surface and the observe screen. The average speckle size s can be expressed as

$$s = 1.22\frac{\lambda z}{d} \qquad (2-34)$$

In actual application, if the speckle is captured by a camera rather than a bare screen, the average speckle size will be scaled. This will be discussed in the next part. For the captured speckle images, people actually use autocorrelation function to calculate the size of the speckles. Figure 2.10 shows a speckle pattern and Figure 2.11 shows its autocorrelation result. Usually, the average speckle size is determined by the width of the with half of the maximum

autocorrelation value.

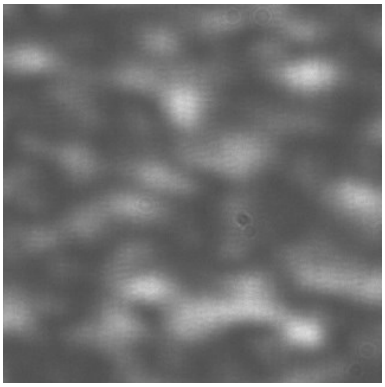

Figure 2.10 Captured speckle template with width and height of 320 pixels

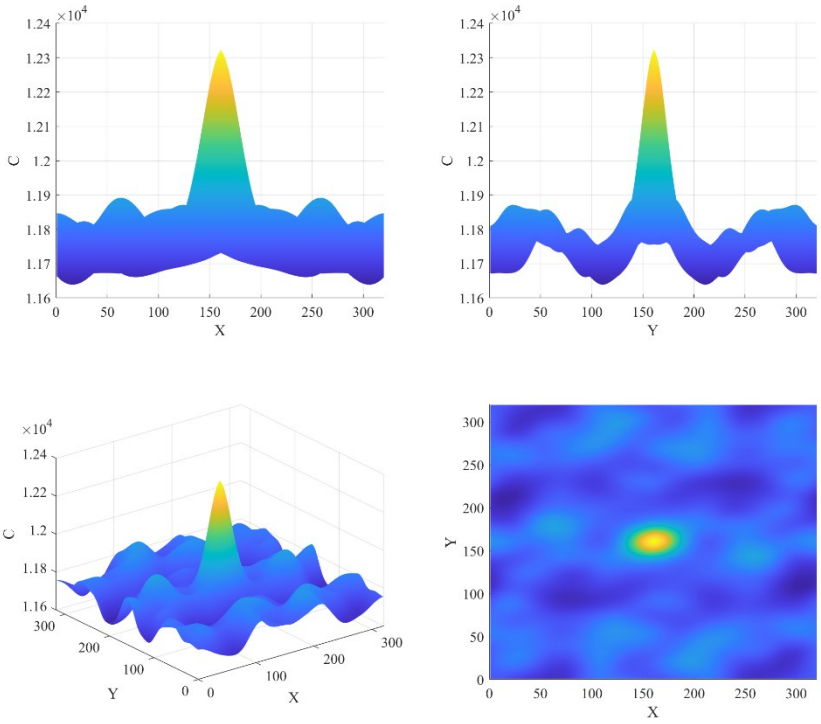

Figure 2.11 Autocorrelation result of captured speckle pattern

2.3 Laser Speckle in Camera

In the former part, we introduced that speckle is generated by the superposition of a large

number of light rays with random amplitude and phase in free space, and also the statistical properties of the resultant phasor are introduced. The speckle phenomenon can be found in many optical devices, such as holography, optical coherence tomography, optical projection display, or even human eyes. In free space, a laser speckle image scattered from an optical rough surface be recorded using a bare sensor. However, in a speckle-based measurement method, the laser speckle image is usually recorded by camera rather than a bare sensor. A typical camera is consistent with sensor and lens. In the early years people use film and lens to record dynamic speckle images through multiple exposure, but now the work is done using digital camera.

Since the light is refracted when it passes through lens, the speckle pattern recorded by camera is different from that recorded by a bare sensor in free space. Especially in speckle-based measurement method, people usually defocus the camera to record the speckle pattern, which further changes the intensity and average size of the speckle pattern. In this part speckle properties in camera will be introduced. We will begin with a review of geometric optics knowledges of lens imaging model. Then speckle pattern in camera will be introduced. Finally, the effect of defocus on speckle properties is introduced.

2.3.1 Basic imaging principle of thin convex lens

Generally, a digital camera contains imaging sensor and lens, the lens focuses light from every point of the object onto the sensor plane so the camera can capture a clear image of object. The lens is complex and usually it is consisted with multiple lens, as shown in Figure 2.12. These lenses are combined together to play a role similar to a convex lens with fixed or adjustable focal length. Therefore, the lens can be regarded as a single thin convex lens for convenience. The thin lens means that the thickness of the lens is ignored when discussing the

principle of camera.

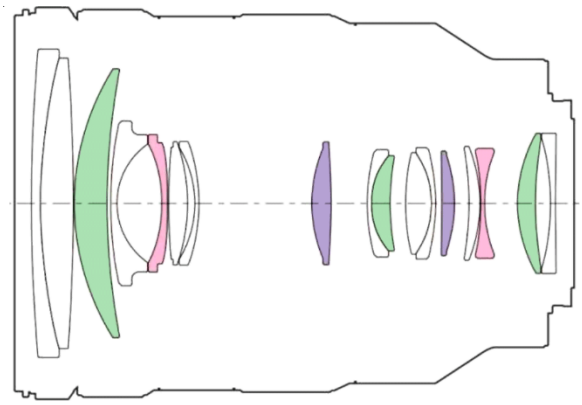

Figure 2.12 Camera lens group composed of multiple lenses

Figure 2.13 illustrates a focused camera based on thin convex lens. The first and foremost parameter of a convex lens is its focal lens f. On the optical axis in front and after lens, the two virtual points at the distance f from the center of the lens are called the front focal point (FFP) and back focal point (BFP). In a focused camera, the lens focuses lights emitted from every point of object on the focal plane onto the image sensor to obtain a clear image of the object. This gives us another important parameter: the focusing length of the lens u, which is the distance between focal plane and the lens. For a focused camera, u is also the object distance. Correspondingly, the distance v between imaging sensor (or image plane) is called image distance.

According to geometric optics, three rules must be followed when light transmit through lens. First, the direction of light passing through the center of lens is unchanged. Second, light parallel to the optical axis is refracted when passing through the lens, and the refracted light must pass through the BFP. Third, light passing through the FFP is refracted by the lens and then travels parallel to the optical axis. The three fundamental rules can help us determine the image's position and size object on the focal plane by tracing three special light rays emitted

from every point of object.

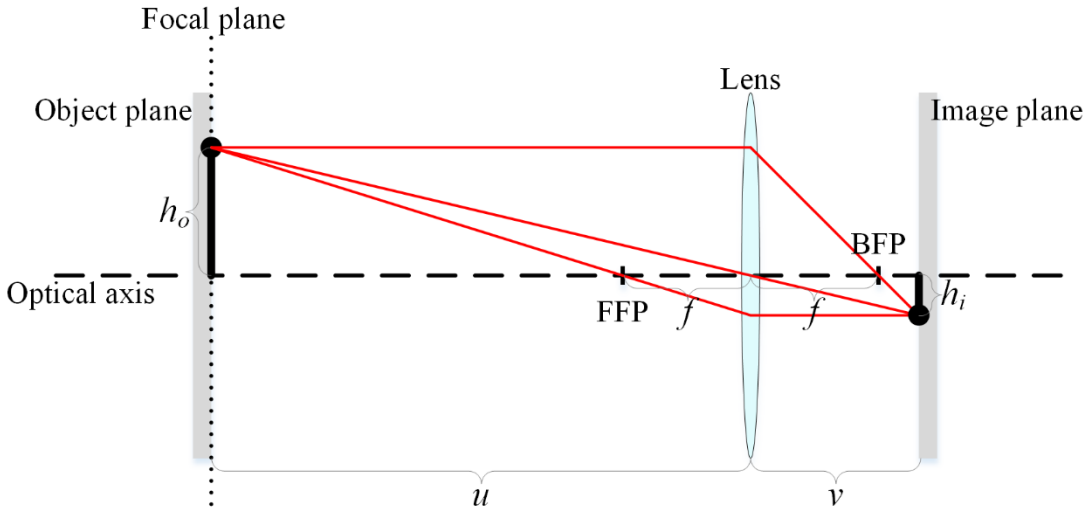

Figure 2.13 Typical principle of focused camera with thin convex lens

It is obviously that if the object distance u is changed, the value of v need to be changed accordingly in order to capture a clear image, and vice versa. The relationship between u, v, and f can be expressed as [49]:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \qquad (2-35)$$

Generally, the focus distance of the camera lens is an adjustable value, and the scale is marked on the lens, as shown in Figure 2.13. Thus, the image distance v can be easily obtained using Equation (2-35). In this way, we can further obtain the relationship between the object size and image size. Assuming that the object size is h_o and the image size is h_i, the ratio of the object-image can be expressed as:

$$\frac{h_o}{h_i}=\frac{u}{v} \qquad (2-36)$$

For the convenience of further discussion, here gives definition of the magnification M of the

camera, which is:

$$M = \frac{h_i}{h_o} = \frac{v}{u} \qquad (2-37)$$

Since the distance between lens plane and imaging sensor plane is hard to figured out when capturing image, M is commonly expressed using focal length and focusing length of a camera:

$$M = \frac{f}{u-f} \qquad (2-38)$$

2.3.2 Description of defocused camera

In most cases, people make camera focus on the object to obtain a clear image when they use camera to capture object. More intuitively, focus means that the object distance u is same as the focusing distance of lens, or in other word, the object is located at the focal plane. However, in speckle-based measurement, people are more concerned about the information of speckle image reflected from the object rather than the object itself. Although our target is measuring the vibration of an object, its information can be reflected by speckle vibration. Even the camera does not collect the details of the object's motion, one can still obtain the information needed. Therefore, in speckle sensing method the camera is usually defocused. The simple understanding of defocus is that the object distance u is not equal to the focusing distance of camera lens, or in other word, the object is not located at the focal plane. Figure 2.14 illustrate the defocused camera, where the object is offset from the focal plane. It should be claimed here that the defocus situation discussed in this paper refers to the camera's focal plane locates between the object and the camera. The case where the focal plane is on the other side of the object will not be discussed.

A significant parameter in a defocused camera is the amount of defocus ΔL, which is defined as the distance between the object and the focal plane. In a defocused camera, light emitted from a point of the object does not perfectly converge at the sensor plane but are instead spread over a finite area, the captured image appears blurred. The diameter D_i of the blurred area on the sensor can be deduced using the theory from part 2.3.1.

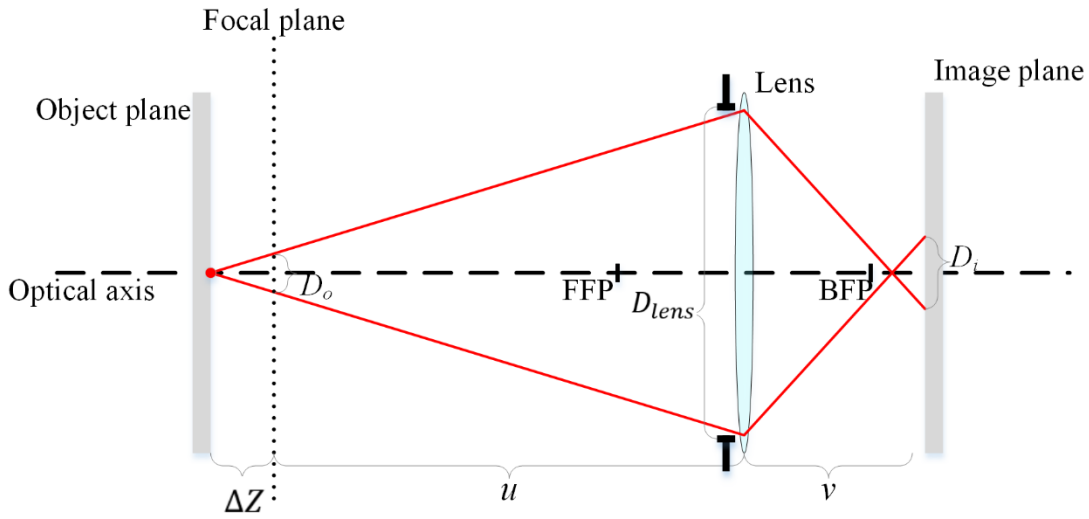

Figure 2.14 Description of a defocused camera

Assuming the light rays emitted from the point on the object are spread into a cone with diameter D_o at the focal plane. According to the proportional relationship of similar triangles, we have:

$$\frac{\Delta Z}{D_o} = \frac{\Delta Z + u}{D_{lens}} \qquad (2-39)$$

where D is the aperture of lens.

Assuming there is a virtual object with diameter D_o at the focal plane, in other word, the blurred area on the sensor is the image of the virtual object at focal plane. With the magnification factor M of camera, we can calculate the diameter D_i of the blurred area on

the sensor, which is:

$$D_i = D_o M = \frac{D_{lens}\Delta Z}{\Delta Z + u} M \qquad (2-40)$$

2.3.3 Speckle in defocused camera

In part 2.2.2 we give the expression for the average speckle size in free space, which is Equation (2-34). In camera, the speckle in the space is focused onto the imaging sensor by camera lens. The image sensor is composed of a two-dimensional pixel array. Therefore, people usually use pixels as a unit to measure the speckle size and the speckle displacement. In this part we will deduce the average speckle size of speckle pattern captured by a defocused camera. We also did experiment to reveal that how defocus affects the intensity and contrast of a speckle pattern.

It has been mentioned before that in a defocused camera, it could be considered that the camera captures virtual object on the focal plane. The laser lights scattered from the object spread into space, and the average speckle size on the focal plane s_o is $1.2(\lambda\Delta Z/d)$ according to Equation (2-34). The virtual speckle is then captured. Thus, the size of the virtual speckle is magnified by factor M. Therefore, the average speckle size on the sensor is:

$$s_s = 1.22\frac{\lambda\Delta Z}{d} M = 1.22\frac{\lambda f}{d} \times \frac{\Delta Z}{u} \qquad (2-41)$$

In the camera, the $F^{\#}$ also determines the observed speckle number at imaging sensor. Assume the number of observed speckles is N and the diameter of lens aperture is D, so we have:

$$N = \frac{D}{s_s} = \frac{Dd}{1.22\lambda f} \times \frac{u}{\Delta Z} \qquad (2-42)$$

The $F^{\#}$ is defined as the ratio between lens focal length and the lens aperture, which is given by:

$$F^{\#} = \frac{f}{D} \qquad (2-43)$$

Based on Equation (2-42) and (2-43), we have the number of speckles at the imaging sensor, which is given by:

$$N = \frac{d}{1.22F^{\#}\lambda} \times \frac{u}{\Delta Z} \qquad (2-44)$$

From Equation (2-44) we can see that the $F^{\#}$ does not change the average speckle size but change the number of observed speckles. As shown in Figure 2.15, with small $F^{\#}$, the number of observed speckles reduced, but each speckle's size keeps same.

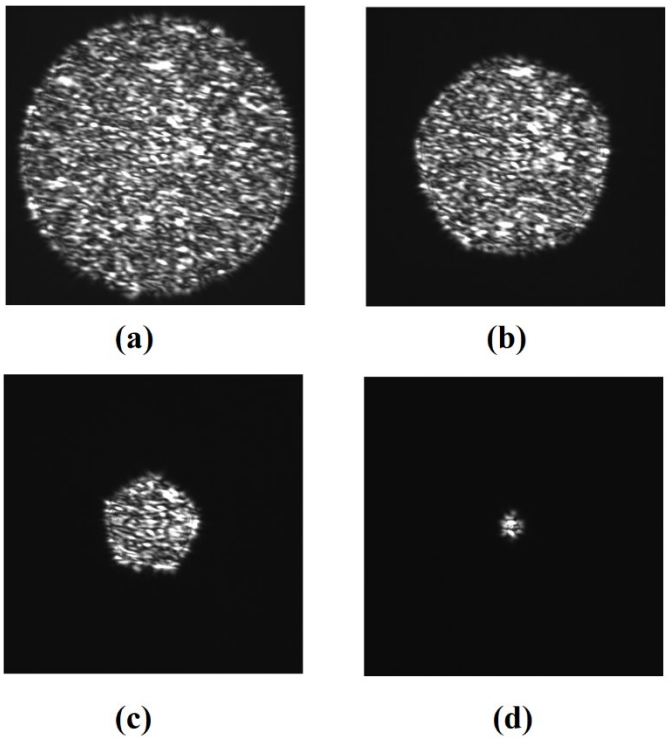

Figure 2.15 Captured speckle patterns with different F#. (a) F#=1.4 (b) F# = 2.0 (c) F#=4.0 (d) F#=8.0

In order to verify the theory experiment has been conducted, we analyzed the autocorrelation of the actual captured speckle pattern to obtain the average speckle size in a defocused camera. The wavelength λ of the laser source is $650nm$, and the beam spot size is about $5mm \times 2.5mm$. First the camera without lens is located $70cm$ away from the object to capture speckles in free space. The exposure time is $20\mu s$ and the image size is $2048 \times 1088\, pixels$. The captured speckle pattern is shown as Figure 2.16. The autocorrelation result of the captured image is shown in Figure 2.17

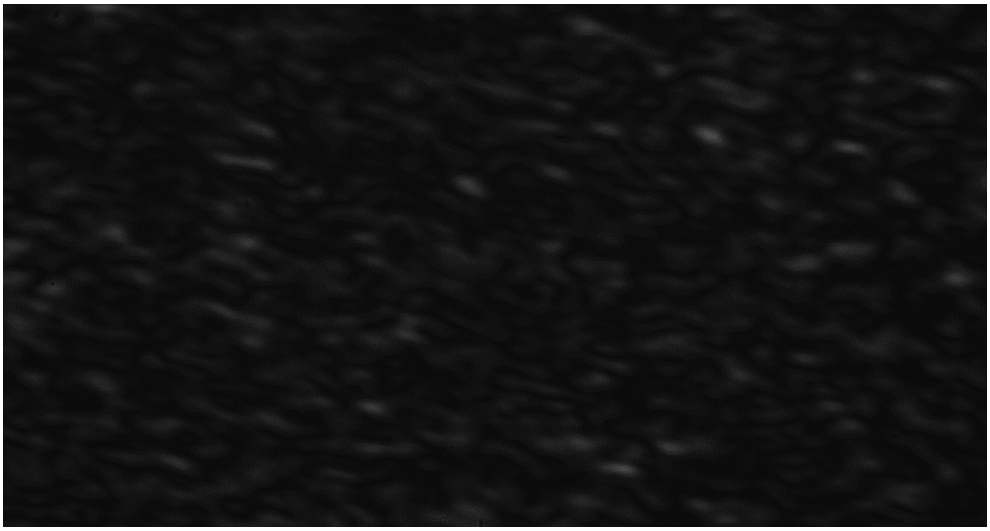

Figure 2.16 Speckle pattern captured by bare sensor with exposure time of 20μs and image size of 2048×1088 pixels

In this situation $\lambda = 650nm$, the object-sensor distance $\Delta Z = 70cm$. The shape of the laser spot is not a round, so the size is expressed as $d_x = 2.5mm$ and $d_y = 5mm$. According to Equation (2-34), theoretically the average speckle size captured by bare sensor should be $222um \times 111\mu m$. The pixel size of our camera is $5.5\mu m \times 5.5\mu m$. Thus, the pixel size of the captured speckle pattern should be $40\, pixels$ in X-direction and $20\, pixels$ in Y-direction. From the Figure 2.17(c) and Figure 2.17(d) one can see that the actual captured speckle size is basically the same as the theoretical value. The error may come from the inaccuracy of the laser

spot size and the object-sensor distance.

Then the speckle pattern is captured by camera with lens. In this time the camera is located $1m$ away from the object. However, the camera is defocused with $\Delta L = 70cm$, which is the location of the bare sensor in the last shot. Obviously the focusing length u of camera is $30cm$. The captured image is shown as Figure 2.18.

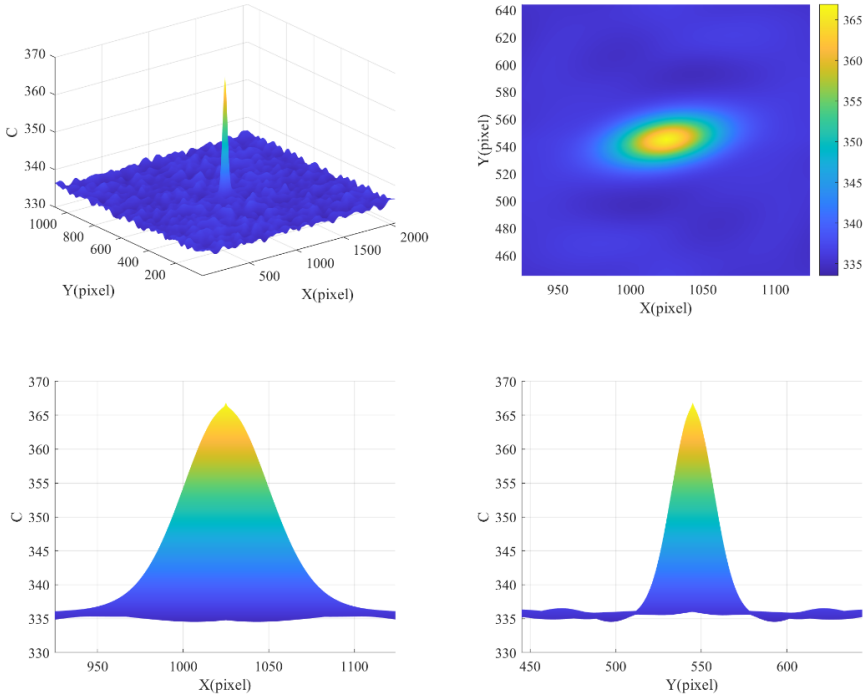

Figure 2.17 Autocorrelation of speckle pattern captured by bare sensor

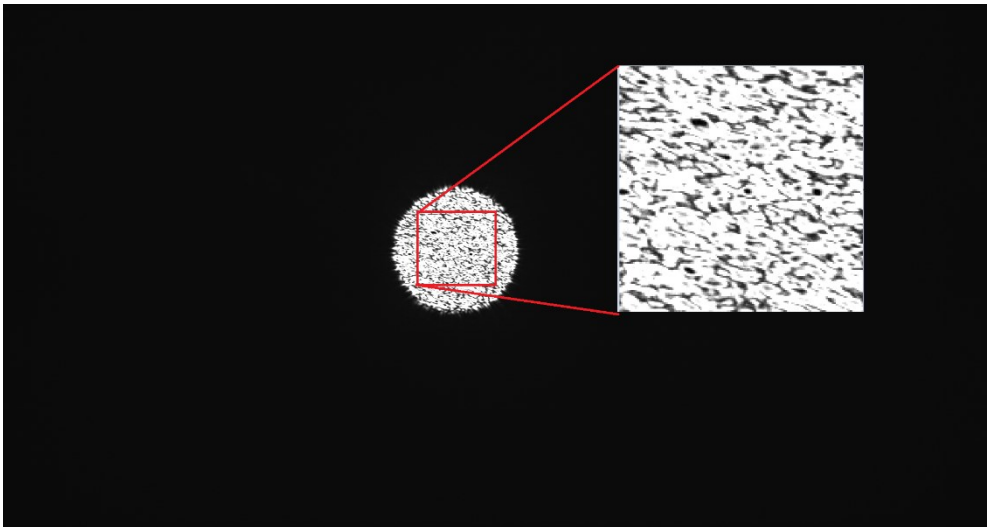

Figure 2.18 Speckle pattern captured by camera with lens, the lens is defocused

Since the lens focus the light on a small area of the imaging sensor, this time the speckle pattern only occupies a small area of the sensor. We take $160 \times 160\ pixels$ area of the speckle pattern for analyzation. The autocorrelation result of the template is shown as Figure 2.19.

Theoretically, the average size speckles captured by the sensor can be equivalent to the free space speckle size on the focal plane after magnified by the camera. In this situation the focusing distance $u = 30cm$ and the lens focal length $f = 25mm$. Thus, the magnification M of the camera is 0.091. The average speckle size on the sensor is $20\mu m \times 10\mu m$, or $3.67\ pixels$ in X-direction and $1.81\ pixels$ in Y-direction. Our experiments basically verify the correctness of the speckle size calculation formulas in the free optical space and defocused camera.

Let us consider Equation (2-41) again. In speckle sensing method, the camera focusing length is usually much longer than the lens focal length, i.e., $u \gg f$. In this situation Equation (2-42) can be approximately expressed as:

$$s_s = \frac{1.2\lambda f}{d\Delta s} \times \frac{\Delta Z}{u} \qquad (2-45)$$

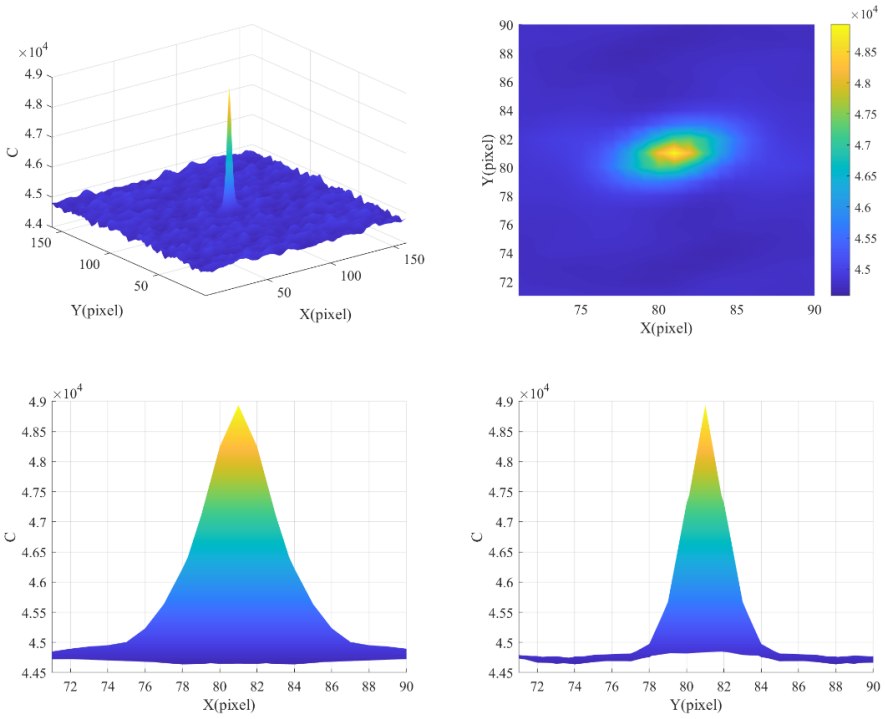

Figure 2.19 Autocorrelation result of speckle pattern captured by camera with defocused lens

It could be seen that in a defocused camera the captured speckle size is determined by several parameters. Among them, parameters like laser wavelength λ, laser spot diameter d, lens focal length f and camera pixel size Δs are usually fixed. The average speckle size s_p is mainly determined by the defocus component $\Delta Z/u$. ΔZ is the object-focal plane distance and u is the focal plane-camera distance. $\Delta Z/u$ can be used to indicate the degree of defocus, and it can be easily adjusted. In the next experiment, speckle patterns under different degree of defocus are captured. The camera is located $1m$ away from object, and the camera's focusing lengths are settled to be $0.5m$, $0.4m$, $0.3m$ and $0.25m$ respectively. In other words, the defocus levels $\Delta Z/u$ are 1.00, 1.50, 2.33 and 3. The situation and the captured speckle patterns are shown in Figure 2.20. Figure 2.21 shows the autocorrelation result of the captured speckle patterns. From the width of the autocorrelation result one can see

that the average speckle size in horizontal and vertical direction increase with the defocus level. In the later part the relationship between defocus level and the image properties will be further discussed.

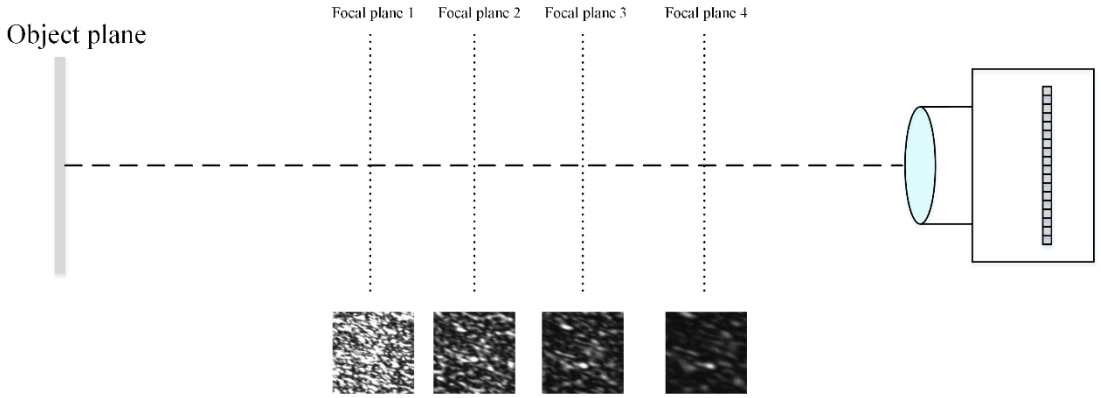

Figure 2.20 Captured speckle patterns with different amount of camera defocusing

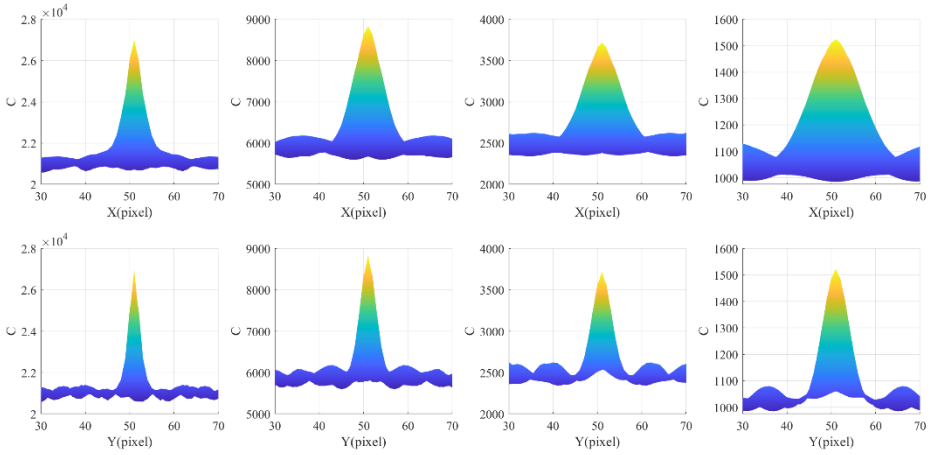

Figure 2.21 Autocorrelation result of captured speckle patterns with different amount of camera defocusing

2.4 Motion sensing model using speckle image

Speckle motion sensing method uses speckle images reflected from object illuminated by laser to measure the object motion, and usually a defocused camera is adopted to capture the

speckle images. In this part we will introduce the motion model of speckle image in a defocused camera. We will analyze the motion of speckle images when the object undergoes in-plane translation, axial translation, and rotation. Generally, the vibration object does not show in-plane rotation, this situation will not be discussed. Because the camera is defocused, the captured speckle pattern presents only shifting or vibrating rather than pattern changing when the object moves. This is significant because it allows the extraction of the trajectory movement by tracking the speckle motion.

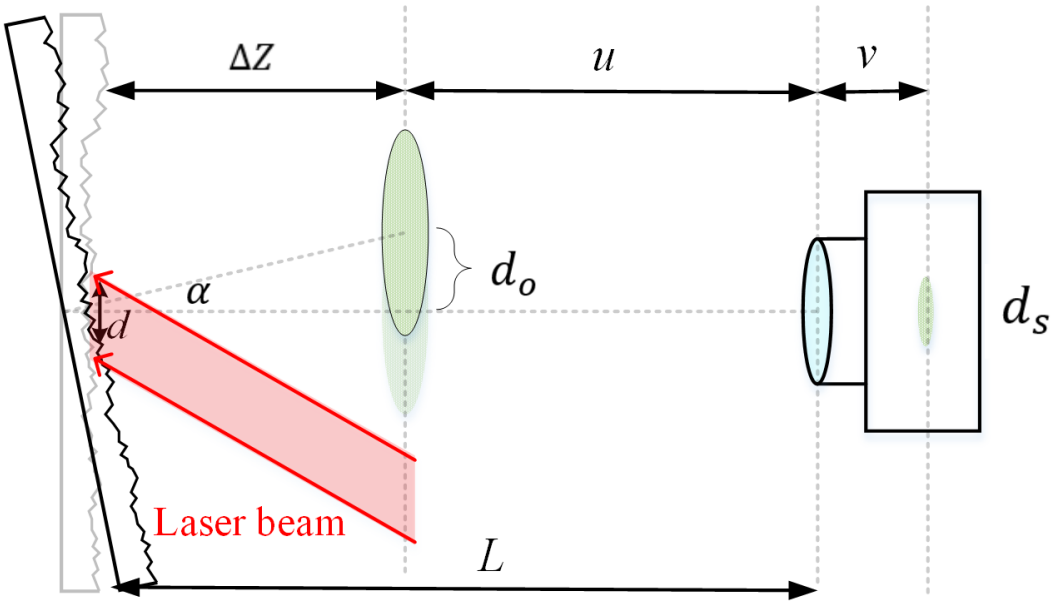

Figure 2.22 Schematic description of the defocus speckle sensing method

Figure 2.22 shows a defocused camera capturing speckle pattern reflected from tested object. At first the slightly defocus situation will be discussed, here assuming $\Delta Z = Z_1$, and the distance Z_1 between object plane and camera focal plane is small. The coordinates of the object plane are denoted by (x, y), whereas the coordinates of the speckle plane in the space are denoted by (x_o, y_o). The axial axis is denoted by Z. When the object is illuminated by laser beam with diameter D and wavelength λ, the laser will be scattered into space randomly due to the rough surface. We might regard each point (x, y) of the illuminated area on the

object surface as a virtual light source, which emits light with random phases into space. For an individual virtual light source, when its light travels to point (x_o, y_o) in the near field, its optical path ΔL is:

$$\Delta L = \sqrt{(x - x_o)^2 + (y - y_o)^2 + {Z_1}^2} \tag{2-46}$$

The phase difference $\Delta\varphi$ is:

$$\Delta\varphi = \frac{2\pi\Delta L}{\lambda} = \frac{2\pi}{\lambda}\sqrt{(x - x_o)^2 + (y - y_o)^2 + {Z_1}^2} \tag{2-47}$$

In the object plane, assuming the phase at point (x, y) is $\varphi(x, y)$. When its light travels to (x_o, y_o), the phase is:

$$\varphi(x_o, y_o) = \varphi(x, y) + \Delta\varphi \tag{2-48}$$

Since the amplitude at (x_o, y_o) is determined by all virtual light source in object plane, its amplitude distribution equals to the Fresnel integral performed over all points (x, y), which is:

$$\begin{aligned} T_m(x_0, y_0) &= \iint \exp\{i[\varphi(x, y) + \Delta\varphi]\}dxdy \\ &= \iint \exp\left[i\varphi(x, y)\right]\exp\left[\frac{\pi i}{\lambda Z_1}((x - x_o)^2 + (y - y_o)^2)\right]dxdy \\ &= A_m(x_o, y_0)\exp[i\psi(x_o, y_0)] \end{aligned} \tag{2-49}$$

Here paraxial approximation is assumed. The speckle on the focal plane is then captured by camera, the intensity $I(x_s, y_s)$ at (x_s, y_s) on the sensor is:

$$I(x_s, y_s) = \left|\iint T_m(x_0, y_0)h(x_o - mx_s, y_o - my_s)\right|^2 \tag{2-50}$$

where h is the spatial impulse response and m is the inverse of the magnification factor of the camera. Here the expression of m is:

$$m = \frac{(Z - Z_1) - f}{f} \qquad (2-51)$$

In remote sensing situation usually the object-lens distance is much longer than other distance. i.e., $Z \gg f$ and $Z \gg Z_1$.

The small area of the object illuminated by laser is regarded as rigid body. In other word, deformation is not considered when the illuminated area vibrates. As mentioned before, the object motion mode can be classified into three types: in-plane transverse, axial movement, and tilting. The three types of motion cannot be separated, and a general consideration is these three types of motion simultaneously occur during object vibration. However, we can separately discuss the response of the captured speckle pattern to a single mode of motion.

For in-plane transverse, i.e., the point (x, y) in the object plane undergoes a transversal movement $(\delta x, \delta y)$. According to Equation (2-47) the amplitude distribution T_m on the focal plane will simply shift by $(\delta x_o, \delta y_o)$, and the amount is equal to $(\delta x, \delta y)$. The shift $(\delta x_o, \delta y_o)$ is usually very small when the object vibrates, and the small amount is then demagnified by the camera so that the shift on the imaging sensor is barely detectable.

For the axial movement, the displacement caused by axial vibration can be regarded as a small value change of Z_1. Assume the amount of displacement is δz. As proved in part 2.3.3, when object undergoes axial displacement, the defocusing level is changed, which changes the average speckle size. Therefore, the captured speckle pattern is scaled. However, in vibration situation δz is usually very small, i.e., $\delta z \ll Z$. The change is very small for the final imaged speckle pattern. In conclusion, the axial vibration of the object has almost no effect on the defocus, and the change of the speckle image can also be ignored.

For the tilting motion, assume the tilting angle in x direction and y direction is α_x and

α_y, the object shift amount β_x and β_y along axial direction caused by tilting can be expressed as:

$$\beta_x = \frac{4\pi \tan \alpha_x}{\lambda} \quad (2-52)$$

$$\beta_y = \frac{4\pi \tan \alpha_y}{\lambda} \quad (2-53)$$

Also, the speckle amplitude at (x_o, y_o) on the focal plane is obtained by Fresnel integral:

$$A_m(x_o, y_o) = \left| \iint \exp\left[i\varphi(x,y)\right] \exp\left[i\left(\beta_x x + \beta_y y\right)\right] \exp\left[\frac{\pi i}{\lambda Z_1}\left((x-x_o)^2 + (y-y_o)^2\right)\right] dxdy \right| \quad (2-54)$$

Compared with Equation (2-47), the integral function is changed because of the tilting component $\exp\left[i\left(\beta_x x + \beta_y y\right)\right]$. Therefore, the speckle distribution is also changed. The difference is magnified by imaging sensor, so the captured speckle pattern looks totally different when object undergoes tilting motion.

Based on the analyzation above, when the object vibrates, the three types of motion determines that the speckle pattern is varied randomly. Actually, when Z_1 is very small, 1) the average speckle size $\sigma = 1.22(\lambda Z_1/d)$ at focal plane is very small and 2) the demagnification $m = [(Z - Z_1) - f]/f$ is very large. Thus, the speckle pattern is almost undistinguishable on the imaging sensor.

Now we turn to large defocus, here assuming $\Delta Z = Z_2$, which means the distance between object and camera focal plane is determined by a large value Z_2. In this situation the demagnification factor is:

$$m = \frac{(Z - Z_2) - f}{f} \quad (2-55)$$

Since the defocus amount increases, the demagnification factor m becomes much smaller

compared with that in slightly defocus situation. Besides, the camera focal plane is far away from object, in this situation Equation (2-47) became:

$$T_m(x_o, y_o) = \iint \exp[i\varphi(x,y)]\exp\left[\frac{-2\pi i}{\lambda Z_2}(xx_0 + yy_0)\right]dxdy$$
$$= A_m(x_o, y_o)\exp[i\psi(x_o, y_o)] \quad (2-56)$$

and

$$I(x_s, y_s) = \left|\iint T_m(x_0, y_0)h(x_o - mx_s, y_o - my_s)\right|^2 \quad (2-57)$$

Therefore, for in-plane transverse the speckle pattern is almost unchanged since it does not affect the amplitude of the Fourier transform and magnification of h is much smaller. For axial displacement, the vibration amount is still very small compared with the defocus amount, so axial movement does not change the speckle pattern, either.

For tilting motion, the amplitude distribution on the focal plane is:

$$A_m(x_o, y_o) = \left|\iint \exp[i\varphi(x,y)]\exp\left[i\left(\beta_x x + \beta_y y\right)\right]\exp\left[\frac{-2\pi i}{\lambda Z_2}(xx_o + yy_o)\right]dxdy\right| \quad (2-58)$$

still

$$\beta_x = \frac{4\pi \tan\alpha_x}{\lambda} \quad (2-59)$$

$$\beta_y = \frac{4\pi \tan\alpha_y}{\lambda} \quad (2-60)$$

Further deriving of Equation (2-56), we can get:

$$A_m(x_o, y_o) = \left|\iint \exp[i\varphi(x,y)]\exp\left[i\left(\beta_x x + \beta_y y\right)\right]\exp\left[\frac{-2\pi i}{\lambda Z_2}(xx_o + yy_o)\right]dxdy\right|$$
$$= \left|\iint \exp[i\varphi(x,y)]\exp\left[i\left(\frac{4\pi \tan\alpha_x}{\lambda}x + \frac{4\pi \tan\alpha_y}{\lambda}y\right)\right]\exp\left[\frac{-2\pi i}{\lambda Z_2}(xx_o + yy_o)\right]dxdy\right|$$

$$= \left| \iint \exp\left[i\varphi(x,y)\right] \exp\left(\frac{4\pi i \tan\alpha_x}{\lambda} x - \frac{2\pi i}{\lambda z_2} x x_o\right) dx \cdot \exp\left(\frac{4\pi i \tan\alpha_y}{\lambda} y - \frac{2\pi i}{\lambda z_2} y y_o\right) dy \right|$$

$$= \left| \iint \exp\left[i\varphi(x,y)\right] \exp\left[\frac{2\pi i}{\lambda z_2}(2Z_2 \tan\alpha_x - x_o)x\right] dx \cdot \exp\left[\frac{2\pi i}{\lambda z_2}(2Z_2 \tan\alpha_x - x_o)x\right] dy \right| \quad (2-61)$$

From the analyzation we can see that when the camera is strongly defocused, two of the three types of movement are negligible and only tilting motion affects the speckle motion. Assuming the tilting angle is α. Since α is very small, the shift amount d_o of speckle pattern on the focal plane is:

$$d_o = Z_2 \tan\alpha = Z_2\alpha \quad (2-62)$$

The shift amount d_s on the imaging sensor is:

$$d_s = \frac{Z_2\alpha}{m} = \frac{Z_2 f}{(Z - Z_2) - f}\alpha \quad (2-63)$$

Here let $u = Z - Z_2$, which is the focusing distance of camera that indicates the camera-focal plane distance of camera. Again, one can assume that u is much longer than lens focal length f, i.e., $u \gg f$. In this way Equation (2-61) can be approximately expressed as:

$$d_s = \frac{Z_2}{u} f\alpha \quad (2-64)$$

From the simplified expression one can see that the observed speckle shift d_s is linearly proportional to object tilting angle α. This proves that the speckle will linearly shift when object vibrates. Besides, the sensitivity of the speckle motion is determined by the defocus factor Z_2/u of the camera. For the same vibration, more intense speckle movement can be observed with greater defocus. This is significant for speckle sensing method.

2.5 Experimental Verification

2.5.1 Verification for speckle image statistical properties in defocus

imaging method

In the previous part of this chapter, we used mathematical methods to analyze the statistical characteristics of speckle images and the image characteristics of speckle in the defocused camera. The image characteristics of the speckle, including the intensity, contrast, and average size, are very important to the measurement results, and these characteristics will be affected by various parameters. In Figure 2.23 the concerned speckle properties and the factors that determines these properties are listed. Speckle properties are mainly affected by the camera parameters, for example, the defocus level, the camera focal length and the shutter time. Besides, laser and tested object also play important roles in speckle images.

In the next part, we will use the method of experimentally collecting speckle images + speckle image numerical analysis to study the effects of different parameters on speckle characteristics. When a specific characteristic is discussed, it is ensured that other experimental conditions are the same. For all experiment same camera is used, as well as the relative position of laser-object-camera, where the camera and laser are located $4m$ away from the object.

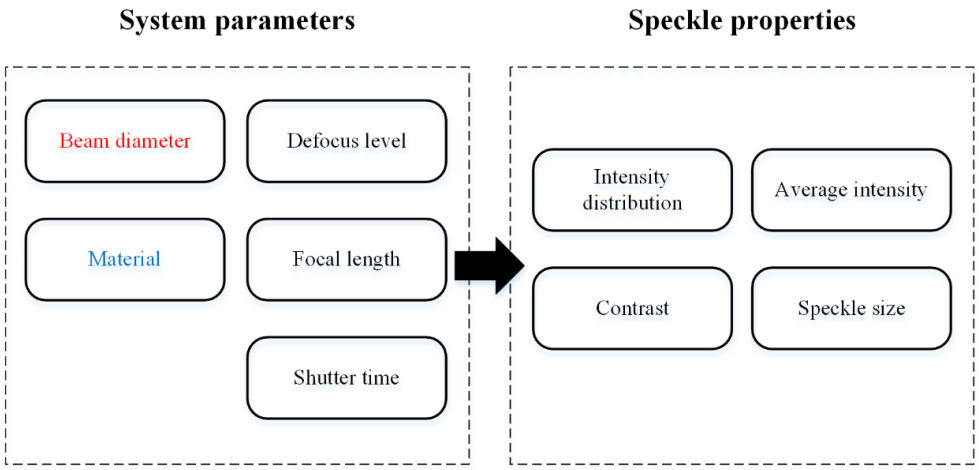

Figure 2.23 Speckle image characteristics and some parameters that affect speckle image characteristics

2.4.1.1 Defocus level

First speckle patterns under different camera defocus level have been captured. The camera defocus level is adjusted by changing the camera focusing length. In this experiment the camera focusing length is settled to be $0.4m$, $0.3m$, $0.25m$, and $0.15m$ respectively, which means the defocus level is gradually increasing. The defocus value is defined as the ratio of the object-focal plane distance and the focal plane-camera distance, which means the defocus levels are 9, 12.3, 15, and 25.6, respectively. The captured speckle patterns are shown in Figure 2.24, the intensity distributions of the captured speckle patterns are shown in Figure 2.25. In Table 2.1the statistical properties, including the mean intensity, image contrast, and the average speckle size in horizontal and vertical direction are listed.

According to Equation (2-42), the average speckle size is proportional to the defocus level of the camera. It can be intuitively seen from the images that the image is bright when the defocus level is slight, and a large number of speckles can be captured in the image window. From the mathematical analysis results, it can be seen that many speckles overlap together, resulting in a high mean intensity of the image, and the intensity distribution of the image is also concentrated in the area with gray value near 255. At the same time, the overlap of speckles results in low contrast of the speckle image because the image is too bright. Finally, it can be seen from the analysis results that the average size of speckles is small when the camera is slightly defocused.

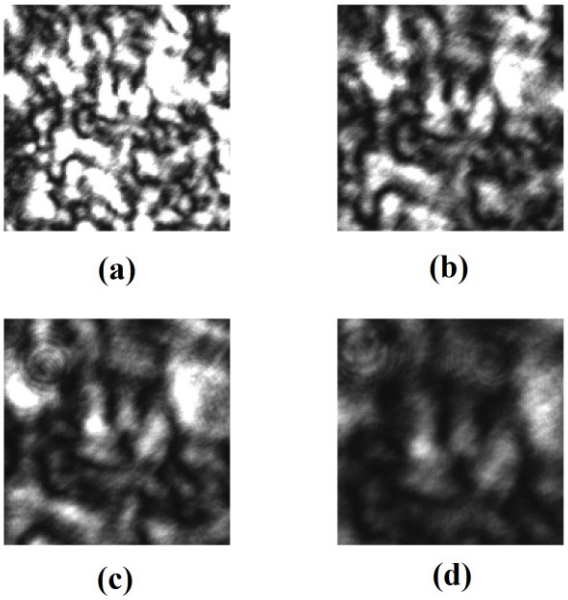

Figure 2.24 Captured speckle patterns under different camera defocus level. (a) Defocus level = 9 (b) Defocus level = 12.3 (c) Defocus level = 15 (d) Defocus level = 25.6

As the defocus level increases, the size of the speckles gradually becomes larger and separated away from each other, the mean intensity of the image becomes lower, and the intensity distribution gradually conforms to the concentration toward the dark part. At the same time, the contrast of speckle is also gradually increasing. The variation of the statistical characteristic of speckle with defocus basically matches with the theoretical analysis.

Based on the result above, it is not difficult to see that the appropriate degree of defocus is very important for the speckle sensing method. If the defocus level is too slight, the speckle size will be very small and aliased together on the same pixel, the entire image will be too bright with very low contrast. In this situation the image shift can neither be observed nor calculated. On the other hand, increasing the defocus can separate the speckles from each other on the sensor, so that high-contrast speckle pattern can be observed. In this case, one can use speckle to sense the motion of the object. However, it should be pointed out that when the

defocus is too large, the speckle size may exceed the image size, the brightness and contrast of the image also decrease simultaneously. In this situation, the displacement of the image is also difficult to be observed and calculated. Therefore, for the speckle sensing method, it is necessary to adjust the defocus according to the actual parameters to obtain the ideal measurement effect.

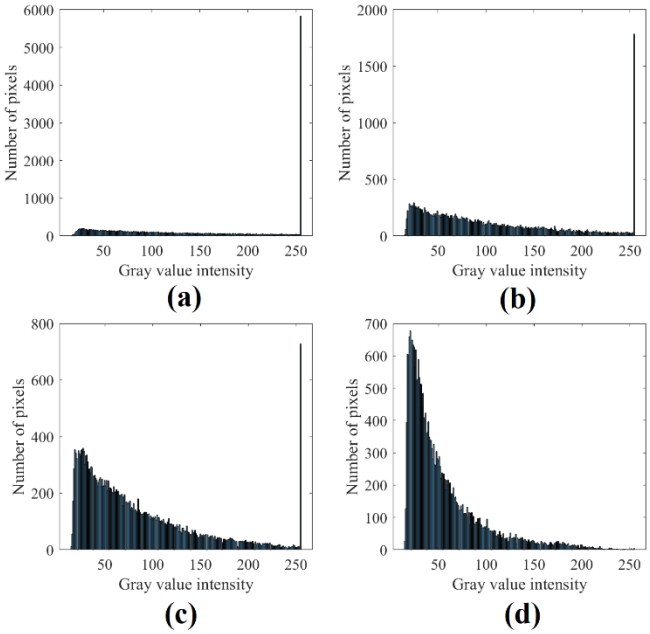

Figure 2.25 Intensity distribution of captured speckle patterns under different camera defocus level. (a) Defocus level = 9 (b) Defocus level = 12.3 (c) Defocus level = 15 (d) Defocus level = 25.6

Table 2.1 Statistical properties of speckle patterns with different defocus level

Defocus level	9	12.3	15	25.6
Mean intensity	139.4223	103.1657	82.1707	54.375
Contrast	0.6043	0.687	0.7185	0.7191
Average size (Horizontal, in pixel)	7.2659	9.6485	11.6242	13.27
Average size (Vertical, in pixel)	9.315	12.1031	15.0785	18.5102

2.4.1.2 Focal length

In this experiment a focal length adjustable lens is used, and the speckle pattern is captured with different $f = 20mm$, $f = 30mm$, $f = 50mm$ and $f = 75mm$ respectively. Here the camera focusing distance is settled to be $1m$ for all time. The captured speckle patterns are shown in Figure 2.26, whereas the intensity distribution of the speckle area is shown in Figure 2.27. In Table 2.2 the statistical properties of the speckle pattern are listed.

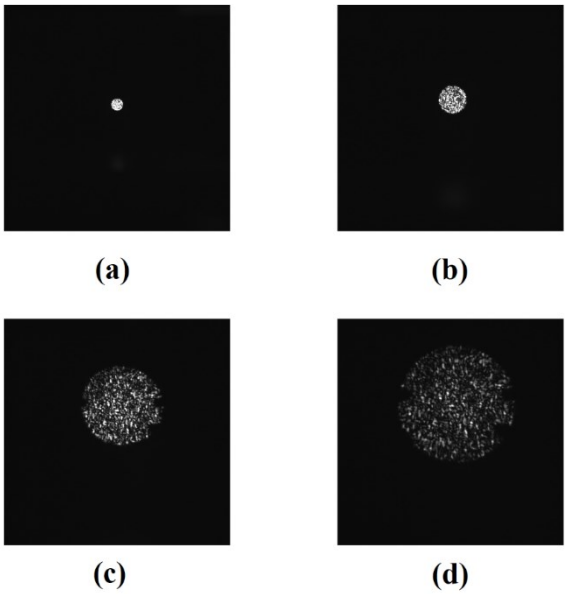

Figure 2.26 Captured speckle patterns with different lens focal length f. (a) f=20mm (b) f=30mm (c) f=50mm (d) f=75mm

The focal length of the camera affects the properties of the speckle image by affecting the magnification of the camera. According to Equation (2-38), the magnification factor M increases with the focal length f when the focusing distance u of the camera is fixed. It can be seen from the captured images that the area occupied by the speckle pattern on the imaging sensor increases with the increase of focal length f. On the same time, it can be seen from Equation (2-42) that the average speckle size also increases with the increase of f. Besides, when f is small, spatial aliasing occurs again due to the crowding of the speckle pattern in a

small area. The gray value of most pixels is very high, and the entire image shows high brightness but low contrast, which is not suitable for observing image motion. On the other hand, when f becomes large the speckles locate in a large area on the imaging sensor. A high-contrast and clear speckle pattern could thus be observed.

In summary, lens focal length f is also crucial for observing clear speckle images. Generally, f is determined by the actual measured distance. When the measurement distance is relatively short, people can choose the lens with shorter f, while when the tested object locates at a long distance, lens with long f is usually needed to achieve the ideal measurement effect.

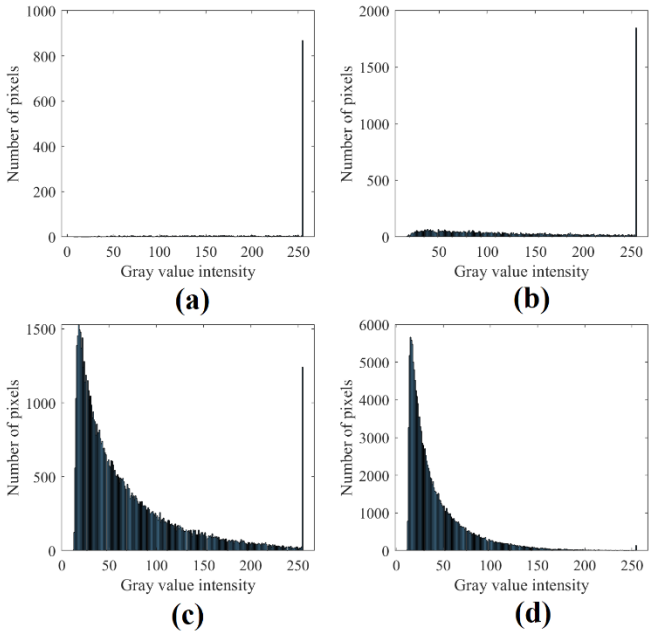

Figure 2.27 Intensity distribution of speckle patterns with different lens focal length f. (a) f=20mm (b) f=30mm (c) f=50mm (d) f=75mm

Table 2.2 Statistical properties of speckle patterns with different lens focal length

Focal length (mm)	20	30	50	75
Mean intensity	205.975	139.4697	67.6053	42.1123
Contrast	0.3302	0.5867	0.8135	0.782
Average size (Horizontal, in pixel)	1.7732	3.4414	5.4784	7.4272
Average size (Vertical, in pixel)	1.9503	4.3552	7.3375	9.6509

2.4.1.3 Shutter time

In this experiment speckle patterns captured with different camera shutter time are investigated. The camera shutter time is settled to be $20\mu s, 100\mu s, 500\mu s$ and $1000\mu s$ respectively. It shall be noted that this experiment is equivalent to investigating the effect of laser power on the speckle pattern, since the laser power and exposure time together determine the luminous flux received by the pixels on the sensor per unit time. By the way, the above shutter time can be regarded as the corresponding exposure time when the camera captures images at the frame rate of 50000 fps, 10000 fps, 2000 fps and 1000 fps. The captured speckle patterns are shown in Figure 2.28, the intensity distributions of the captured speckle patterns are shown in Figure 2.29. In Table 2.3 the statistical properties, including the mean intensity, image contrast, and the average speckle size in horizontal and vertical direction are listed.

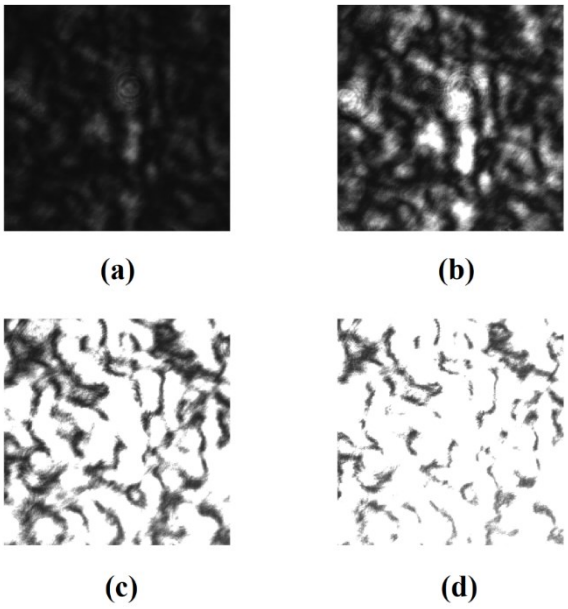

Figure 2.28 Captured speckle patterns with different shutter time. (a) 20μs (b) 100μs (c) 500μs (d) 1000μs

Different from camera defocus level and focal length, changing of the camera's shutter time does not affect the parameters of the camera. It can be intuitively seen from the images that the speckle pattern does not change but only becomes brighter with longer shutter time. From Table 2.3 people can see that the speckle size keeps the same for different shutter time. Again, if the shutter time is too long the imaging sensor is over-exposure. The gray value of most pixels falls in the high-brightness area and the image is too brighter so that the contrast decreases. This could be the reason that why the mathematical analysis result of the average speckle size became smaller.

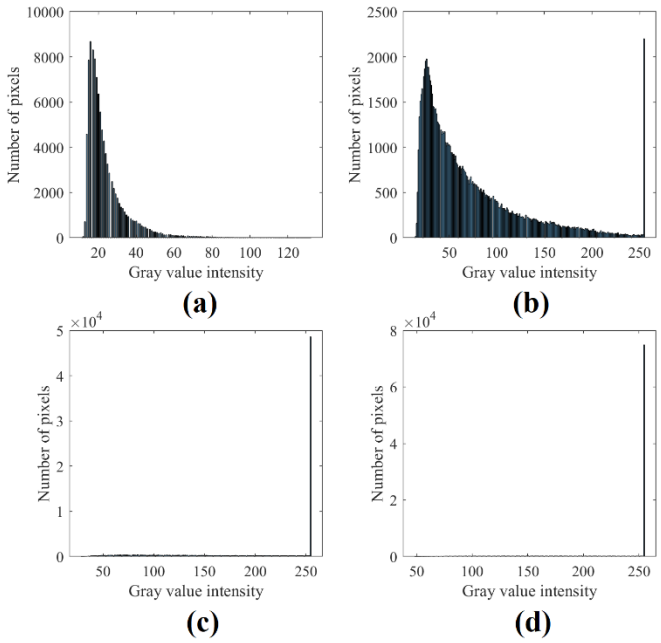

Figure 2.29 Intensity distribution of speckle patterns captured with different shutter time (a) 20μs (b) 100μs (c) 500μs (d) 1000μs

In actual situations, the camera's shutter time t is first determined by the camera's frame rate. For example, if the camera's frame rate is 8000 fps, then t must less than $125\mu s$, i.e., $t_{min} < t < 1/FPS$, where t_{min} is the minimum shutter time of the camera. Since the shutter time does not affect the imaging properties of the camera and only changes the brightness and contrast of the image, the shutter time can be adjusted within the above range to compensate or weaken the light intensity, so as to obtain a clear speckle image with high contrast.

Table 2.3 Statistical properties of speckle patterns captured with different shutter time

Shutter time (μs)	20	100	500	1000
Mean intensity	23.9364	72.8927	193.4305	229.6864
Contrast	0.466	0.7551	0.3786	0.2184
Average size (Horizontal, in pixel)	15.5703	15.7383	12.2431	9.1876
Average size (Vertical, in pixel)	19.4373	20.3375	14.9853	10.6119

2.4.1.4 Beam diameter

In this experiment the object is illuminated by an expanded laser beam rather than a laser spot. The wavelength of the laser is $650nm$ and the diameter of the expanded laser beam is $50mm$. The other parameters are the same as in the previous experiment, and the shutter time is $500\mu s$. The captured speckle pattern is shown in Figure 2.30(a), while the intensity distributions of the captured speckle pattern is shown in Figure 2.30(b).

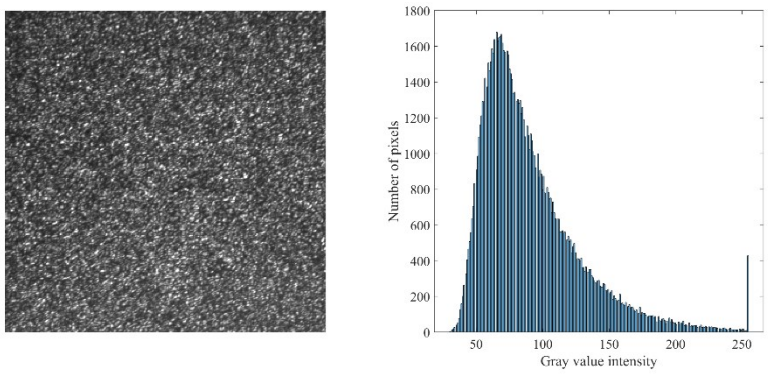

Figure 2.30 (a) Captured speckle pattern with expanded laser beam. (b) Intensity distribution of the captured speckle pattern

Compared with illuminating object with laser spot, the most obvious change in the speckle pattern captured the expanded laser illumination is the average size of the speckles. According to Equation (2-34), the increasing diameter of the illuminated area on the object results in the decreasing average size of the speckles. Compared to Figure 2.28(c), the speckles are smaller and denser is Figure 2.30(a). In fact, by mathematical analysis, it can be found that the average size of the speckle 2.0699 pixels and 1.6401 pixels in horizontal and vertical direction.

For speckle sensing method, using of expanded laser beam can increase the coverage area of the light source at the detection end, while reducing the optical power density per unit area

to avoid laser damage. However, it is necessary to pay attention to the negative effects of expanding the beam diameter, such as the decrease of the light intensity at the receiving end and the reduction of the speckle size.

2.4.1.5 Object material

In the last test, we tried to use the laser to illuminate a paper cup and capture the speckle pattern reflected from it, where in the previous tests the object is Aluminum plate. The laser is spot laser, and the shutter time is $100\mu s$. The captured speckle pattern is shown in Figure 2.31(a), while the intensity distributions of the captured speckle pattern is shown in Figure 2.31(c). For reference the speckle pattern reflected from Aluminum plate and its intensity distribution are shown in Figure 2.31(b) and Figure 2.31(d).

It can be seen from the image that the brightness of the captured speckle pattern is very low because the reflectivity of the paper cup is lower than that of the metal plate. The intensity distribution of speckle is mostly concentrated in the dark area. The results of numerical analysis show that the average intensity of the speckle pattern reflected from the paper cup is only 21.7474. Since the picture is dark overall, the contrast is only 0.2006. The average size of speckles is 15.8599 pixels and 19.8455 pixels in horizontal and vertical direction.

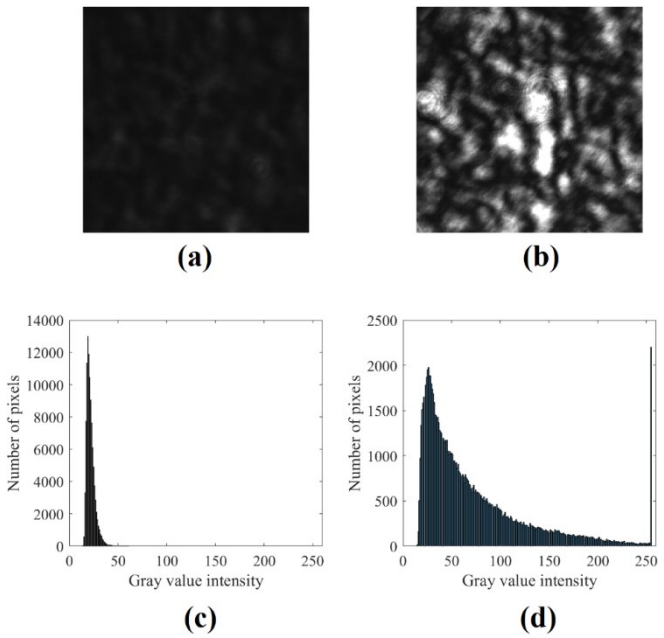

Figure 2.31 Captured speckle patterns and intensity distributions. (a) Speckle pattern reflected from paper cup. (b) Speckle pattern reflected from Aluminum plate (c) Intensity distribution of speckle pattern of paper cup (d) Intensity distribution of speckle pattern

Obviously, it is very difficult to capture clear speckles from objects with low reflectivity at a high frame rate (low shutter time). In practical applications, it is necessary to adjust various parameters, such as laser power, lens parameters, etc., to compensate for the loss of light energy caused by reflectivity.

2.5.2 Verification for speckle motion model in defocused camera

In 2.4 a theoretical analysis for speckle motion model in defocused camera has been illustrated. The conclusion is that in a defocused camera, the speckle motion is mainly occurred by object tilting although the object motion usually contains different modes. Here experiments have been conducted to verify the theory. We put an Aluminum plate on an electric platform and simulate different motion modes by controlling the displacement/rotation of the electric

platform. A camera with lens is adopted to record dynamic speckle sequences under different motion modes. Through the analysis of the speckle displacement, the difference in the sensitivity of speckle images under different object motion modes is obtained. The camera with a $650nm,\ 5mW$ laser is located $4m$ away from the object. The focal length of the lens is $f = 75mm$. The PI M-403.4PD linear stage is used to create linear translation for the object, whereas the NSK rotation stage is used for tilting experiment.

For in-plane translation and axial translation, the linear stage is translated by $1mm$ along the horizontal and axial directions with a speed of $0.1mm/s$, simulating the small-scale movement of the object. For object tilting, we hope that the object can be rotated by a tiny angle of the same magnitude, and the tilting angle is determined by the length of the object and the position of the laser spot. As shown in Figure 2.32, the laser spot illuminates at the edge of the Aluminum plate, and the distance from the plate's edge to the center (axial of rotation) is $4cm$. Assuming the amount of axial translation caused by object tilting is $1mm$, therefore the tilting angle can be approximately calculated, which is $1.43°$.

In this way, we can control the object to transverse in plane, transverse along axial and tilt with approximately same amount, and record video of the speckle motion. The speckle motion is then analyzed, and the result is shown in the Figure 2.33.

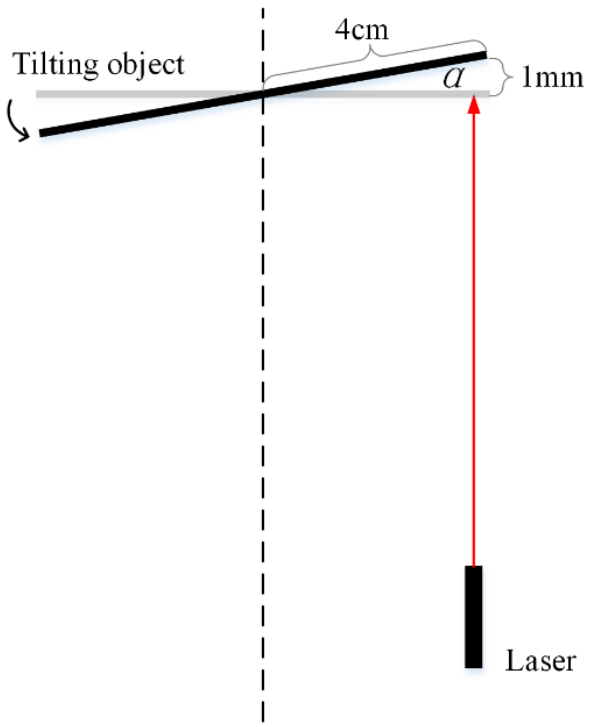

Figure 2.32 Methods for determining object tilting angle

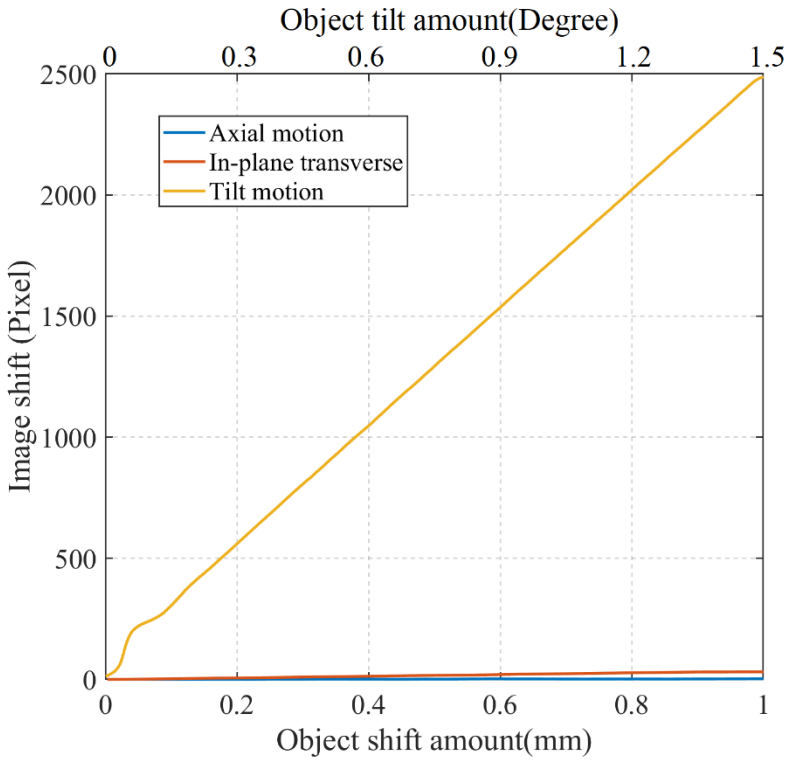

Figure 2.33 Shift amount of speckle image caused by different motion mode

From the analysis results, it can be seen that in a defocused camera, the captured speckle image is almost static when the object translates along the axial direction. The in-plane translation of the object causes the speckle image to move but also the motion is very slight. When the object is translated by $1mm$, the total displacement of the speckle image is about 31 pixels. However, compared to the speckle changes caused by object tilting, both of the two motions are negligible. From the result one can see that when the object is tilted by $1.43°$ (equivalent to $1mm$ along axial direction), the shift amount of speckle image is about 2491 pixels. The displacement of the speckle image caused by tilting is about 80 times the displacement caused by in-plane translation, and more than 660 times the displacement caused by axial translation. Actually, the in-plane transverse almost does not occur when the object is vibrating. In the remaining two motion modes, the speckle image displacement caused by tilting is absolutely dominant. Therefore, it can be considered that the theoretical model of the object-image motion relationship described by the Equation (2-62) is basically valid.

This experiment result is important because it told us two important facts: First, it told us in a defocused camera, the speckle image can amplify the subtle object vibration, convert it into obvious speckle motion. This ensures that we can use speckle image to observe object vibration even it is a very subtle motion. Second, even the object vibration is complex, including motion of multiple modes, the speckle motion is only sensitive to the tilt motion of the object, which means, if we use the speckle image to sense object vibration, the speckle motion is a simple linear motion. We only need to deal with the simple speckle movement to restore the audio signal. And furthermore, by change the camera defocusing, we can control the amplitude of image motion. These two facts allow us to use speckle image to sense subtle and complex object vibration.

2.6 Chapter Conclusion

In this chapter we introduced the optical principles related to speckle. Currently, it is well-accepted that speckle generates from the interference of the scattered light when the laser is irradiated on the surface of the rough object. Due to the randomness of this process, the specific distribution of the speckle pattern cannot be calculated. When studying speckle, people usually start from a statistical point of view and analyze the statistical characteristics of a large number of components with random amplitude and phase through mathematical methods, so as to obtain the image characteristics of the speckle pattern produced by multi-beam interference. It includes important characteristics such as the intensity distribution (corresponding to the first-order statistical characteristics) and the average size (corresponding to the second-order statistical characteristics) of the speckle patterns that we are more concerned about.

In our proposal a digital camera is used as the sensor. Therefore, this chapter then introduces the imaging principle of the camera, discusses the image characteristic (speckle size) and motion characteristic of the speckle pattern in a defocused camera. We also conducted experiments to verify the variation of image characteristics in the defocused camera and analyzed the influence of various parameters on these characteristics. In general, the purpose of studying these optical characteristics is to enable us to understand how to select the appropriate laser light source and camera lens according to the test conditions Our goal is to ensure that speckle patterns with proper size, moderate contrast, and regular motion can be observed in the imaging sensor by adjusting the optical characteristics, so that the vibration characteristics of the object can be accurately measured. Also, the motion characteristic of the speckle image in a strongly defocused camera is investigated by experiments. The experiment result proves that in a defocused camera, the speckle image can amplify the subtle object

vibration, convert it into obvious speckle motion. Besides, even the object vibration is complex, including motion of multiple modes, the speckle motion is only sensitive to the tilt motion of the object, which means the speckle motion is a simple linear motion. This is very crucial to sensing object vibration with speckle images.

CHAPTER 3 IMAGE PROCESSING IN SPECKLE SENSING METHOD

Image displacement calculation is another key point in speckle sensing method. In the last chapter it has been illustrated that object motion can be extracted by observed speckle's motion using defocused camera. Speckle image magnifies the micro motion of object, object's micro motion that is difficult to be directly observed can be clearly expressed using speckle images. However, if one wants to further obtain information about the object motion, the displacement of the speckle image must be calculated first. In this chapter the methods for displacement calculation of speckle image are introduced.

3.1 Speckle Image Shift Measurement in Speckle Photography

Before the invention of digital camera, speckle image has been used for measuring object motion and object deformation. According to the measurement principles, speckle measurement methods can be classified into two types: speckle interferometry and speckle photography. In 1970s, the electronic speckle interferometer proposed by Butters and Leendertz for the first time used a TV camera to record the speckle on the surface of the object and obtained the interference fringes using electronic circuit subtraction and filtering methods [50]. A classical optical setup of speckle interferometry is shown as Figure 3.1. The probe beam irradiates on the object, then the beam reflected from the object interferes with reference beam,

and the interference pattern can be observed on the screen. When the object vibrates, the optical path changes, resulting in phase variation. The object motion thus can be detected from the interference pattern variation. Although this method can be used for measuring object displacement, deformation, and surface roughness, the method is complex compared with speckle photography method.

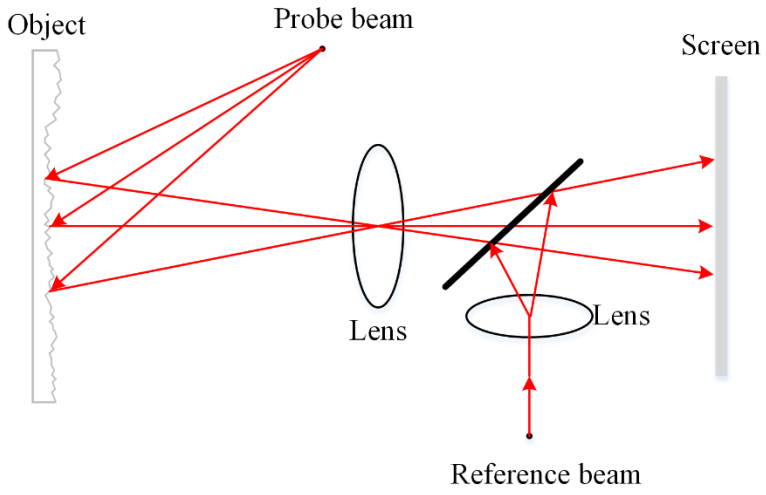

Figure 3.1 Schematic of electronic speckle interference method

On the other hand, the structure of speckle photography is shown as Figure 3.2. This method is very similar to modern digital speckle camera methods, except that film is used to capture speckle patterns, because digital sensors were not invented at that time. The method does not need a reference beam, only a laser beam is irradiated on the surface of the object to be measured, and two speckle patterns are respectively recorded before and after the object moves. This is the so-called "double exposure" measurement.

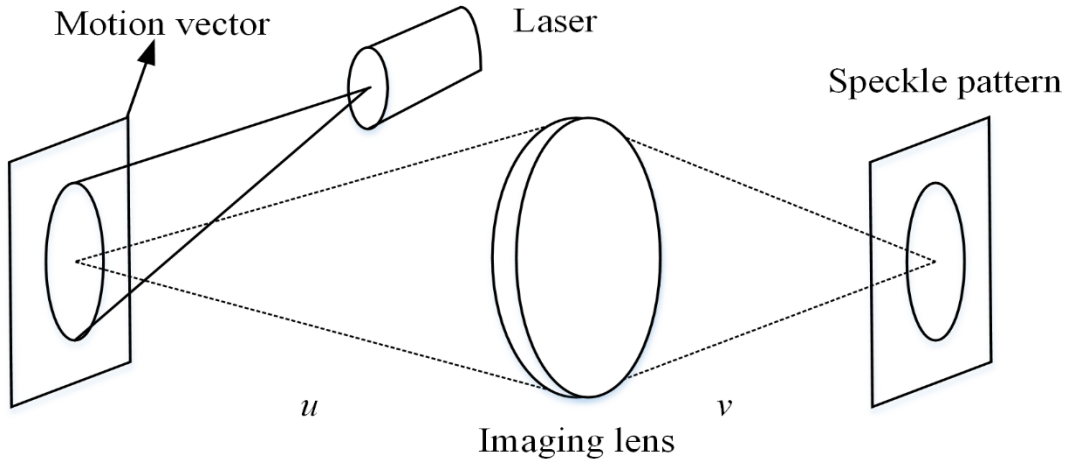

Figure 3.2 Basic optical setup for recording speckle pattern

In the optical path, the speckle image focused by the convex lens will pass through the Fourier lens, and the two exposures recorded on the film are the spatial frequency domain of the speckle image after Fourier transform. A shift occurs between the two pictures due to the movement of the object. This picture shift can be obtained from the analysis of the superimposed exposure stripes. The direction of the fringe is the direction of displacement, and the spatial period of the fringe is the amount of displacement.

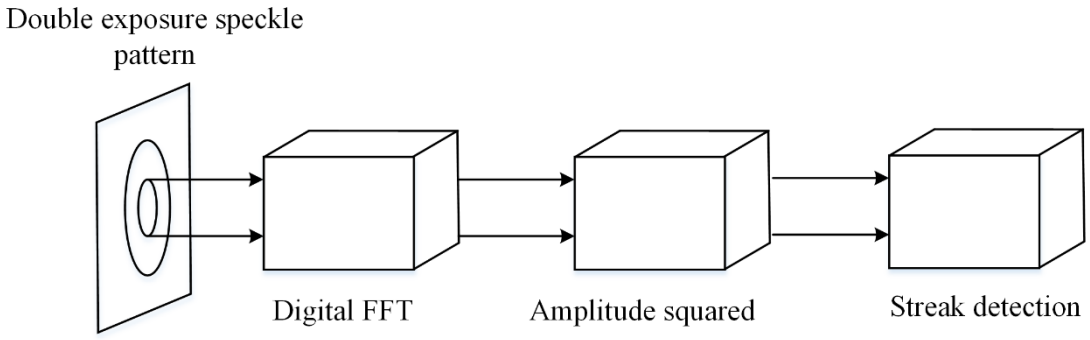

Figure 3.3 Fourier transform processing for double exposure speckle pattern

Although we now use digital imaging sensors to record pictures, this analysis method can still be verified. The Fourier change of the lens can be equivalently obtained by the Fourier

transform of the digital speckle images. The process of the analysis method of displacement in speckle photography is shown in Figure 3.3. Assuming the speckle image obtained by two exposures, and the displacement between them is (x, y). If two exposures occur on a film, the speckle image is shown in Figure 3.4.

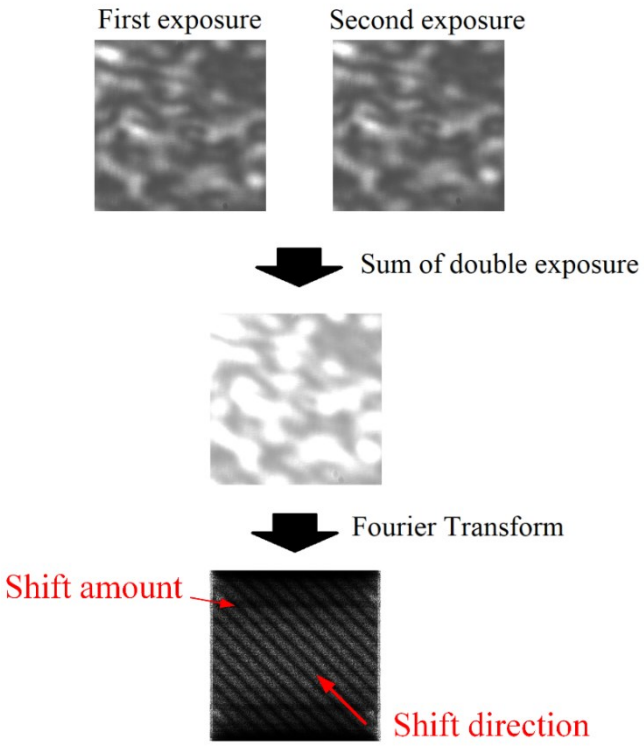

Figure 3.4 The double exposure speckle pattern and its Fourier transform pattern

Using double exposure and Fourier transform can determine the object displacement by analyzing the spatial period and the direction of fringe pattern. However, it is almost impossible to achieve real-time analyzation using such method. Fortunately, with the invention of digital cameras and the advancement of digital image processing algorithms, real-time displacement analysis became possible [51-56].

3.2 Image in Frequency Domain

3.2.1 Basic concept of Fourier transform

In the signal processing field, Fourier transform plays a basic and important role. Through Fourier transform and inverse Fourier transform, the time-frequency domain transform of the signal can be realized, which helps us analyze the characteristics of the signal from different views. Digital image can be regarded as a two-dimensional discrete digital signal. Fourier transform also plays a crucial role in analyzing spatial displacement of digital images [57].

Let us start from a brief review of some basic knowledges. For a one-dimensional continuous signal $f(t)$, its Fourier transform $F(\omega)$ is:

$$F(\omega) = \int f(t) \exp(-i\omega t) \qquad (3-1)$$

Through the Fourier transform, one can obtain the frequency domain information of the signal. Figure 3.5(a) illustrate signal composed of $100Hz$ and $200Hz$ sine components in time domain, and Figure 3.5(b) is the Fourier transform of the signal. The frequency domain result clearly presents the frequency information of all the components contained in the signal.

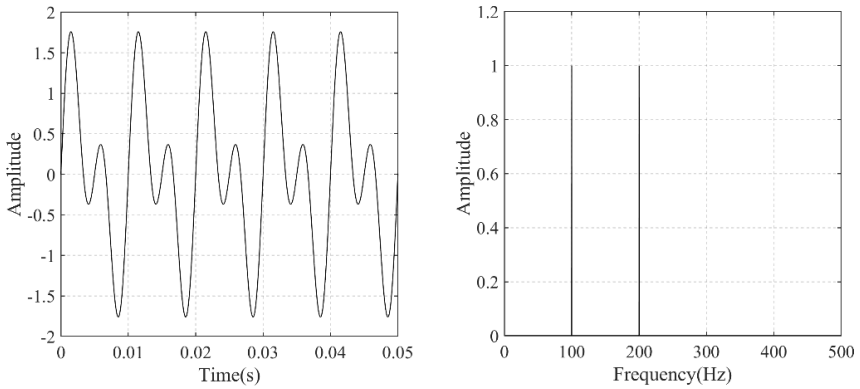

Figure 3.5 The signal waveform composed of two single-frequency sinusoidal signals is in (a) the time domain and its (b) Fourier transform

Now if the signal is two-dimensional continuous signal, the Fourier transform becomes:

$$F(u,v) = \iint f(x,y)\exp[-i(ux+vy)] \qquad (3-2)$$

Digital image is two-dimensional discrete signal. Assuming the size of the image is $M \times N$ pixels, then the discrete Fourier transform of an image is:

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right] \qquad (3-3)$$

where $f(x,y)$ is the image intensity at pixel (x,y). Figure 3.6 illustrates a speckle image and its result of Fourier transform. In time domain, the meaning for coordinate (x,y) of two-dimensional image is easy to understand. It indicates the spatial location of a pixel. In frequency domain, the coordinate (u,v) represents the frequency information. Figure 3.6(b) shows a shifted Fourier transform result. The coordinates near center area represent low-frequency component while coordinates near edge area represent the high-frequency component. As a comparison, here the "Lenna image" and its Fourier transform result are shown in Figure 3.7. It can be seen from the Fourier transform result that the spectrum of speckle pattern is more dispersed, locating in both the center area (low frequency) and edge area (high frequency). On the other hand, the frequency spectrum of a normal picture (Lenna image) is more concentrated in the center area (low frequency). The frequency of the image can be understood as the speed of gray value variation. Compared with normal images, the gray value of speckle images changes more frequently and drastically, which means that they have more high-frequency components in the frequency domain.

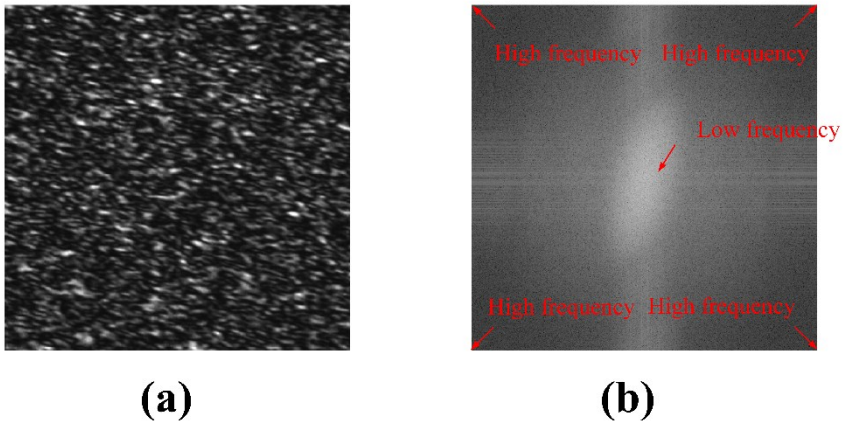

Figure 3.6 Captured speckle pattern and its Fourier transform result. (a) Speckle pattern in time domain (b) Speckle pattern in frequency domain

Figure 3.7 The Lenna image and its Fourier transform result. (a) Image in time domain (b) Image in frequency domain

The process of Fourier transform is reversible. The inverse Fourier transform can be expressed as:

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right] \qquad (3-4)$$

Again, M and N is the width and height of the image. Using inverse Fourier transform the

image can be restored in time domain, as shown in Figure 3.8.

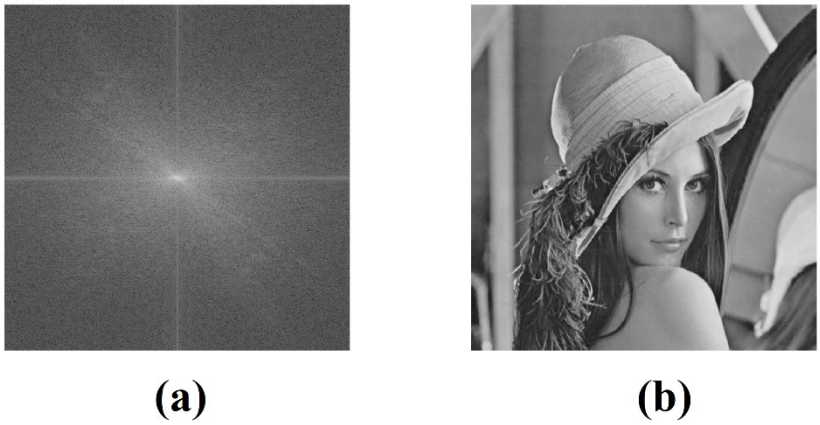

Figure 3.8 Inverse Fourier transform and the recovered image in time domain. (a) Spectrum of the image (b) Recovered image in time domain

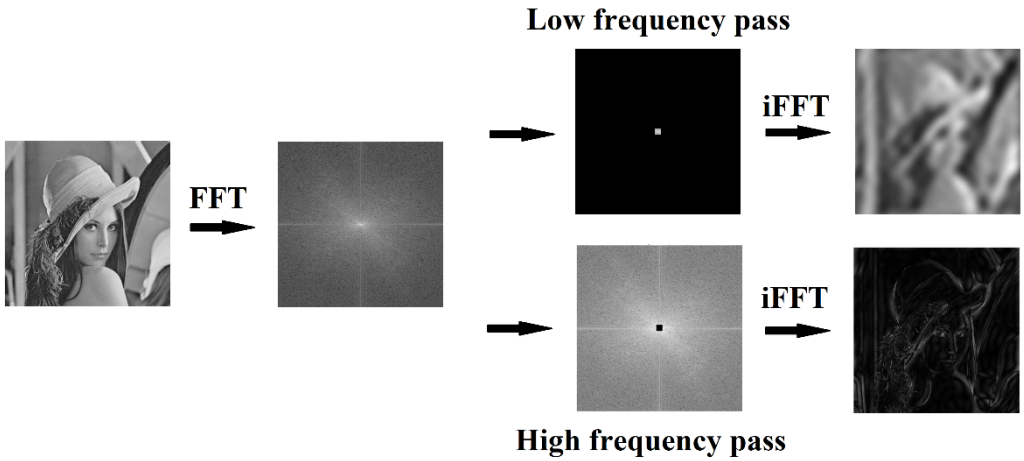

Figure 3.9 Image filtering in the frequency domain

In the process of time -frequency -time domain transform, one can easily process the images in frequency domain. For example, by removing the high-frequency or low-frequency part of the image in frequency domain after Fourier transform, one can realize the filtering of the time-domain image. The image could be blurred by removing the high-frequency part in frequency domain, or the image outline can be extracted by removing the low-frequency part in frequency

domain, as shown in Figure 3.9.

3.2.2 Image translation in the frequency domain

Through the operation in frequency domain, people can realize the image processing of time domain images. Conversely, if image change happens in the time domain, the corresponding changes will also be reflected in the frequency domain. Figure 3.10 shows that the image is rotated in time domain, and its spectrum also shows rotation in frequency domain.

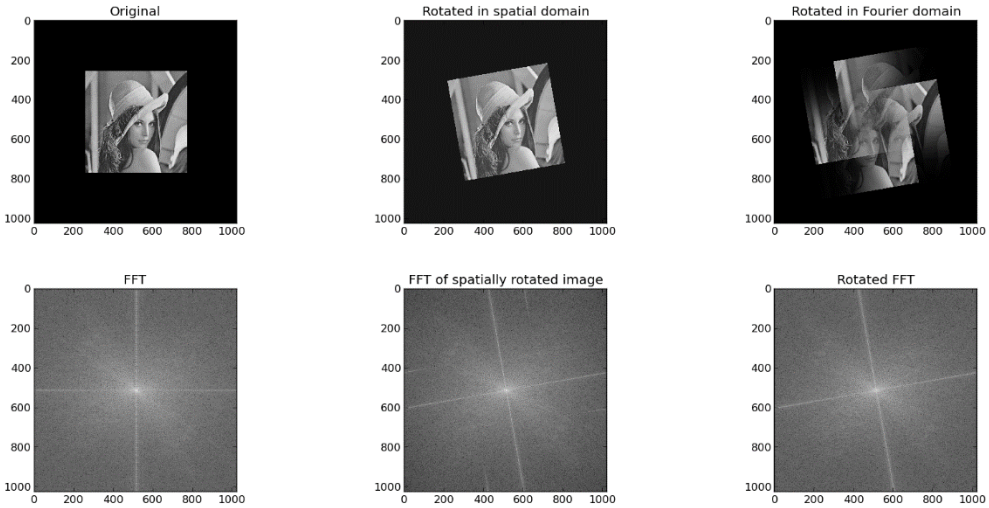

Figure 3.10 Rotated images in time domain and the corresponding changes in frequency domain

For speckle measurement methods, we are more concerned about the changes in the time and frequency domain when the image is translated. According to the two-dimensional discrete Fourier transform formula, assuming the shift amount of the image is (dx, dy), then the Fourier transform result $F_1(u, v)$ after image translation is:

$$F_1(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x+dx, y+dy)\exp\left[-i2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right] \qquad (3-5)$$

The relationship between spectrums before and after image translation is:

$$F_1(u,v) = F(u,v)\exp\left[-i2\pi\left(\frac{udx}{M}+\frac{vdy}{N}\right)\right] \qquad (3-6)$$

In Equation (3-6) there are only two unknown parameters, which are the displacement components dx and dy. Because the spectrum $F_1(u,v)$ and $F(u,v)$ before and after the displacement are known, we can obtain the values of the displacement components dx and dy by simple calculation. The displacement of the image can be calculated using the image's Fourier transform. Actually, using fringes to calculate image displacement in the speckle photography described earlier is based on this principle.

On the other hand, by operating the Fourier transform result in the frequency domain, the image displacement can also be achieved reversely. The mathematics for shifting image in frequency domain is straightforward. The image $f(x,y)$ is transferred into frequency domain $F(u,v)$. Generally, the image can be re-transformed to the time domain using Equation (3-4), but here before doing inverse Fourier transform a linear phase shift is applied in the complex plane to obtain a shifted image in time domain. The amount of phase shift equals to the desired image displacement. After that, the inverse Fourier transform is carried out to obtain the shifted image $f_1(x,y)$ in time domain. Assuming the target shift amount is $(\Delta x, \Delta y)$, the expression for $f_1(x,y)$ is:

$$f_1(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[i2\pi\left(\frac{ux}{M}\Delta x+\frac{vy}{N}\Delta y\right)\right] \qquad (3-7)$$

In Figure 3.11 the result of image shifted by (5, 5) pixels using Fourier transform is illustrated. Using this method can not only achieve integer pixel displacement, but also achieve high-precision sub-pixel displacement of the image. It is of great significance to manually extract precise shifting of image. It can provide a ruler for the follow-up research on the accuracy of

various displacement calculation algorithms.

(a) (b)

Figure 3.11 Shift image in frequency domain. (a) Original image (b) Shifted image

3.3 Methods for Image Shift Calculation

In speckle sensing method, one crucial step is displacement calculation for the captured speckle image sequence. In the past, digital image correlation (DIC) is widely used in speckle displacement calculation. Recently, using some other image processing algorithms for speckle image displacement calculation has been reported. For example, feature based algorithms and optical flow algorithms. In this part several representative image displacement analysis algorithms are introduced.

3.3.1 Digital image correlation

The Fourier-based DIC method is a powerful image registration technique. It is an area-based algorithm and can achieve sub-pixel level accuracy of displacement calculation. The earliest use of the fast Fourier transform (FFT) form image registration can be found in 1970s [58]. After that, a huge number of DIC based image registration has been proposed, and these works enhanced the accuracy, robustness, and efficiency of this technique [59-63].

Basically, there are two steps for DIC image registration. The first step is calculating the pixel-level image displacement. Pixel-level means the displacement is first determined with integer pixel accuracy. This step can be achieved though the FFT of the image. According to the correlation theorem, the Fourier transform of the correlation of two image is the product of the Fourier transform of one image and the complex conjugate of the Fourier transform of the other image. This theorem indicates the process for calculating integer displacement of two images with DIC. Assuming the camera captures two images $f(x,y)$ and $g(x,y)$ and the displacement between the two image is $(\Delta x, \Delta y)$. i.e., $g(x,y) = f(x-\Delta x, y-\Delta y)$. The correlation function of the two image is given by:

$$CC = F^{-1}\big(F(u,v) * G(u,v)\big) \qquad (3-8)$$

where $F(u,v)$ and $G(u,v)$ is the FFT result of $f(x,y)$ and $g(x,y)$, $*$ stands for complex conjugate, and F^{-1} denotes the inverse Fourier transform. Figure 3.12 shows two speckle images and their cross-correlation result. The true translation amount in x and y directions between two images are both $20\ pixels$. The shift amount can be determined by finding the peak in the correlation result. The discrete coordinates correspond to the integer pixel shifts between two images.

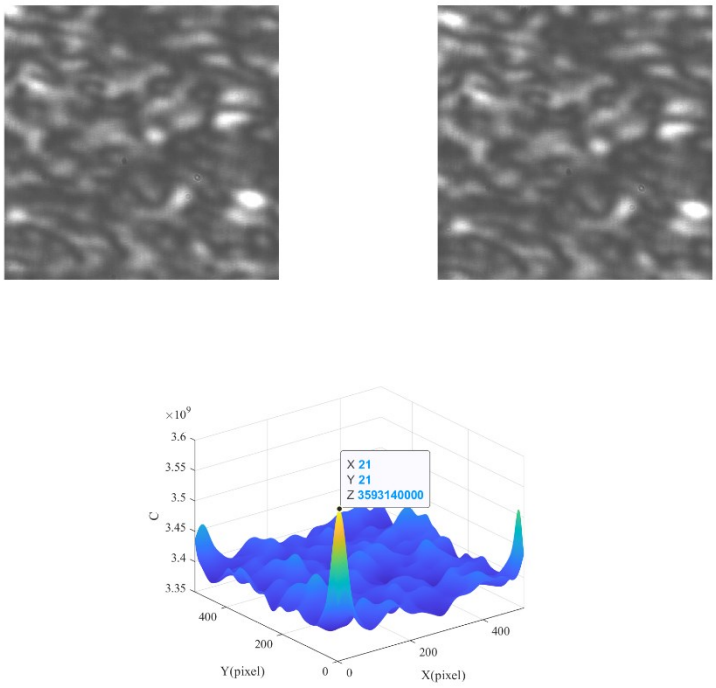

Figure 3.12 Two speckle patterns and their cross-correlation result

It is obvious that the method of finding the discrete coordinates of the peak can only determine the image displacement with integer pixel because the pixel coordinates of the image are all integer values. But in practical applications, we may need to calculate the image displacement with sub-pixel level accuracy, which means the image shift may be less than 1 pixel. Generally, the sub-pixel Fourier-based image correlation methods can be classified into two categories, which can be implemented either in the spatial domain by means of the sharp correlation peak or in the frequency domain by means of the phase difference. The sub-pixel analyzation methods in spatial domain include peak centroid, similarity interpolation, upsampling, and nonlinear optimization, whereas the sub-pixel calculation methods in frequency domain include plane fitting, line fitting, sawtooth signal estimation, frequency estimation, and optimization. As a representative here the upsampling method is introduced, which is an

accuracy and efficient method for determining sub-pixel image shift.

Here after obtaining the correlation result, the upsampling of a small area near the correlation peak is conducted. The upsampling result is shown in Figure 3.13. Assuming the upsampling factor is κ, usually a 1.5×1.5 pixel neighborhood (in units of the original pixels) is upsampled to $1.5\kappa \times 1.5\kappa$ pixels. This is more efficiency compared with upsampling whole image. Then the sub-pixel shift amount is determined by the peak coordinates of the upsampling area. For instance, if the upsampling factor $\kappa = 100$, it means the method can determine shift with a resolution of 0.01 pixels.

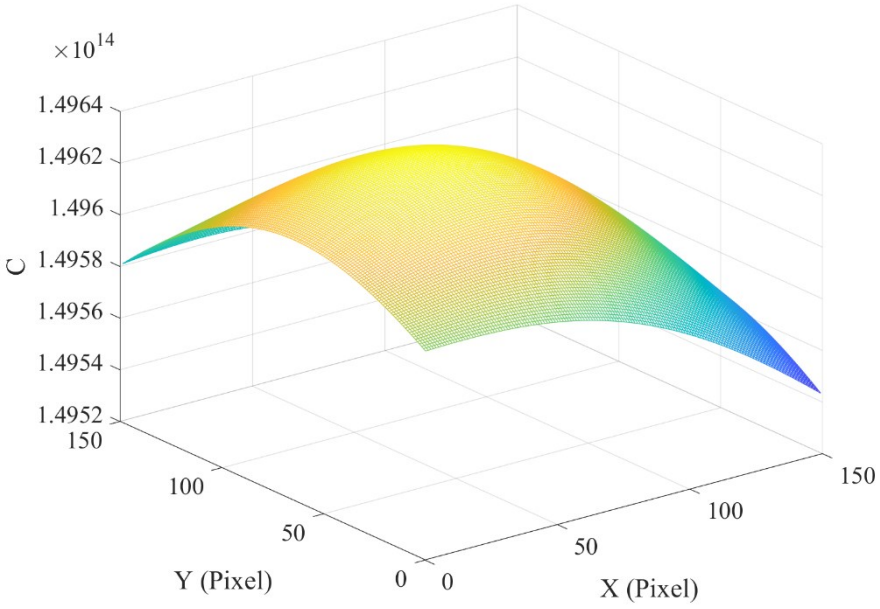

Figure 3.13 Upsampling result on the 1.5×1.5 pixels neighbor area of the cross-correlation peak, the magnification factor is 100

3.3.2 Feature matching

Feature based methods can also be used to determine image translation. Generally, the

feature-based methods are performed in three steps: feature detection, feature description and feature matching. Image feature includes corner, blob, region, edge, and so on. Usually, the most well-used feature is point of image. There are many feature-based algorithms, such as FAST, SIFT, SURF, ORB, and so on [64-67]. Here the ORB (Oriented FAST and Rotated BRIEF) algorithm is selected as a representative to explain the image translation calculation process based on feature points.

The first step of feature matching is detecting feature points in two images. In this step the algorithm scans all pixels in the images, and pixels satisfying the algorithm's conditions are selected as feature points. Different algorithms have different method to determine feature points. With ORB algorithm, given a pixel p and radius r. On the circle around p, if there are more than N consecutive pixels whose gray value is brighter or darker than the gray value of p, then p will be selected as a feature point, as shown in Figure 3.14. Image pyramid is also conducted to make sure that the feature points are scale-invariant.

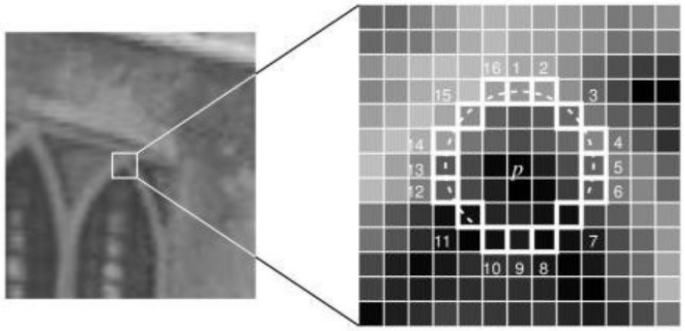

Figure 3.14 Method for determine ORB feature point based on pixel intensity

In the next step, the descriptor is generated for each feature point. The descriptor can be regarded as an ID of the point, and with this the points in different images can be matched

together. In ORB algorithm, each feature point is given a binary feature descriptor consisting of 0 and 1. In this way, the points in the two images with similar detectors can be matched. By comparing the coordinates (x, y) and (x', y') in original image and translated image, one can determine the shift amount of two images, as shown in Figure 3.15.

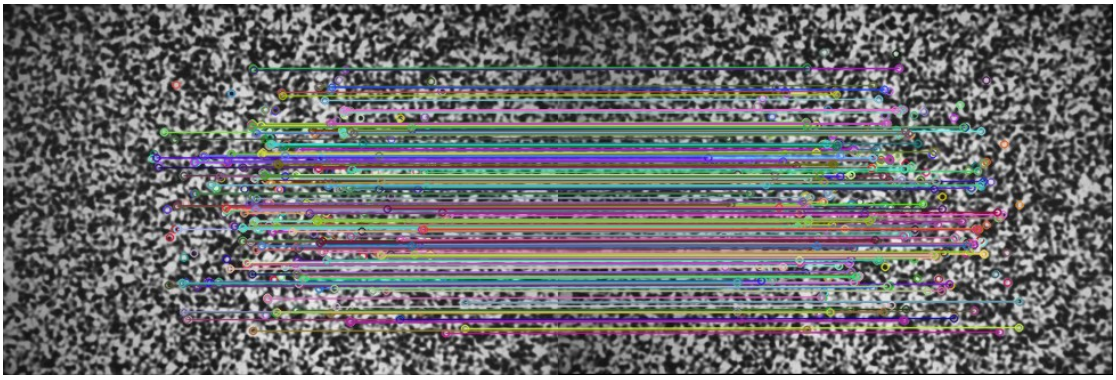

Figure 3.15 Matching of ORB feature points in two images

3.3.3 Optical flow method

3.3.3.1 General principles of optical flow method

The optical flow method uses the variations in the time domain of the pixels in the image sequence and the relationship between adjacent frames to find the correspondence between the previous frame and the current frame and then calculate the movement of objects between adjacent frames [68]. Generally, the instantaneous variation of gray value at a specific coordinate point on a two-dimensional image plane is defined as an optical flow vector. In general, the "optical flow" is the instantaneous velocity, which is equivalent to the displacement of the target point when the time interval is small (for example, between two consecutive frames in the video). In space, motion can be described by a motion field, and on an image plane, the motion of an object is often reflected by the different gray distributions of different images in the image sequence. Therefore, the motion field in space transferred to the image is

expressed as optical flow field. The optical flow field is a two-dimensional vector field, which reflects the changing trend of the gray level of each point on the image and can be regarded as the instantaneous velocity field generated by the motion of pixels with gray levels on the image plane. The information it contains is the instantaneous velocity vector information of each image point.

The purpose of studying the optical flow field is to approximate the motion field that cannot be directly obtained from the sequence of images. There are two basic assumptions for optical flow method. (1) The brightness is constant. That is, when the same target moves between different frames, its brightness will not change. This is the assumption of the basic optical flow method (all variants of the optical flow method must be satisfied), which is used to obtain the basic equation of the optical flow method. (2) Time continuity or exercise is "small exercise". That is, changes in time will not cause drastic changes in the target position, and the displacement between adjacent frames should be relatively small. It is also an indispensable assumption of the optical flow method.

Consider the light intensity of a pixel $I(x, y, t)$ in the first frame (where t represents its time dimension). It moved the distance of (dx, dy) to the next frame, which took dt time. Because it is the same pixel, according to the first assumption mentioned above, we believe that the light intensity of the pixel before and after the movement is constant, namely:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \qquad (3-9)$$

Carrying out Taylor expansion on the right end of Equation (3-9), we get:

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + \varepsilon \qquad (3-10)$$

where ε represents the second-order infinitesimal term, which can be ignored. According to

Equation (3-9) and Equation (3-10), we can get:

$$\frac{\partial I}{\partial x}\frac{dx}{dt}+\frac{\partial I}{\partial y}\frac{dy}{dt}+\frac{\partial I}{\partial t}\frac{dt}{dt}=0 \qquad (3-11)$$

Suppose u and v are the velocity vectors along the X axis and Y axis respectively, and we get:

$$u=\frac{dx}{dt} \qquad (3-12)$$

$$v=\frac{dy}{dt} \qquad (3-13)$$

Let $I_x=\frac{\partial I}{\partial x}$, $I_y=\frac{\partial I}{\partial y}$ and $I_t=\frac{\partial I}{\partial t}$, which represent the partial derivatives of the gray value of the pixels in the image along the X, Y and T directions. Thus, Equation (3-11) can be written as:

$$I_x u+I_y v+I_t=0 \qquad (3-14)$$

Among all the parameters, I_x, I_y and I_t can be obtained from the image data, and (u,v) is the optical flow vector. There is only one constraint equation, and there are two unknowns in the equation. In this case, the exact values of u and v cannot be obtained. Under this situation, additional constraints need to be introduced, and constraints are introduced with different considerations, resulting in different optical flow field calculation methods. According to the difference between theoretical basis and mathematical methods, the optical flow methods can be classified into five types: gradient-based (differential) method, matching-based method, energy (frequency)-based method, phase-based method, and neurodynamic method. Besides, optical flow methods can also be divided into dense optical flow and sparse optical flow according to the degree of density of the two-dimensional vector in the formed optical flow field. Dense optical flow is an image registration method for point-by-point

matching of an image or a specified area. It calculates the offset of all points on the image to form a dense optical flow field. Through this dense optical flow field, pixel-level image registration can be performed, as shown in Figure 3.16. Since the optical flow vector is dense, the registration effect is obviously better than that of sparse optical flow registration. But its side effects are also obvious. Because the offset of each point needs to be calculated, the calculation amount is also significantly larger, and the timeliness is poor.

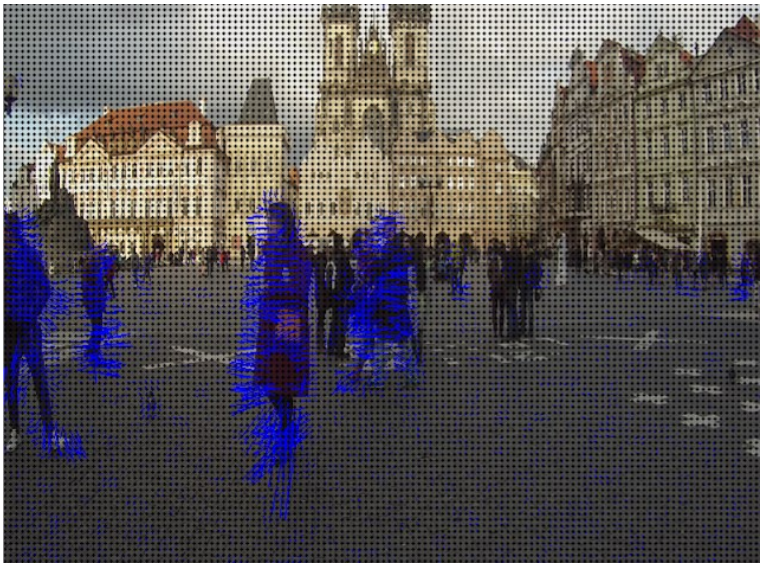

Figure 3.16 Dense optical flow

Contrary to dense optical flow, sparse optical flow does not perform point-by-point calculation for each pixel of the image. It usually needs to specify a set of points for tracking, as shown in Figure 3.17. This set of points should have some obvious characteristics, such as Harris corners, etc., then the tracking will be relatively stable and reliable. The computational overhead of sparse tracking is much smaller than that of dense tracking.

Figure 3.17 Sparse optical flow

The advantage of the optical flow method is that it can accurately detect and recognize the position of the moving day mark without knowing the information of the scene, and it is still applicable when the camera is in motion. Moreover, optical flow not only carries the movement information of the moving object, but also carries rich information about the three-dimensional structure of the scene. It can detect the moving object without knowing any information about the scene. The applicable conditions of the optical flow method, that is, the two basic assumptions, are not easy to meet in reality.

Hypothesis 1: The brightness is constant.

But the actual situation is that the optical flow field does not necessarily reflect the actual movement of the target, as shown in Figure 3.18. In the picture, the light source does not move, and the surface of the object is uniform, and autobiographical motion is generated, but no light flow is generated. In the picture, the object does not move, but the light source and the object move relative to each other, but there is light flow. Therefore, it can be said that the optical flow method is sensitive to light, and the change of light easily affects the recognition effect.

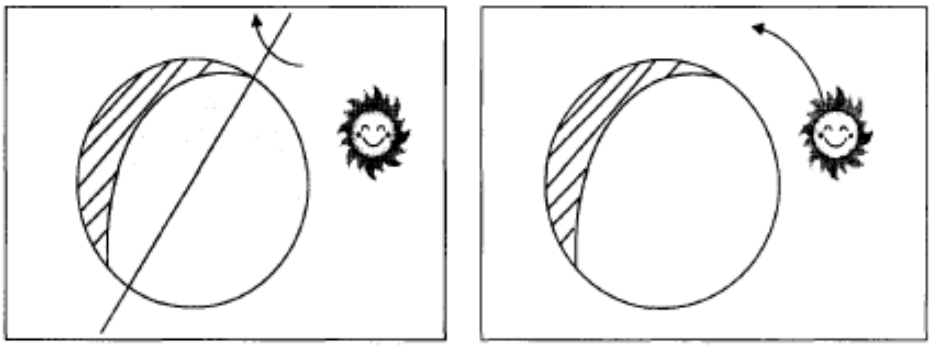

Figure 3.18 Intensity constant problem for optical flow algorithm

Hypothesis 2: Small exercise.

As mentioned in the previous part, in reality, large-distance movement is also common. Therefore, when the speed of the target to be detected is too fast, the traditional optical flow method is not applicable.

3.3.3.2 Introduction to Farneback optical flow algorithm

Farneback optical flow method is a dense optical flow algorithm that calculates image displacement based on image intensity gradient variations, which was proposed by Farneback in 2003 [69, 70]. The basic concept of calculating displacement is as follows.

In the Farneback optical flow algorithm, a two-dimensional gray image is regarded as a two-dimensional quadratic function $f(x, y)$, where the independent variable is the pixel position (x, y) of the image, and the dependent variable is the image intensity at that point. For the pixel P of interest, take it as the center and select a neighborhood with a radius of n pixels to build local coordinate. According to the coordinates and gray values of all points in the $(2n + 1) \times (2n + 1)$ area, the coefficients at the center point P can be fitted to obtain the local image model:

$$f_1(\mathbf{x}) = \mathbf{x^T A_1 x} + \mathbf{b_1^T x} + \mathbf{c_1} \tag{3-15}$$

where $\mathbf{x} = (x, y)$, which indicates the coordinate of center pixel P. $\mathbf{A_1}$, $\mathbf{b_1}$, and $\mathbf{c_1}$ are coefficients matrix, their dimensions and the elements are given by:

$$\mathbf{A_1} = \begin{pmatrix} r_4 & \frac{r_6}{2} \\ \frac{r_6}{2} & r_5 \end{pmatrix} \tag{3-16}$$

$$\mathbf{b_1} = \begin{pmatrix} r_2 \\ r_3 \end{pmatrix} \tag{3-17}$$

$$\mathbf{c_1} = r_1 \tag{3-18}$$

The fitting based on the gray value in the neighborhood ensures each point in the image have its own coefficients $r_1 \sim r_6$. Then, assuming the image is then translated to a new image f_2, and the translation vector is $\mathbf{d} = (\Delta x, \Delta y)$. Then the relationship between f_2 and f_1 can be expressed as:

$$\begin{aligned} f_2(\mathbf{x}) &= f_1(\mathbf{x} - \mathbf{d}) \\ &= (\mathbf{x} - \mathbf{d})^\mathbf{T} \mathbf{A_1} (\mathbf{x} - \mathbf{d}) + \mathbf{b_1^T} (\mathbf{x} - \mathbf{d}) + \mathbf{c_1} \\ &= \mathbf{x^T A_1 x} + (\mathbf{b_1} - \mathbf{2A_1 d})^\mathbf{T} \mathbf{x} + \mathbf{d^T A_1 d} - \mathbf{b_1^T d} + \mathbf{c_1} \\ &= \mathbf{x^T A_2 x} + \mathbf{b_2^T x} + \mathbf{c_2} \end{aligned} \tag{3-19}$$

Since the corresponding coefficients of each term in the quadratic function are equal, we can get:

$$\mathbf{A_1} = \mathbf{A_2} \tag{3-20}$$

$$\mathbf{b_2} = \mathbf{b_1} - \mathbf{2A_1 d} \tag{3-21}$$

$$\mathbf{c_2} = \mathbf{d^T A_1 d} - \mathbf{b_1^T d} + \mathbf{c_1} \tag{3-22}$$

According to Equation (3-21), the translation vector $\mathbf{d}$ can be determined, which is:

$$\mathbf{d} = -\frac{1}{2}\mathbf{A_1^{-1}}(\mathbf{b_2} - \mathbf{b_1}) \qquad (3-23)$$

In this way one can calculate the displacement vector for each point in the image, theoretically. However, such a simple calculation cannot provide precise result, and also the calculation is time-consuming. Therefore, the algorithm is optimized, as follows:

First of all, a weighting algorithm is used when performing polynomial expansion at point P, which artificially makes the gray value of the points close to the center take a higher weight and reduces the weight of the gray value of the edge points, so as to obtain more precise coefficients $r_1 \sim r_6$. Meanwhile, space conversion and dual conversion are used when solving the coefficients to improve the efficiency of calculation

Second, according to Equation (3-20) the coefficients of the quadratic term are supposed to be same but in actual situation which is not. Here the average value of the quadratic coefficients is adopted:

$$\mathbf{A} = \frac{\mathbf{A_1} + \mathbf{A_2}}{2} \qquad (3-24)$$

and

$$\Delta\mathbf{b} = -\frac{1}{2}(\mathbf{b_2} - \mathbf{b_1}) \qquad (3-25)$$

In this way, Equation (3-23) could be expressed as:

$$\mathbf{A} \times \mathbf{d} = \Delta\mathbf{b} \qquad (3-26)$$

Third, although the displacement at each pixel can be determined using Equation (3-26), the fact is that pixel wise solving Equation (3-26) only performs very noisy result. Hence a

hypothesis is made here, which is the optical flow field changes smoothly. The hypothesis can be understood in this way. For two points with similar spatial positions in the image, it can be considered that their motion vectors $\mathbf{d}$ are also approximately equal. As shown in Figure 3.19, the pixels representing the mountain and the pixels representing the flowing water are far away in space, and their motion vectors are also very different. But for points that are relatively close in space, such as pixels that both represent flowing water, their motion vector difference is very small, and the motion variance in space is sooth.

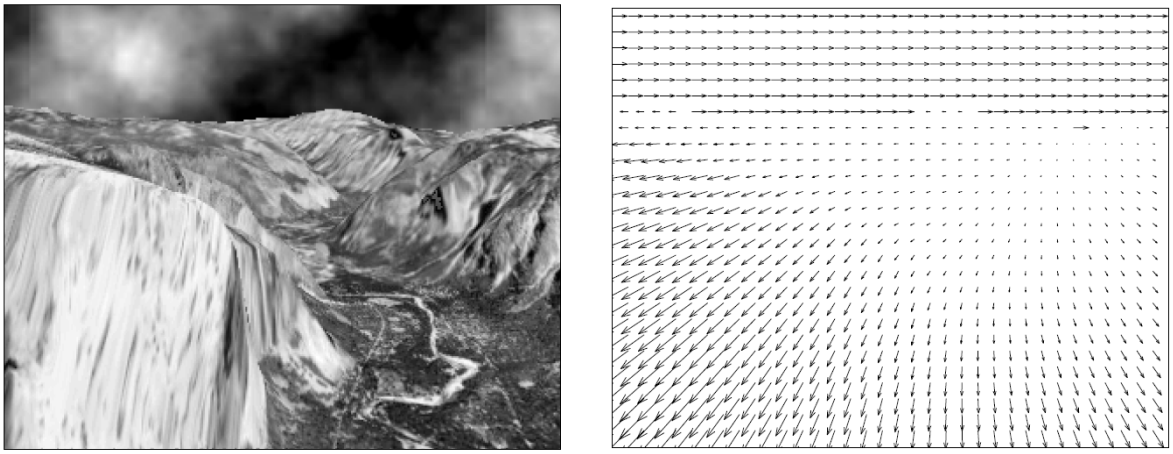

Figure 3.19 One frame of the Yosemite sequence and the corresponding true velocity

Therefore, when calculating the motion vector, people do not solve one equation for a pixel P, but use multiple sets of equations in the neighborhood I of the point to construct a weighted least squares equation:

$$e(\mathbf{x}) = \sum_{\Delta x \in I} \omega(\Delta x) \|\mathbf{A}(\mathbf{x} + \Delta x)\mathbf{d} - \Delta\mathbf{b}(\mathbf{x} + \Delta x)\|^2 \qquad (3-27)$$

The assumption that the optical flow field variation is smooth is very important. It allows us to use the least square method to eliminate the noise in the calculation, which is a guarantee for the high accuracy of the optical flow method. In fact, we can make a further assumption for speckle images, that is, the optical flow field of the speckle image is the same at all points,

which can be derived based the theoretical analysis in the previous chapter. When the area illuminated by the light source is very small, the area can be regarded as a rigid body during the movement, that is, only the global movement occurs to the area without deformation. In this case, all pixels in the speckle image will show the same motion characteristics. When using the optical flow method to calculate the speckle displacement, the neighborhood size can be set to the size of the entire speckle image in the step of solving the displacement, so as to obtain an optical flow field with the same displacement everywhere.

Let us go back to the optical flow algorithm. There is a problem in the previous analysis: if one wants to solve the optical flow at a point $\mathbf{x}$ of the image, then it is necessary to know the polynomial expansion coefficients at this point and the coefficients at the corresponding points $(\mathbf{x}+\mathbf{d})$ in the translated image. However, $\mathbf{d}$ is exactly the unknown quantity that need to be solved. In other words, we cannot know the corresponding point of a certain point in the translated image.

To solve this problem, optical flow algorithm proposed two solutions. First, similar to the smooth variation of the optical flow field within an image, for a continuous image sequence, the optical flow algorithm assumes that the optical flow variation between images is also smooth. This is easy to understand. For example, for high-speed camera, the time interval between adjacent frames is generally very small, so it can be considered that the image motion does not suddenly change drastically. Under this assumption, we can use the motion vector $\mathbf{d}$ obtained in the previous calculation as the prior input of the current calculation to obtain the correspondence between the pixels in the two images. After obtaining the current motion vector, the calculated result can be iterated for several times to get an accurate result of optical flow field.

It is clearly that in some cases we cannot get an exact prior input, for example, when calculating the displacement of the first two frames. In these cases, the optical flow algorithm assumes that the prior input is zero. Meanwhile, the optical flow algorithm uses the image pyramid method, as shown in Figure 3.20. For the two input images, several downsamplings are carried out to reduce the image size. At first prior input is used to calculate the displacement at the top level with smallest image size, and then the output from top level is passed as the prior input to the next level of the pyramid and repeat this process. Finally, the displacement output with full image size is obtained at the bottom pyramid layer. With image pyramid, even if the displacement of the images is large, after several downsamplings the displacement at the top level will become small and close to our prior input, which is zero. This enhances the ability of optical flow to calculate large displacements to a certain extent.

So far, the main calculation idea of Farneback optical flow method has been explained. The calculation process of the algorithm is shown in Figure 3.21, which shows the optical flow algorithm with a three-layer image pyramid. In this figure, f_1 and f_2 represent two input images. LP stands for low pass filter, which means that the image is Gaussian blurred to eliminate the high-frequency noise at each layer. The mark "↓" means downsampling of the image, and the mark "↑" means upsampling the optical flow field of the calculation result with the same factor. At each image pyramid level, PE stands for polynomial expansion, which means estimating the polynomial coefficients for each pixel using the local image model. DE stands for displacement estimation, which means estimating the motion vector based on the polynomial estimation result. The processing of PE and DE are explained before. In addition, although it is not shown in the figure, multiple iterations will be performed on each layer of the image pyramid.

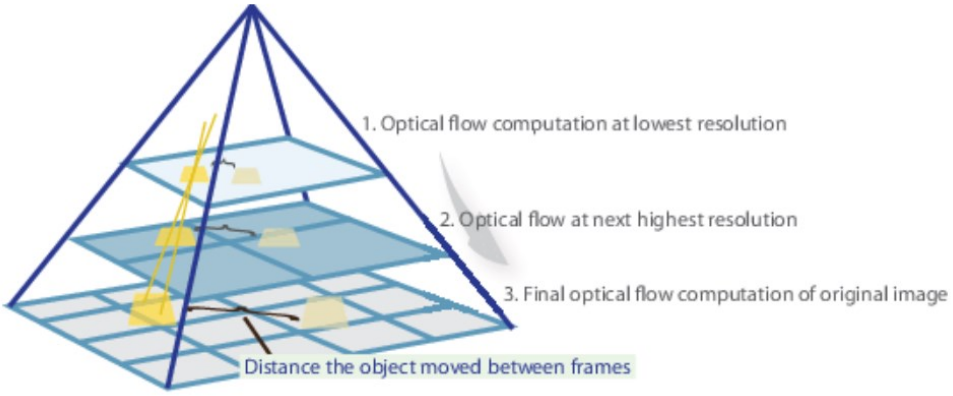

Figure 3.20 Image pyramid in optical flow algorithm

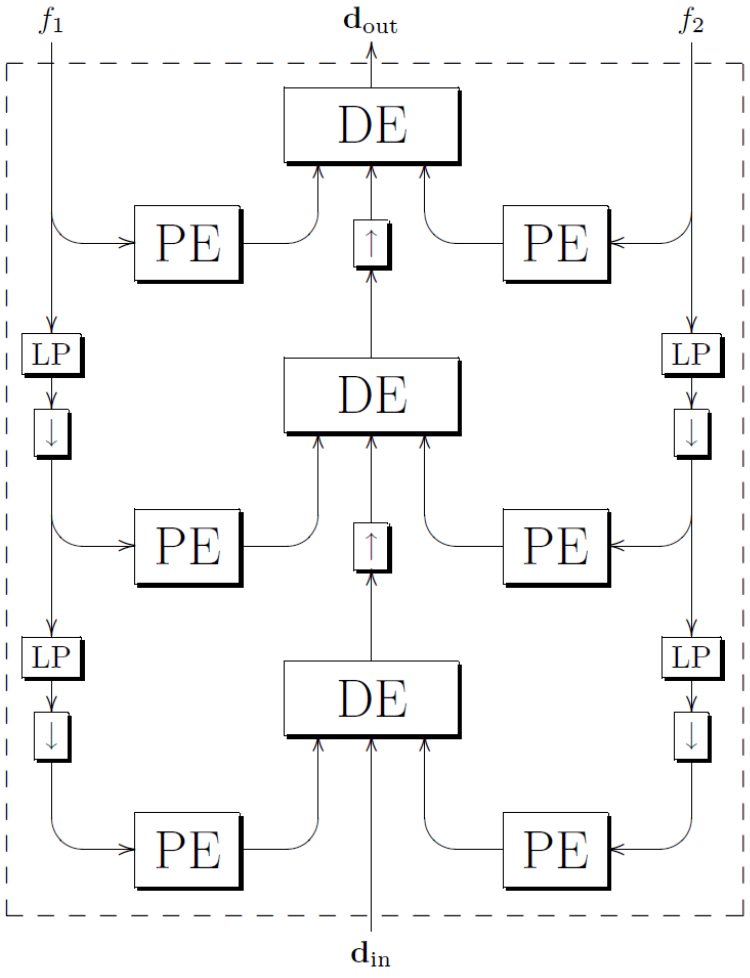

Figure 3.21 Processing flow chart of Farneback optical flow algorithm

3.4 Comparison of Different Methods

3.4.1 Requirements for image processing in real-time sensing method

Calculating image displacement has a number of applications in the computer vison fields, such as visual navigation and visual detection. In different applications, the requirements for the accuracy and speed of image displacement analysis are also different. With respect to the audio vibration measurement method using laser speckle images, our design goal is to restore the audio signal in real time while measuring vibration. The audio signal comes from the speckle displacement obtained by the image processing analysis. Therefore, if one wants to restore the audio signal in real time, the most basic requirement for the image processing algorithm is high processing speed. Specifically, the sound ranges that human can hear is 20~20000Hz, and the typical fundamental frequency of voice speech is 85~180Hz for adult male and 165~255Hz for adult female. Certainly, the human voice signal is not mono frequency but contains multiple frequencies. Thus, the sampling speed is much faster than several hundred hertz. For reference, in telephone the available voice frequency band ranges from approximately 300 to 3400 Hz. For this reason, the ultra-low frequency band of the electromagnetic spectrum between 300-3000 Hz is also called voice frequency, which is electromagnetic energy representing baseband sound energy. According to the Nyquist-Shannon sampling theorem, the sampling frequency (8 kHz) must be at least twice the highest component of the speech frequency before it can be sampled in discrete time (4 kHz) in order to effectively reconstruct the speech signal. For higher quality voice signal, usually a higher sampling frequency is needed. For example, the sampling frequency in music CD is usually 44100Hz.

Nevertheless, for an audio vibration measurement method based on speckle images, the first

challenge is that the image processing speed must reach at least 8kHz in order to reproduce the audio signal in real time. This image processing speed is much faster than the general real-time visual detection method. In order to achieve this target, the most direct and effective approach is reducing the speckle image size only use a small part of the speckle image for image displacement analysis. Certainly, the reduction of the image size means that less image information can be used, which is the second challenge to the image displacement analysis. At last, different from the normal visual tracking and image calculation under large-scale motion, the displacement of the vibrating object is usually very small, so the displacement scale of the image is also very small even the image is captured by defocused camera. In addition, due to the ultra-high-speed sampling frame rate of the camera, the displacement between adjacent images will be very small, and the change between adjacent images is almost indistinguishable. Therefore, the last challenge for image processing is the calculation accuracy.

In the following part, we will discuss the characteristics of different image displacement calculation algorithms. Through experiments, we can obtain the difference in calculation speed and accuracy of different algorithms to find the image processing algorithm suitable for the speckle vibration measurement method.

3.4.2 Feature-based method versus area-based method

Although both DIC, feature points and optical flow methods can analyze the displacement of images, these three methods also have their own unique features. Among them, the feature point method is quite different from the other two methods. The feature point method uses the local information in the image to realize the image displacement analysis based on the matching of the invariant features of the image in the changing process. In contrast, the DIC method, and the dense optical flow method both use the information of a local or a whole area in the image

to obtain the image displacement, which is a region-based analysis algorithm.

The feature points extracted by the feature point method usually have scale or rotation invariance, as shown in Figure 3.22. This ensures that the feature points extracted when the image undergoes large-scale translation, rotation or scale transformation can still be tracked. Because of this ability, the feature point method is suitable for low-speed (referring to the slow camera frame rate) and large-scale motion scenes. The slow frame rate of the camera means that the interval between two adjacent images is longer. In this case, the possibility of a sudden large change in the object or the camera's motion state during tracking is also greater. For example, in visual slam, the feature point method is widely used to calculate the movement posture of the camera itself.

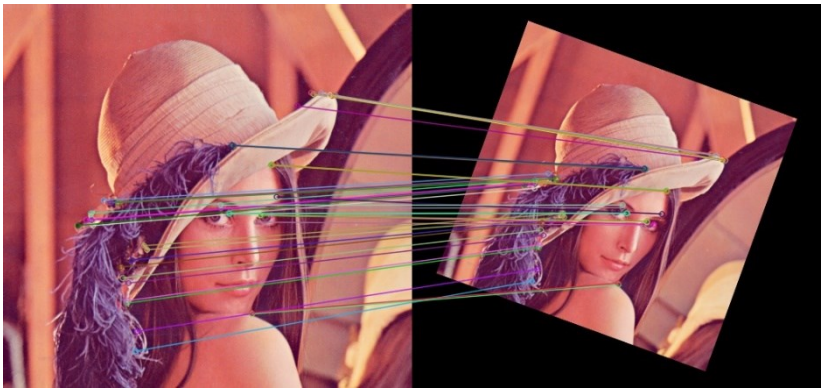

Figure 3.22 Rotation invariance and scale invariance of feature point matching method

However, if people want to use the feature point method to determine the displacement of the picture, generally speaking, the following procedures must be done: scanning the two images pixel by pixel to find the feature points → calculate the descriptors for each feature point → matching the feature points which the descriptor distance is within the threshold → determining the image shift according to the image coordinates of the matched points. At the same time, mismatching often occurs in the process of matching feature points, so special algorithms must be used to filter out mismatched points. Such cumbersome steps determine

that the speed of the feature point algorithm is very low, and it is basically unable to meet the needs of high-speed image processing. This image processing speed of feature point method can already meet real-time needs in robot navigation and other fields. But if one want to sample audio vibrations up to several kHz, this image processing rate cannot reach the level of real-time processing. In addition, because the speckle pattern is different from the general pattern, mismatches will occur more often. Accuracy also cannot be guaranteed.

In general, the application of using the feature point algorithm to calculate the image displacement is more common in the fields of visual slam and 3D modeling. There are few cases of applying the feature point method in the field of speckle pattern analysis, only in [71, 72]the possibility of speckle patterns was explored. This method is not suitable in the field of ultra-high-speed calculation of the displacement of the speckle pattern.

3.4.3 DIC method versus optical flow method

DIC method and dense optical flow method are both area-based image displacement calculation algorithms. The DIC algorithm performs discrete Fourier transform on the image and calculates the image displacement through the upsampled cross correlation result. On the other hand, optical flow analyzes the image displacement by establishing a two-dimensional quadratic function model of the partial image based on the image gradient. Both of these two algorithms can achieve sub-pixel accuracy. Matlab provides functions to implement the two algorithms. We use these two algorithms to conduct experiments in Matlab to obtain the performance of these two algorithms in terms of speed and accuracy.

3.4.3.1 Image processing speed

First, the calculation speed of the two algorithms is investigated. The speed test is carried

out under different image size, obtaining the average computational time of 10000 times calculations. It shall be noted that the upsampling factor is 100, which means the displacement determination resolution is 0.01 pixels. For Farneback optical flow method, the number of image pyramid and the iteration times at each pyramid are both 3. The test result is shown in Figure 3.23.

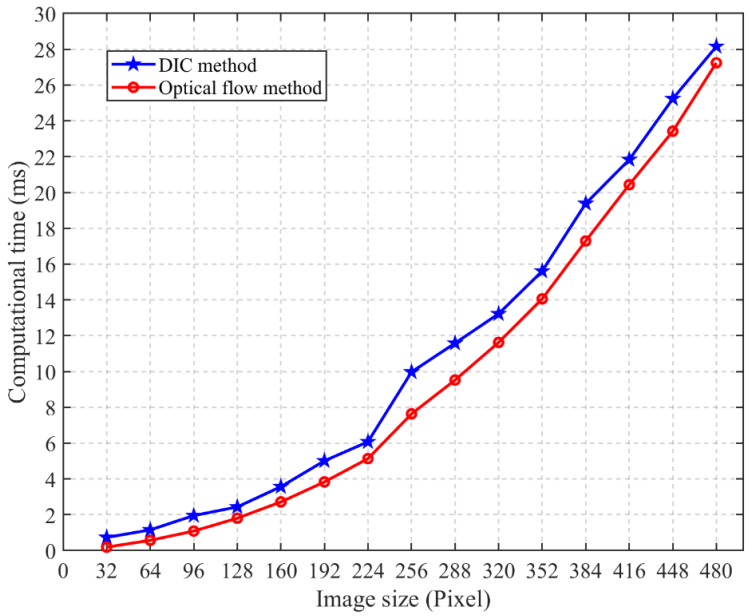

Figure 3.23 Dependence of the computational time consumption on the image size for DIC algorithm and optical flow algorithm (pyramid=3, iteration =3)

From the result one can see that the image size directly affect the calculation time for both two algorithms. In this experiment the tested image size gradually reduced from 480×480 pixels to 32×32 pixels with a step size of 32 pixels. It can be seen from Figure 3.23 that the computational time of DIC and optical flow methods both decreases with the decreasing in image size. Specifically, it takes about 28ms for DIC method to calculate image shift with size 480×480 pixels, whereas the time comsumption is 0.73ms with image size of 32×32 pixels. On the other hand, for optical flow method the time comsumption is 27ms and 0.18ms for the

two image size.

With respect to the computational speed of the two algorhtms, the optical flow method shows better performance than DIC method. Especially, as the image size decreases, the calculation speed of the optical flow method is much faster than DIC method. In the case of small image sizes that we are more concerned about, the optical flow method shows more efficient calculation speed. In fact, when the optical flow method is executed in the C++ environment (which is also the working environment for our camera), it shows a faster speed, as shown in Figure 3.24. Since the optical flow algorihtm provide several input parapmeters that are selectable, such as the number of image pyramid levels and iteration times at each level, those parametes are aslo investigated and in Figure 3.24 we also show the computational time cost with different algorithm parameters.

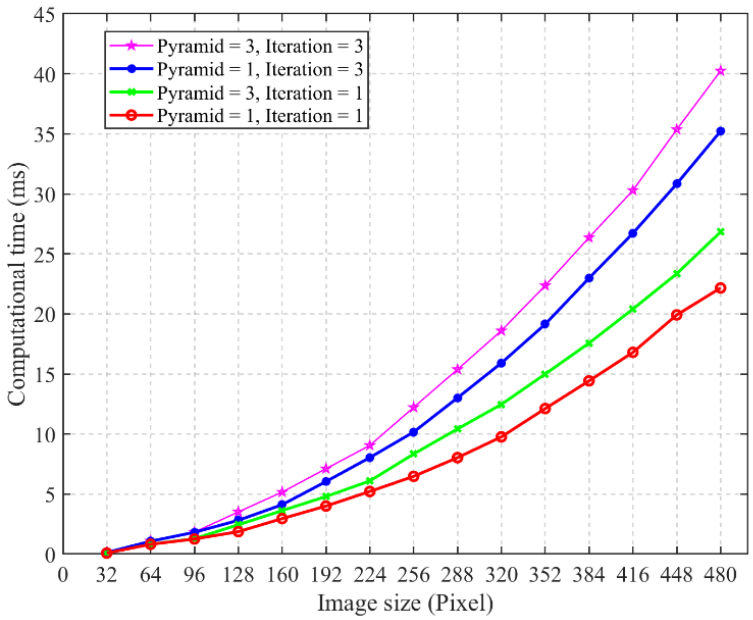

Figure 3.24 Dependence of the computational time consumption on the image size for optical flow algorithm with different parameters in C++ environment

It can be seen that in the C++ environment, reducing the number of image pyramid levels or

reducing the number of iterations both make the calculation faster. This is easy to understand because the calculation process is simplified. Among the two parameters, the number of iterations has a greater impact on the algorithm speed. In the case of small-size images that we are concerned about, the optical flow algorithm takes only 84.4μs to complete one calculation under the condition that the number of image pyramid levels and iterations are both 1. In other words, when the image size is 32×32 pixels and the optical flow method does not perform iteration and image pyramid processing, the algorithm can complete 11848 calculations per second for image displacement analysis. Certainly, the calculation not only includes image displacement analysis, but also includes other parts, the actual calculation rate will be less than this value. However, this fast image processing speed makes it possible to recover the audio signal in real-time.

3.4.3.2 Displacement estimation accuracy

In this part another significant indicator of the algorithm is investigated, which is displacement estimation error. In order to obtain results as accuracy as possible, an image sequence with known shift is needed. In [73] a method for producing high-precision experimentally subpixel shifted images has been created. The author provided a set of speckle samples which are used for this test. The known shift amount of the images is 0~0.9 pixels in both horizontal and vertical directions, with a step size of 0.1 pixels. Thus 100 images are used for determining the accuracy of DIC and optical flow algorithm, and again the test is conducted under different image size.

Figure 3.25 presents the dependence of the displacement estimation error on the image size for the two algorithms. From the result one can see that the calculation by optical flow algorithm gives better result than DIC method, especially when the image size is reduced, the

DIC algorithm's calculation error increases drastically, but the optical flow algorithm can still maintain a high calculation accuracy. When the image size is 32×32 pixels, the average calculation error of DIC algorithm is 0.13 pixels, while the calculation error of the optical flow algorithm is only 0.0087 pixels. The DIC algorithm and the optical flow algorithm have little difference in calculation time, but in terms of calculation accuracy, especially when the image size is small, the reliability of the optical flow method far exceeds that of the DIC algorithm.

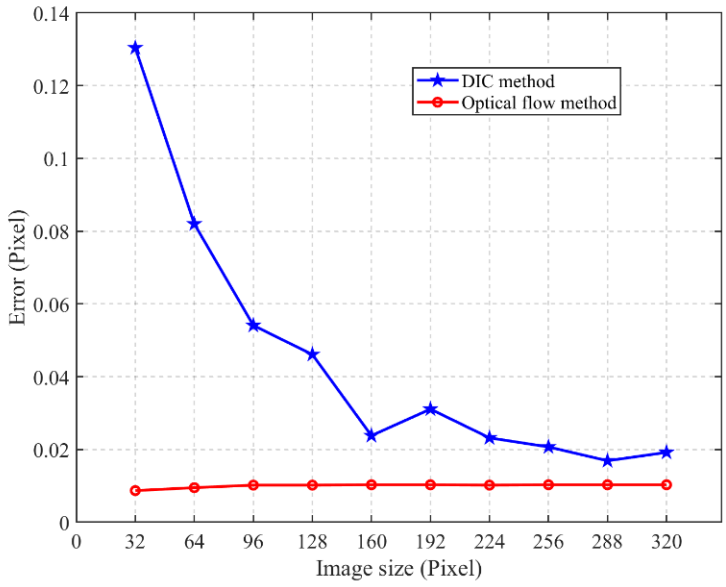

Figure 3.25 Dependence of the calculation error on the image size for DIC and optical flow algorithms

Next, we test the dependence of the calculation error on the image size for optical flow algorithm with different parameters, and the result is shown in Figure 3.26. The result proves that simplifying the processes in the optical flow algorithm, such as the number of image pyramid levels or the number of iterations, will not affect the final calculation accuracy. Even if the image is not down-sampled and iteratively calculated, the calculation result is not worse than that of the three-level image pyramid and three iterations per level. The reason for this phenomenon is that the tested displacement is very small, even the maximum displacement is

only 0.9 pixels. On the other hand, the function of the image pyramid is to make up for the shortcomings of the optical flow algorithm when calculating large-scale displacements, so that the image displacement after the pyramid downsampling matches the prior input of the displacement in the optical flow method as much as possible ($\mathbf{d} = 0$), and further improve the accuracy through iterative calculation. However, when the displacement of the image is small, the displacement of the full-resolution image is already very close to the prior input ($\mathbf{d} = 0$). Therefore, the accuracy of the image can be guaranteed even if the pyramid downsampling and iterative calculation are not performed on the image.

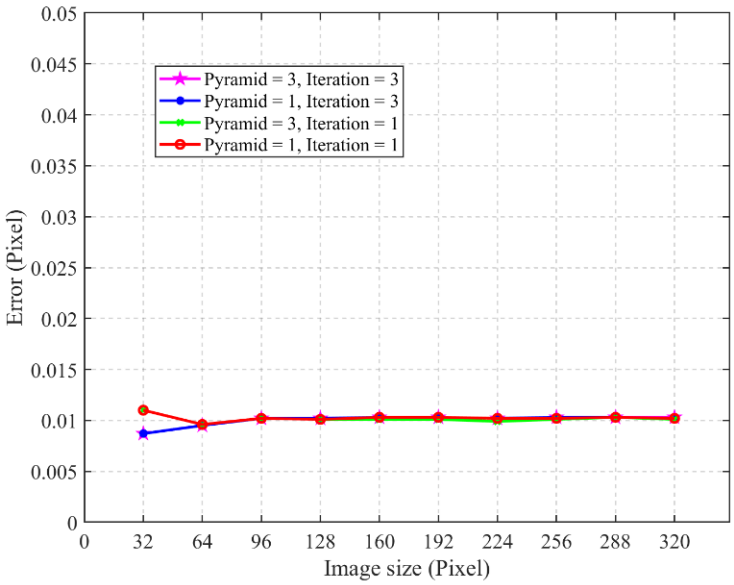

Figure 3.26 Dependence of the calculation error on the image size for optical flow algorithm with different parameters

3.4.3.3 Large displacement problem

It can be seen from the previous experimental results that when the image displacement is small, the optical flow algorithm is slightly faster than the DIC algorithm in speed, and far exceeds the DIC algorithm in calculation accuracy. However, based on the previous analysis

of the optical flow algorithm, it is not difficult to infer that the optical flow method has problems when calculating large displacement. Even if the optical flow algorithm uses image pyramid to try to make up for this defect, the large-scale displacement estimation is still an issue for the optical flow algorithm. In the next experiment, we manually shift the speckle image with increasing amount from 1 pixel to 10 pixels. The image size is 32×32 pixels. Then optical flow method is employed to determine the image displacement. For comparison, the same test is also conducted with DIC method. The dependence of the calculation error on the shift amount for the two algorithms is shown in Figure 3.27.

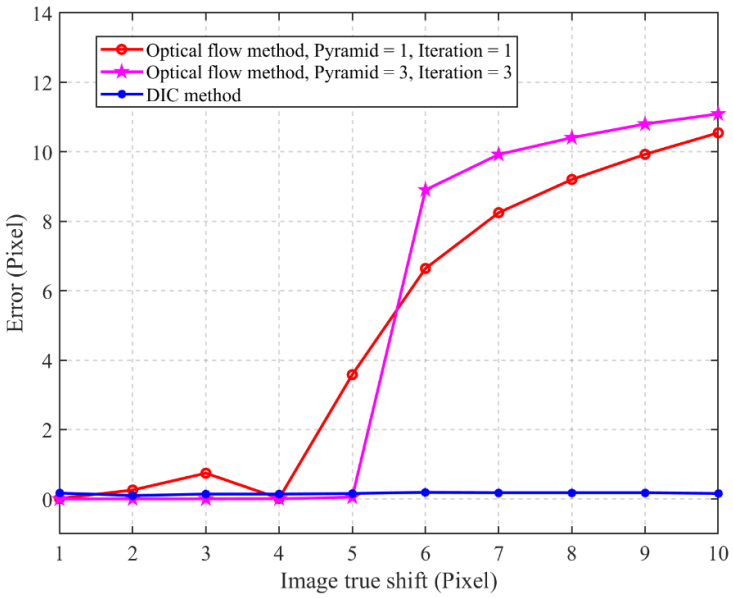

Figure 3.27 Dependence of the calculation error on the shift amount for the DIC algorithm and optical flow algorithm

It can be seen from the result that the optical flow method, especially without image pyramid processing, is very unstable under large-scale image displacement. When the displacement is larger than 4 pixels, the calculation result given by optical flow algorithm is no longer reliable. Using image pyramids and iterative in the algorithm can increase the stability of calculation

accuracy and obtain more accurate displacement estimation results when the displacement is less than 5 pixels. However, this performance enhancement has limitation. When the displacement is larger than 5 pixels, the calculation result of the optical flow algorithm again become unreliable. In contrast, although the accuracy of the DIC algorithm is not as good as the optical flow method when the displacement is small, DIC algorithm has the advantage that it can always give a stable displacement analysis result. Even if the displacement continuously increases, the error of the DIC algorithm does not increase significantly.

In the past, we used an industrial camera with a frame rate of 2300fps under 32×32 pixels resolution as the image receiving device of the speckle vibration measurement method. Because the frame rate of the camera is low, the displacement between the adjacent images may sometimes exceed the effective calculation range of the optical flow algorithm especially when the defocus is high and the object vibration amplitude is large, resulting in distortion of the restored audio signal waveform. In this case, the range of speckle amplitude can be controlled by adjusting the defocus of the camera, so that the optical flow algorithm can accurately calculate the displacement of the images. Normally, the displacement between consecutive pictures in a speckle-based vibration measurement method is very small if the camera captures image with very high frame rate, so not only do not have to worry about the distortion of the optical flow algorithm calculation results, but you can boldly abandon the image pyramid and iteration to obtain higher image processing speed.

3.5 Chapter Conclusion

Image processing, specifically, calculating the displacement of the speckle image is an important part of the speckle vibration measurement method. For the proposed method, our target is restoring audio signal in real time while using laser speckle images to detect audio

vibration. Therefore, according to the Nyquist theorem, the proposed algorithm must have a calculation rate at the sound frequency level and be able to calculate a small image displacement at the sub-pixel level. In order to achieve this goal, reducing the size of the captured speckle image is the most direct and effective approach, which can greatly increase the image processing rate.

Among the image displacement calculation algorithms, the feature point methods are widely used visual SLAM, and the image features can be matched correctly even the image is rotated or the scale changes. However, for speckle sensing method, there is no such calculation requirements, therefore the feature point method is rarely used in speckle image processing. The DIC algorithm has been widely used in speckle image analysis in the past and it is a very effective calculation method. However, the experimental results show that the accuracy of the estimated displacement by the DIC algorithm is unstable when the image size is small. We finally decided to adopt the optical flow algorithm as the image processing algorithm of the proposed speckle vibration measurement method. The optical flow algorithm can provide the dense optical flow field of the image, that is, the displacement vector of each pixel in the image, and the optical flow algorithm is efficient, which can meet the needs of audio signal sampling. In addition, since the time interval between speckle images captured by high-speed camera is very short, no large displacement happens in the speckle image sequence. Therefore, the shortcomings of the optical flow algorithm are perfectly avoided. The experiment results in this chapter also show that the optical flow algorithm is suitable for the proposed method.

CHAPTER 4 REAL-TIME ACOUSTIC VIBRATION SENSING METHOD USING 2-D SPECKLE IMAGE

Sound detection with optical means is an appealing research topic. In this chapter, we proposed a laser microphone method allowing simultaneous detection and regeneration of the audio signal by observing the movement of secondary speckle patterns. In the proposed method, optical flow method, along with some denoising algorithms are employed to obtain the motion information of the speckle sequence with high speed. Owing to this, audio signal can be regenerated in real time with simple optical setup even the sound source is moving. Experiments have been conducted and the results show that the proposed method can restore high quality audio signal in real time under various conditions.

4.1 Background Introduction

Sound detection with optical means is an appealing research topic due to its broad application prospects, such as remote monitoring, rescue, and so on [74, 75]. One of the approaches is detecting sound with laser speckle images. The principle of laser speckle method is simple: when a coherent light is reflected by an optically rough surface, a high-contrast grainy speckle pattern can be observed with an image device due to the interferometry of the multiple reflection light waves. A major property of speckle pattern is that the speckle motion is very sensitive to the motion of the object. The captured speckle pattern shows significant

displacement even the object motion is extremely slight. Based on this property, sound vibrations can be detected with speckle images, and the audio signal can be recovered by extracting information from the movement of the captured speckle image sequence. Compared with other method, like the interferometric or holographic measurement method [76, 77], the laser speckle detection method has a simple structure and low hardware cost and can achieve remote sound detection. Previously there were several studies on recovering sound with laser speckle, mainly focused on the applications in remote monitoring. In [31] the authors proposed a remote sound extraction method based on laser speckle. The result shows they can record the speech or heart beats with the distance up to 100 meters. In [33], the authors proposed an intensity variance-based method for sound recovery via the appropriate pixels' gray-value variations from the laser speckle patterns. In the previous studies, people usually record a video of speckle motion and then analyze the video to restore the audio signal. Although these works successfully achieved sound regeneration with laser speckle images, still real-time sound detection and regeneration have not been considered, nor has it been considered for detection under moving sound source situation, which greatly limits the potential applications of this technology.

In this chapter, a real-time sound detection and regeneration method based on laser speckle image is proposed. Different from the previous research, the proposed method for the first time took the real-time processing and regeneration of audio signal with moving sound source into consideration. In our method, after capturing speckle images, high-speed calculation is conducted immediately to obtain the displacement of the captured speckle images instead of saving the images into the computer's hard disk. Thus, the method can output audio signals in real time while sampling. To achieve this, only a small part of the imaging sensor is used to capture the speckle patterns. In this way, not only high camera sampling rates can be achieved

even with a common industrial camera, but also the computation time can be reduced because of the small image size. Moreover, optical flow algorithm is adopted to obtain the displacement between two frames in a short time. These two points enable a real-time processing speed and sub-pixel level accuracy. In addition, some denoising algorithms are proposed to correct the calculation noise in real time, which will be explained in the following part. This not only improves the accuracy of the results, but also enables the method to regenerate audio signal with moving sound sources. Compared with the previous methods, our method works more like a microphone rather than a recorder, which enables our method to have a wider range of potential applications, such as a meeting scenario, monitoring, and so on.

4.2 Methodologies

The algorithm's flowchart of the proposed method is shown in Figure 4.1. The specific description of the whole process is as follows.

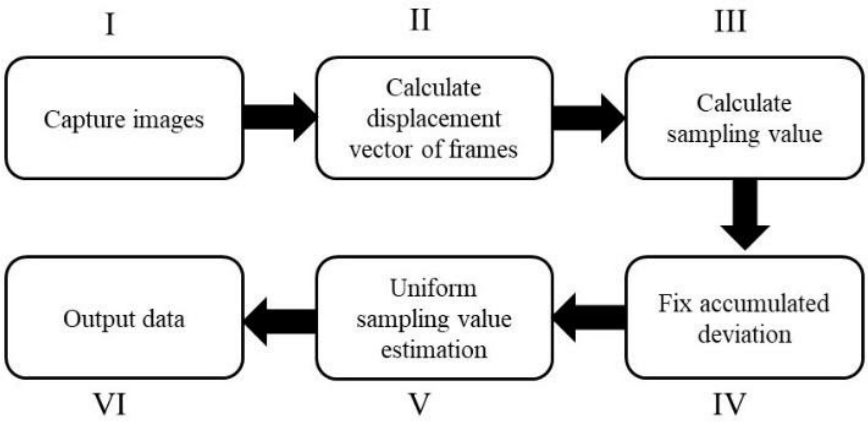

Figure 4.1 Flowchart of the real-time signal processing algorithm of our method

Capture Images: The frame rate of the camera directly determines the detectable frequency

range of the laser microphone. In order to reduce the computational cost and improve the transmission speed, we try to use a common USB3.0 camera to capture the speckle images. The window size is set to be 32x32 pixels. Under this resolution the camera can reach the frame rate of 2300fps.

Calculate displacement vector of frames: After getting frame sequence, the Farneback optical flow algorithm is adopted to obtain the motion between frames. As shown in Figure 4.2(c), the dense optical flow algorithm computes the motion vector for every pixel between two images. Here one parameter "Window Size" in the algorithm is settled to be 32, and in this way the results of each pixel are approximately the same. The average of all vectors is taken as the global shifting $\mathbf{d}(x, y)$ between two frames.

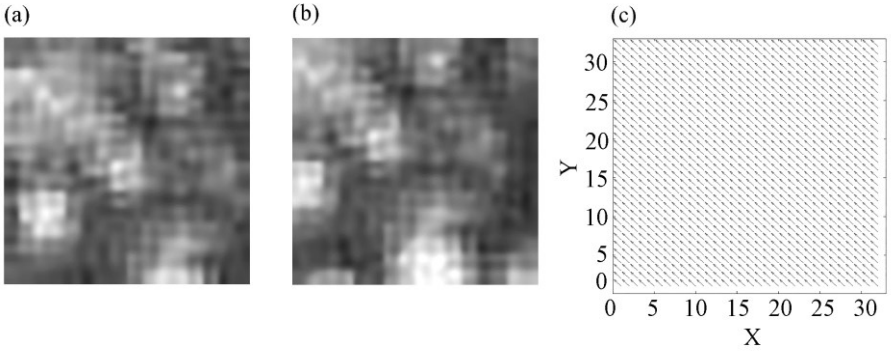

Figure 4.2 Two speckle images and the optical flow field between them. (a) Former frame. (b) Later frame. (c) Optical flow.

Calculate sampling value: After getting vector of displacement $\mathbf{d}(x, y)$ between adjacent frames, the length $|\mathbf{d}|$of the displacement between the two frames can be easily obtained as:

$$|\mathbf{d}| = \sqrt{x^2 + y^2} \qquad (4-1)$$

Then the angle α of the global displacement vector is given by:

$$\alpha = \tan^{-1}\frac{y}{x} \qquad (4-2)$$

According to the speckle motion model, the speckle sequence shows nearly linear vibration when the object of interest vibrates. Figure 4.3 is the statistical histogram that shows angle of displacement of every two frames in 10,000 images. From Figure 4.3 we can see that the angles are clearly distributed in two different intervals when the speckle sequence linearly reciprocates. Therefore, the displacement length $|\mathbf{d}|$is superimposed according to the direction angle of the displacement vector to obtain a sinusoidal waveform representing the original signal.

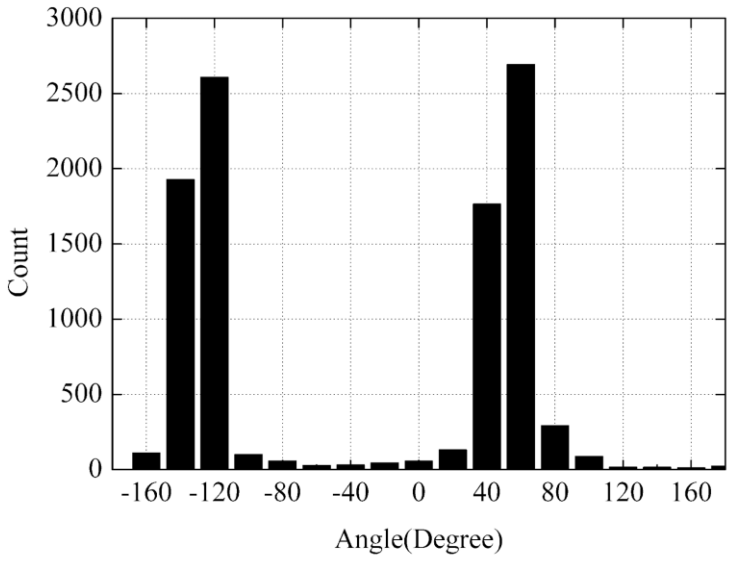

Figure 4.3 Angle statistical histogram of 10 thousand vectors

Fix accumulated deviation: Because of the noise of the image sequence, each calculation of displacement brings a tiny deviation. Especially the high sampling rate causes the deviation to be accumulated very quickly. Assuming the error for each calculation is e_i, the displacement between frame i and frame $i+1$ be $|\mathbf{d}|_i$, the accumulated displacement be s_i, which can be expressed as:

$$s_i = \sum |\mathbf{d}|_i + \sum e_i \qquad (4-3)$$

Taking a single frequency audio signal as an example. Ideally the accumulated result $\sum|\mathbf{d}|_i$ presents a sinusoidal waveform. However, due to the accumulation deviation $\sum e_i$, the regenerated waveform performs global drift. Figure 4.4 shows a waveform of regenerated 50Hz audio signal. As shown with the black line, the regenerated sinusoidal waveform drifts drastically in only 10 seconds.

Fixing the drift problem is crucial. In our method, we always take the latest 100 points, in other words, the sampling data of the latest 0.1 seconds, as sample to estimate the real-time drift slope k. Every time after e_i is obtained, the data from s_{i-100} to s_{i-1} are used to calculate the real-time drift slope k to help fixing the drift and obtain the fixed data S_i, which can be expressed as:

$$S_i = s_i - k \times i \qquad (4-4)$$

The fixed waveform is shown with the red line in Figure 4.4. From the result we can see that although the accumulated deviation will cause drift, it can be fixed, and the flat waveform can be obtained with the denoising algorithm. Besides, real-time fixing of drift makes it possible to detect the audio signal with moving sound source, which will be explained in the following part.

Uniform sampling estimation: With real-time fixing, the drift problem can be solved. However, because the time consumption of capturing image and calculating optical flow for each sampling is not exactly equal, the sampling rate is not uniform. This will lead to noise if the non-uniform sample data is replayed.

To deal with this problem, here an estimation algorithm is proposed to obtain uniform

sampling value. Since the sampling rate is fast, the variety between each two sample points can be approximated as a linear function. As shown in Figure 4.5, every time after obtaining two sample values S_i and S_{i+1} at the time of t_i and t_{i+1}, we estimate the sample value S_{i_u} corresponding to all T_i times in the $[t_i, t_{i+1}]$ interval, where T_i is a uniform time with the interval of 1ms. In this way uniform sampling value S_{i_u} can be obtained.

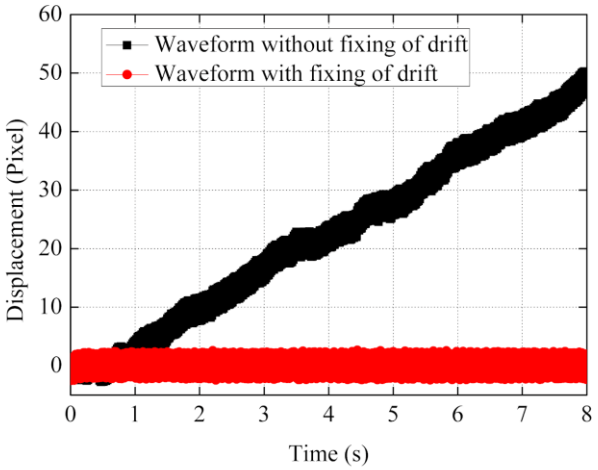

Figure 4.4 Waveform before and after fixing of accumulated drift

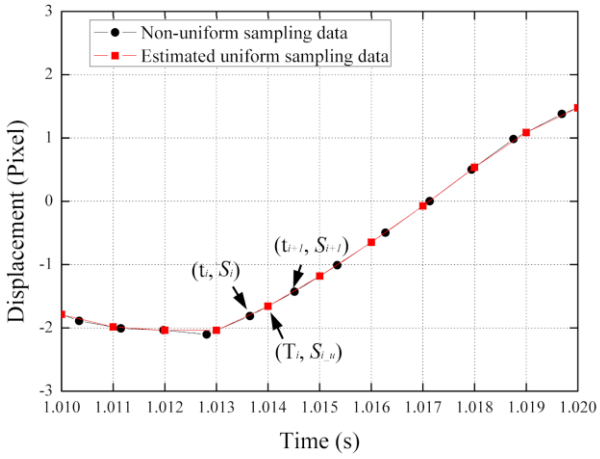

Figure 4.5 Part of the waveform of uniform sampling date estimation

The above is the processing of each sampling data. For our method, it takes about 1ms from step I to step VI to obtain one sampling data. According to Nyquist sampling theorem, it means audio signal under 500Hz can be regenerated in real-time.

4.3 Experiment Result

4.3.1 Single frequency test

In the first experiment we tried to regenerate single frequency audio signal with our method. The experiment setup is shown in Figure 4.6, and the schematic is shown in Figure 4.7. An expanded laser beam with the output power of 100mW at the wavelength of 650nm is illuminated on the membrane of the speaker. The PointGray GS3U3-32S4C-C camera with a $f = 25mm$ lens is used to capture the speckle images. As mentioned above, the image resolution is settled to be 32×32 pixels, and the frame rate of the camera is 2300fps. The camera is connected with a desktop that controlled with python code. Both the laser and the camera are positioned around 1m away from the speaker.

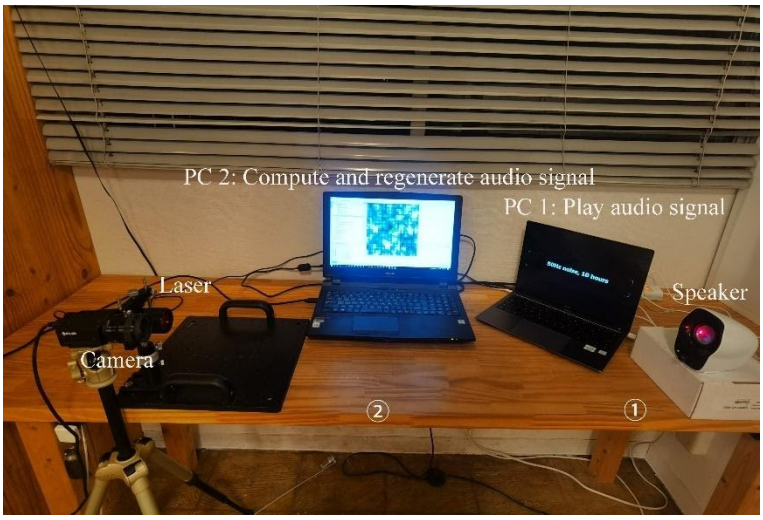

Figure 4.6 Experiment scenario of the proposed laser microphone method

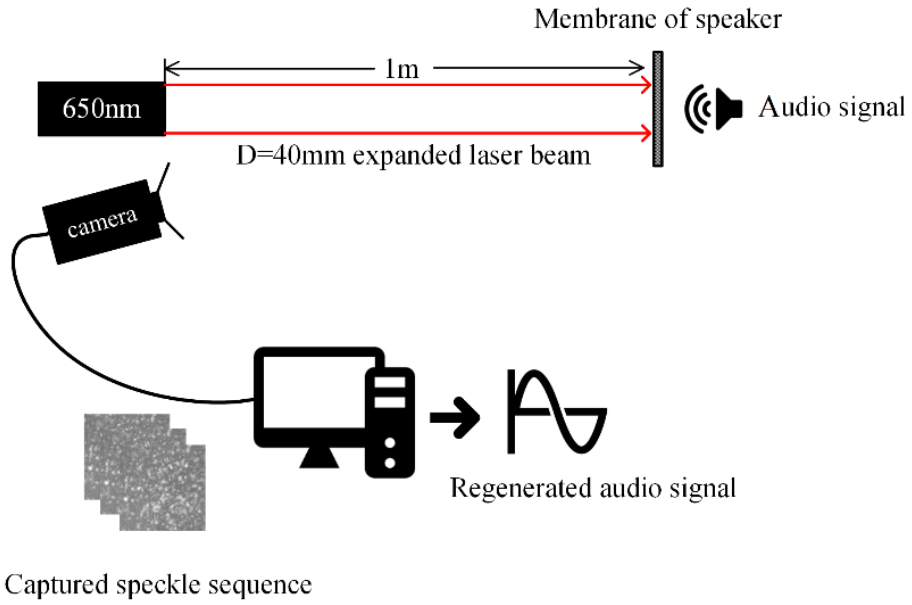

Figure 4.7 Schematic diagram of the proposed laser microphone method

The audio signals with the frequencies of 50Hz, 100Hz, 150Hz and 200Hz are tested, respectively. The waveforms of the regenerated audio signal are shown in Figure 4.8, and Figure 4.9 shows the spectrum after the Fourier transform of each result. From the results we can see that the information extracted from the speckle motion can correctly represent the frequency of the signal source. Especially in the low-frequency region, the regenerated waveforms are very clear. With the frequency increases, the quality of the regenerated audio signal becomes worse. This is mainly caused by the limited sampling rate of the camera. It is foreseeable that if a higher speed camera is employed, the method will be able to regenerate audio signals with higher frequency.

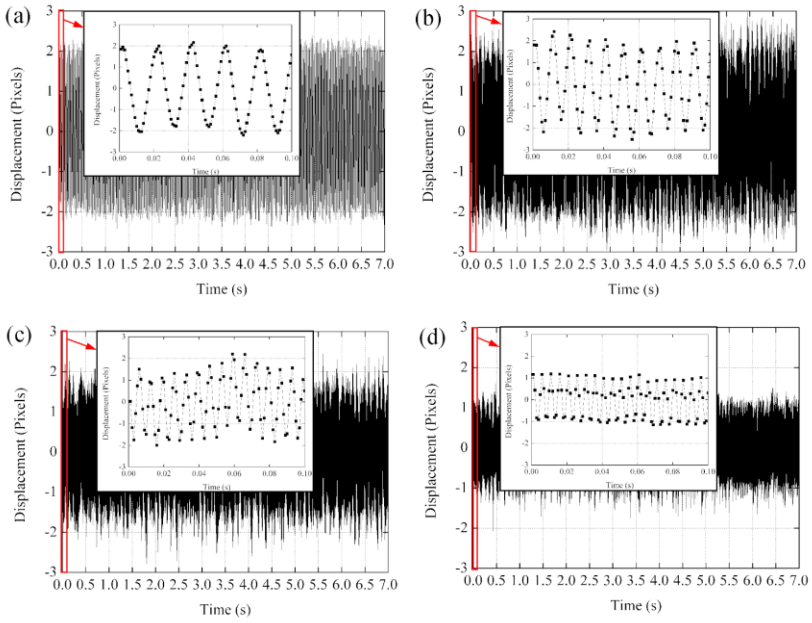

Figure 4.8 Regenerated waveform of different frequency audio signal. (a) 50Hz audio signal. (b) 100Hz audio signal. (c) 150Hz audio signal. (d) 200Hz audio signal

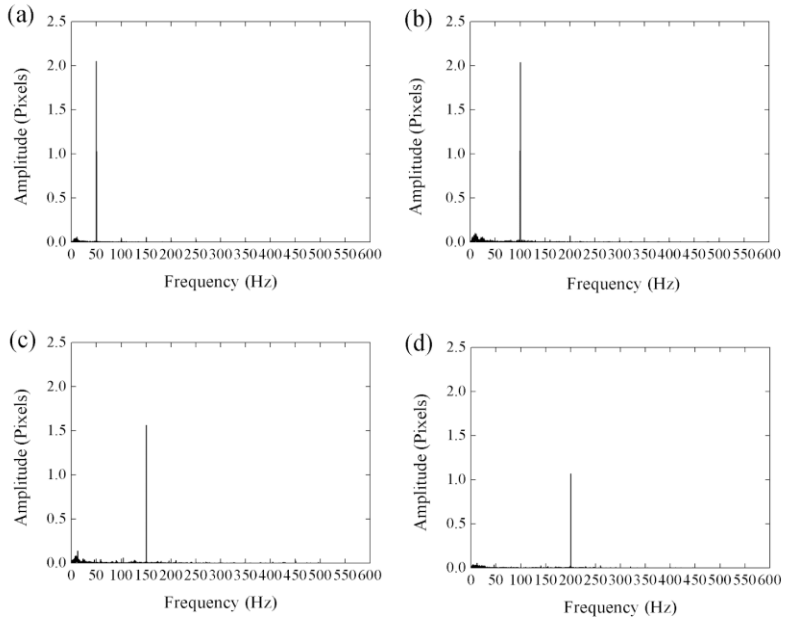

Figure 4.9 Frequency domain diagram of each result. (a) 50Hz audio signal. (b) 100Hz audio signal. (c) 150Hz audio signal. (d) 200Hz audio signal

In the next experiment we present that music can also be regenerated in real-time with the proposed method. We used Matlab to edit and generate the music " Moonlight Sonata No. 14 " for the first forty seconds and played it. Figure 4.10 shows the spectrogram of the created music, and Figure 4.11 shows the spectrogram of the regenerated music. The experiment shows that the music can be regenerated with high quality in real-time by the laser speckle and the proposed algorithm. For reference the audio file of both original music and regenerated music has been provided as the result of this experiment.

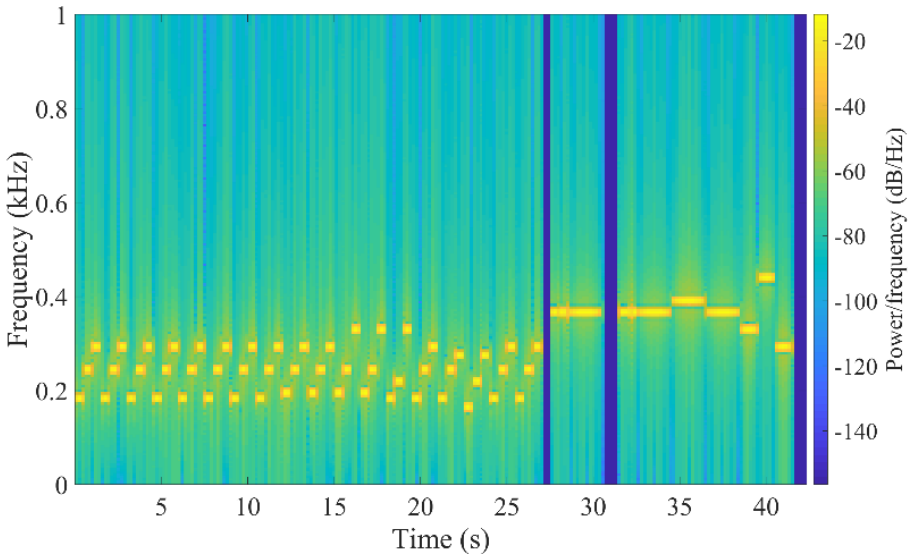

Figure 4.10 Spectrogram of the original audio signal

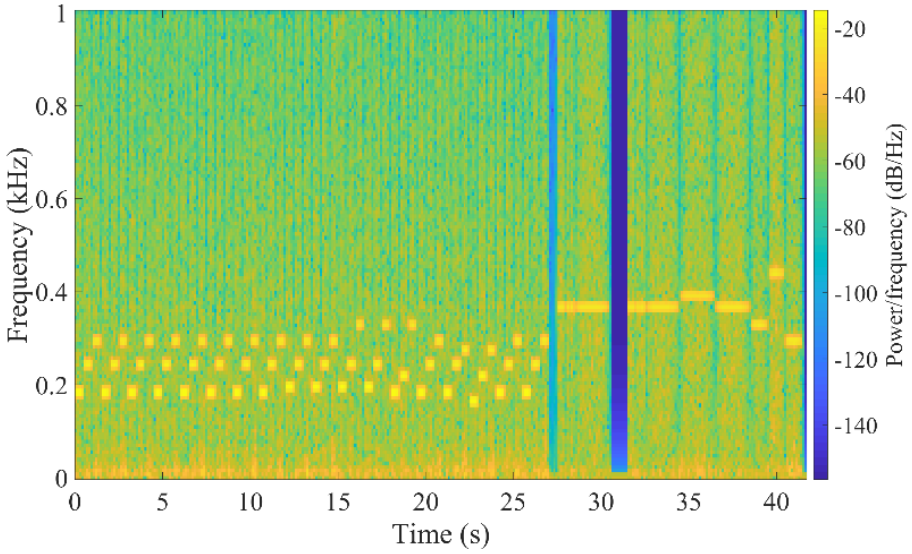

Figure 4.11 Spectrogram of the regenerated audio signal

4.3.2 Effect of amplitude and defocusing on the result

For the proposed laser speckle detection method, it was found that the amplitude of sound source brings challenge to signal regeneration. The increasing amplitude of the object vibration causes the displacement of the speckle to become larger, which in turn leads to a decrease in the degree of correlation between adjacent frames. When the displacement between adjacent frames is large enough, it cannot be correctly calculated since the correlation between two frames are too small. As mentioned above, the image size is set to be 32×32 pixels in our method. Small window size certainly increases the sampling speed. However, on the other hand, it also weakened the ability to observe large speckle motion.

Here the amplitude of the audio signal is gradually increased, and the performance of the method is investigated. The 50Hz audio signal is played by the speaker, the amplitude of the audio signal is adjusted by the volume of the computer which is 48.3dB, 54.3dB, 57.9dB and 61.0dB respectively. Correspondingly, the signal to noise ratio (SNR) of the result is 28.76dB,

31.01dB, 16.83dB, and 1.38dB. Figure 4.12 shows the part of the regenerated waveforms under different amplitude. The result shows that it is easier to recover a clear sinusoidal waveform when the speckle motion is small. However, as the speckle motion becomes larger, the recovered waveform is gradually distorted and the SNR of the regenerated signal decreases.

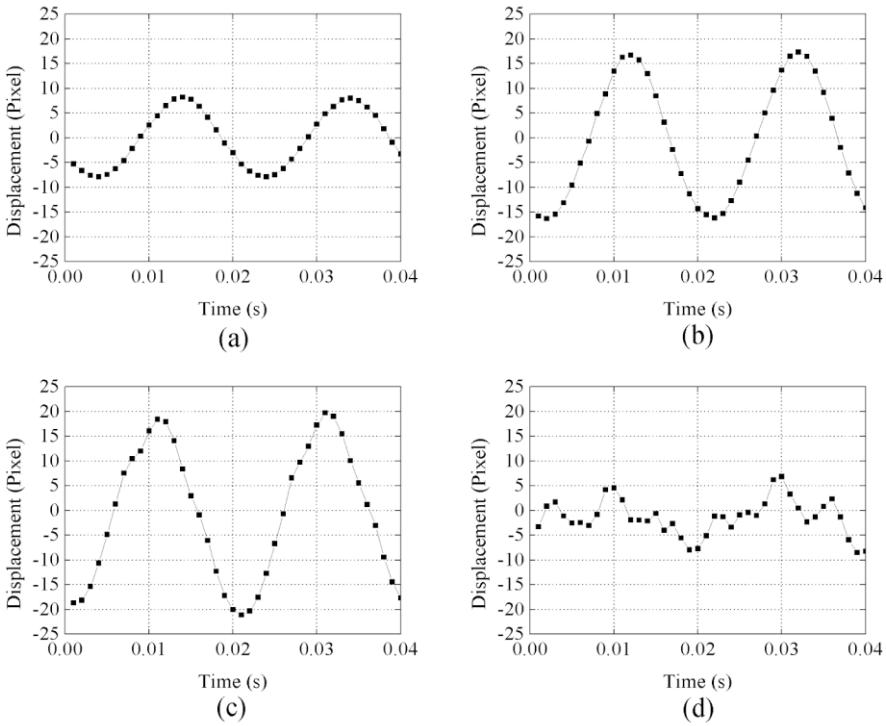

Figure 4.12 Regenerated audio signal with different amplitude. (a) Sound volume is 48.3dB. (b) Sound volume is 54.3dB. (c) Sound volume is 57.9dB. (d) Sound volume is 61.0dB

One way to deal with this problem is using a higher speed camera to get a denser sequence of sampled images so that the distance between every two frames is not too large. On the other hand, adjusting the defocusing amount of the camera can also solve this problem. When the camera focuses on the speckle at the near field from the investigated object (i.e., location ① in Figure 4.6), the amount of defocusing is small. At this situation, the speckles will overlap on the image, and the speckle motion is not sensitive to object motion. Conversely, when the

camera focuses on the speckle at the far field (i.e., location ② in Figure 4.6), the amount of defocus is large. In this situation the bright and dark speckle are distributed on the image clearly, and the speckle motion is sensitive to object motion.

In the next experiment, the distance between camera and object is set to be 1m, and the amount of camera defocusing is set to be 0 (focus), 0.3m, 0.5m, 0.6m, 0.7m, 0.75m and 0.85m respectively. Figure 4.13 shows the image captured under different camera defocusing. In our method, the wavelength of the laser source is 650nm, and a color image sensor is used to capture the speckle pattern. Thus, the captured image shows red grain pattern, as shown in Figure 4.13(g). When the camera is focused, the light intensity received by a single pixel becomes stronger. Because the color pixel has four channels (red, green, green, blue), some pixels of the image become colorful. The speaker plays a 50Hz sine wave with different amplitude (48.3dB, 54.3dB, 57.9dB and 61.0dB respectively). The SNR of the result with different situations is shown in Figure 4.14.

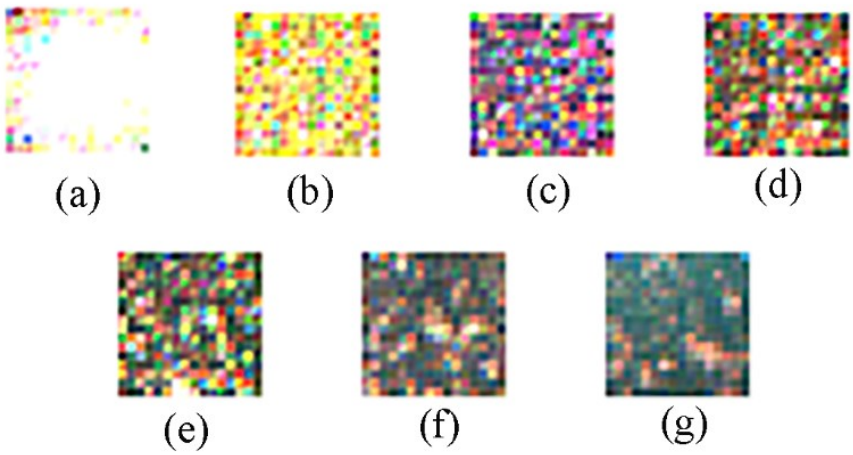

Figure 4.13 Speckle image captured under different amount of camera defocusing amount L. (a) $L = 0$. (b) $L = 0.3m$. (c) $L = 0.5m$. (d) $L = 0.6m$. (e) $L = 0.7m$. (f) $L = 0.75m$. (g) $L = 0.85m$

First, when the camera focuses on the object (amount of defocus equals zero), speckles overlap together and form a featureless bright spot, as shown in Figure 4.13(a). In this situation the SNR of the result is meaningless because object vibration cannot be observed through speckle motion. With the aspect of defocusing situation, the SNR of results can always keep a high level (over 20dB) when the amplitude of audio signal is small. While in case of large amplitude of audio signal, reducing the amount of defocusing can make the motion of the speckle smaller, which in turn makes the result better. For instance, with large amplitude of 61.0dB, if the camera is focused on the speckle field at 0.5m away from object, the SNR of result reaches the optimal value (30.30dB). Figure 4.15 shows the regenerated waveform in this situation. Compared with Figure 4.12(d), the distortion of the result is fixed due to the adjustment of the camera defocusing.

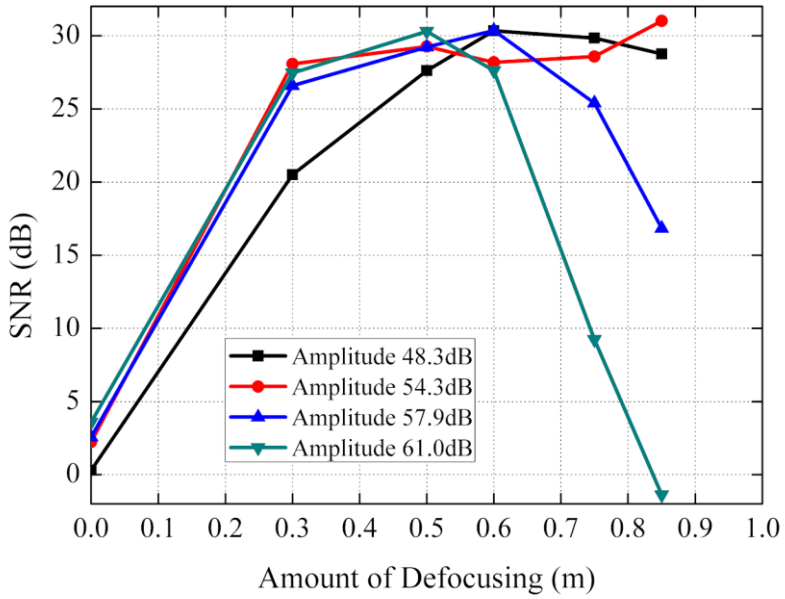

Figure 4.14 SNR of the result with different amount of defocusing under different amplitude of audio signal

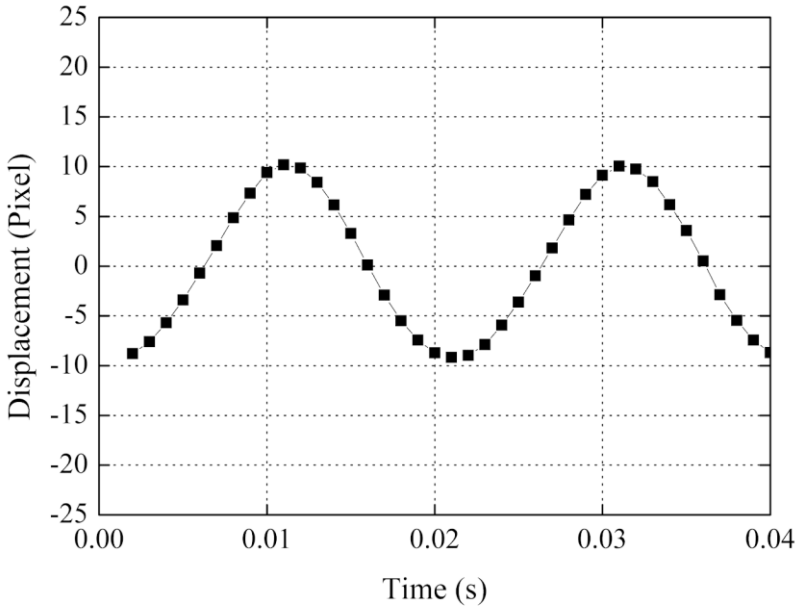

Figure 4.15 Regenerated waveform with 0.5m camera defocusing under the amplitude of 61.0dB

4.3.3 Detection of moving sound source

In the actual situation, usually the sound source cannot maintain absolute stillness. For example, when a person is talking, the body will show a slight movement. Therefore, the detection of moving sound source is investigated. First, we will explain the speckle motion model. The six-degree spatial motion of object can be divided into three categories: transverse, axial, and tilt. According to the previous research, the transverse and tilt motion cause two-dimensional displacement of captured speckle pattern, while axis motion causes scaling variation of captured speckle pattern, as shown in Figure 4.16.

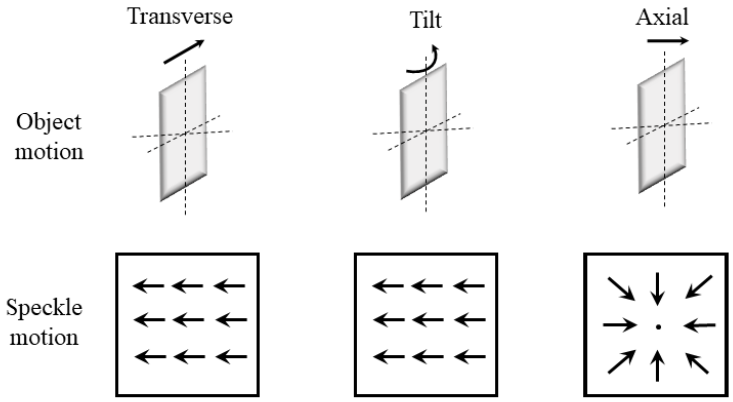

Figure 4.16 Corresponding speckle motion caused by object motion

Therefore, when the sound source undergoes transverse or tilt motion, the motion of the captured speckle consists of two parts: sinusoidal vibration and translational motion, and the calculated waveform will be sine wave with drift. The drift caused by object motion can be fixed in real time by the algorithm mentioned in the previous part.

In the next experiment, the speaker playing the 50Hz audio signal is placed on the linear motor, and motor translates by 30mm at a speed of 5mm/s in transverse direction and then returns to the original position. The black line in Figure 4.17 shows the obtained waveform. The waveform clearly reflects the superposition of sinusoidal vibration and horizontal movement that corresponding to the object motion. Meanwhile the red line shows the fixed waveform with our proposed algorithms, which proves that our algorithm can output the clear sinusoidal waveform in real time without being affected by the motion of the object.

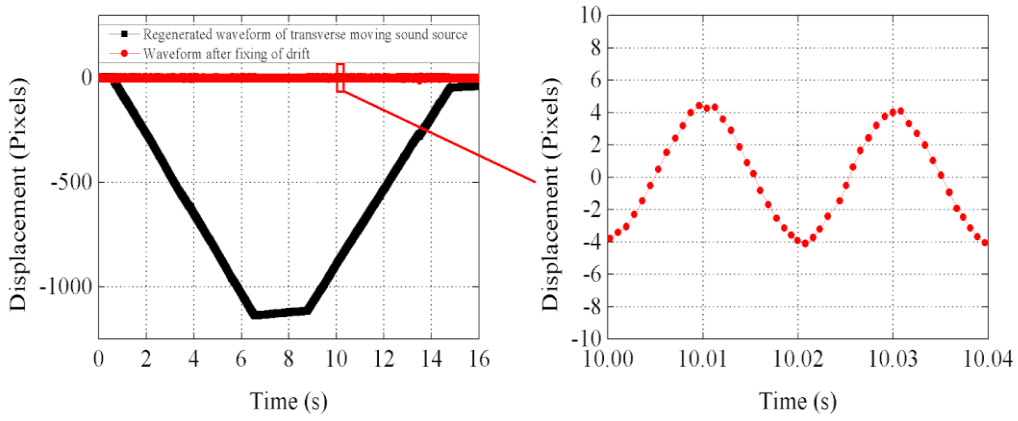

Figure 4.17 Test result of transverse moving sound source

Next the speaker is placed on the rotation motor, and motor rotates by 5° at a speed of 0.5°/s, and then returns to the original position. The black line in Figure 4.18 shows the obtained waveform, and the red line shows the fixed waveform. The result also proves that our method can continuously output a clear audio signal during the tilt motion of sound source.

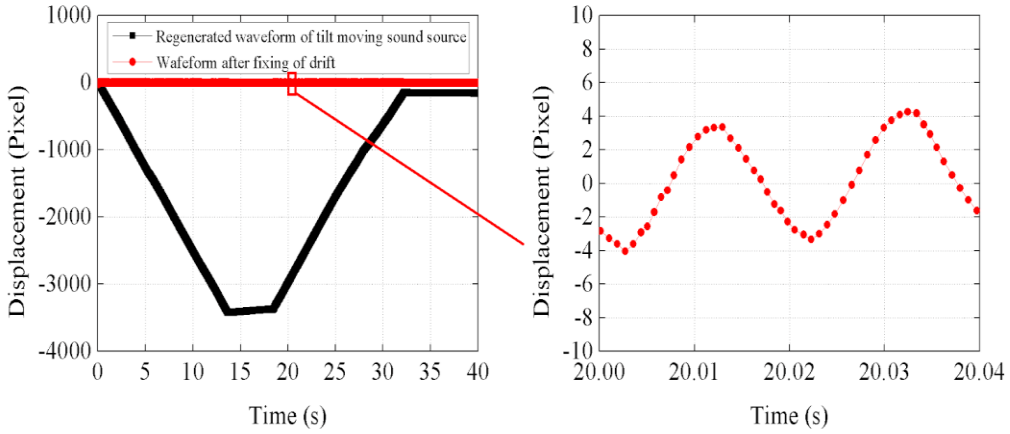

Figure 4.18 Test result of tilt moving sound source

Finally, the axial motion is investigated. The motor translates by 30mm at a speed of 5mm/s in *z*-axial direction and then returns to the original position. The black line in Figure 4.19 shows the obtained waveform. Different from the other two motion, axial motion has little effect on

speckle motion under defocusing situation. It can be seen from Figure 4.19 that the waveform drifts slightly. The drift can also be fixed with our algorithm and the method can output a clear waveform continuously.

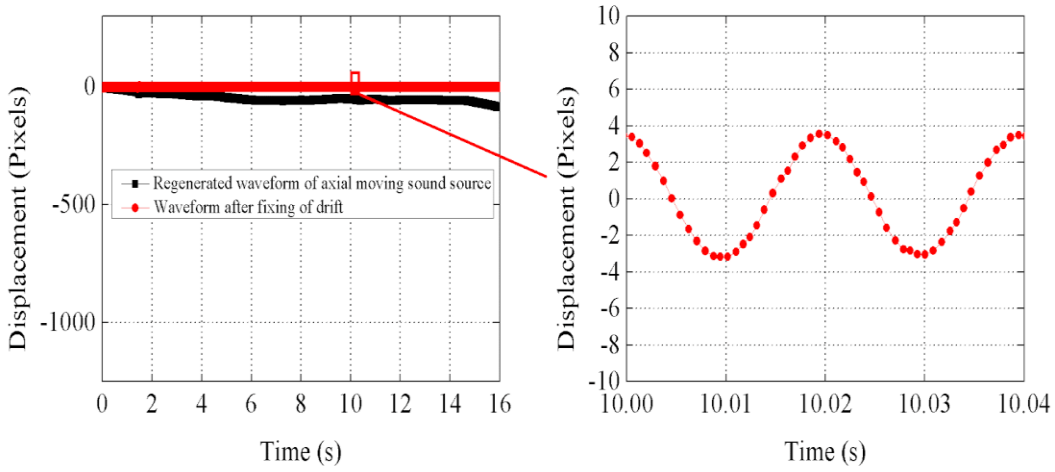

Figure 4.19 Test result of axial moving sound source

4.3.4 Pick up vibrating object from large area using diffuse laser source

In the previous laser speckle-based vibration sensing studies, the laser point is usually used as the light source. The beam diameter of such laser is relatively narrow, usually on the order of several millimeters. At the same time, the area irradiated by the laser on the surface of the object is also very small. In this situation, the energy of the laser is concentrated in a small range, which is suitable for extracting the vibration information of the known target at a long distance, as shown in Figure 4.20(a). However, due to the small irradiation area of the laser light source, this method cannot quickly find the vibrating object in the camera's field of view when the target position is unknown. Here we proposed a new lighting plan, which uses an optical lens to diffuse the laser beam, so that the laser light source can cover a large area of

illumination, as shown in Figure 4.20(b). At the same time, the camera is used to capture speckles within the illumination range. The analysis of the laser speckle optical flow can then be used to determine the source of a vibration within the monitoring range. After the location of sound source is determined, the camera switches to local window with high frame rate, and the audio signal can be restored by the method we mentioned earlier in this chapter.

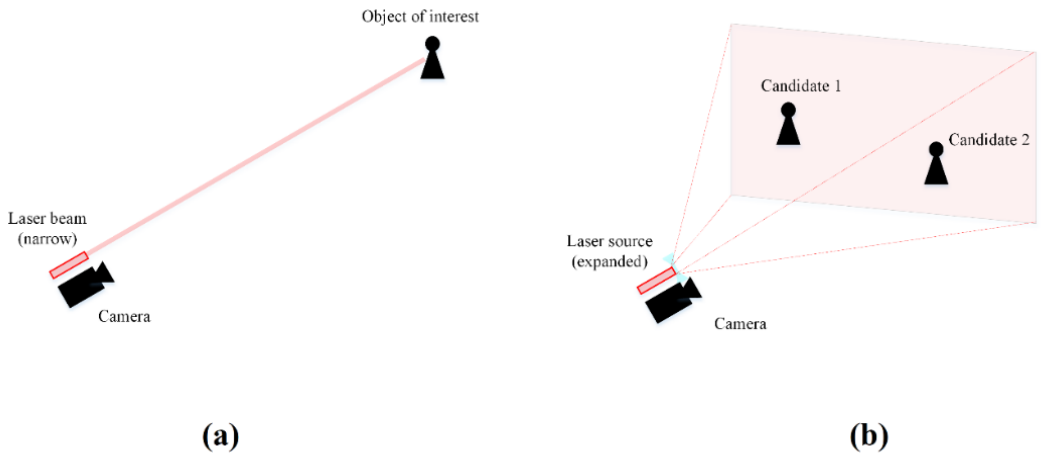

Figure 4.20 Two types of laser source. (a) Laser with narrow beam. (b) Laser with expanded beam.

Expanding the laser bean can increase the coverage area of the light source, but it will also cause variation of the sensing method. First, the power density of the laser beam is decreased due to expansion. For the laser source with same power, the brightness of the captured speckle image is darker with the expanded beam. Second, as the experiment result illustrated in Chapter 2, the average speckle size is small with large illumination area on the object surface. These two points are the effects of using expanded laser beam on characteristics of speckle image. The using expanded laser beam will also cause the motion characteristic of the speckle image.

In Chapter 2 we analyzed the relationship between speckle motion and object motion, proved that in a strongly defocused camera, when the object vibrates the speckle pattern does not change but only shows the displacement. In the previous discussion, the object is assumed to

be a rigid body, which means no deformation happens during object vibration. When the light beam is narrow, the illumination area on the object surface is very small. The small area could be approximately regarded as rigid body. However, if the expanded laser beam is used, a large area on the object surface will be illuminated, and the vibration of the large area, such as speaker membrane's vibration, could not be regarded as rigid body motion. In this situation, although speckle mainly shows displacement, the speckle pattern may change at the same time.

In the next experiment, we tested the restoration of audio signals when laser beam with different diameters irradiates the surface of the object to be measured, which is a speaker membrane. The diameter of the laser beam is 7cm, 3cm, 1.5cm and 0.3cm respectively, as shown in Figure 4.21. Figure 4.22 shows the captured speckle patterns, and Figure 4.23 shows the restored waveform, where the frequency of the original signal is 100Hz.

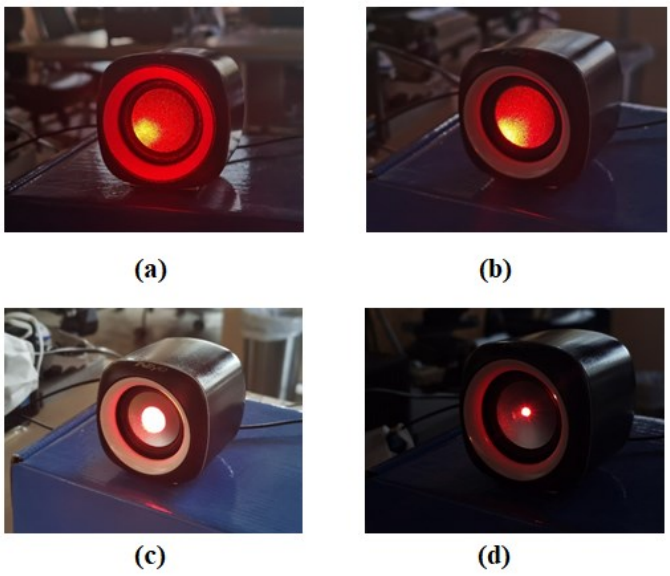

Figure 4.21 Laser illumination with different beam diameter D. (a) D=7cm. (b) D=3cm. (c) D=1.5cm (d) D=0.3cm

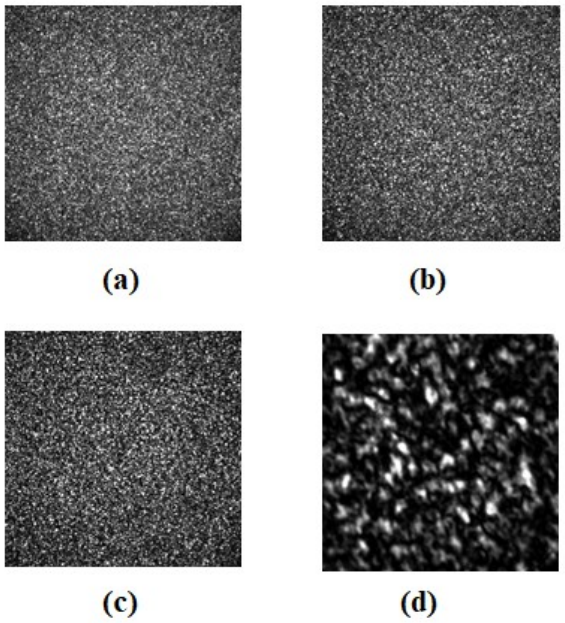

Figure 4.22 Captured speckle patterns with different laser beam diameter D. (a) D=7cm. (b) D=3cm. (c) D=1.5cm (d) D=0.3cm

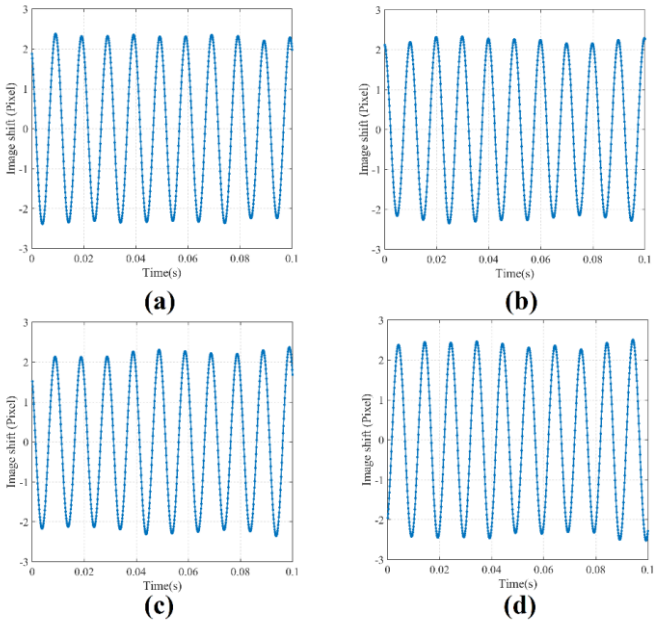

Figure 4.23 Restored waveform with different laser beam diameter D. (a) D=7cm. (b) D=3cm. (c) D=1.5cm (d) D=0.3cm

Based on the results, one can see that a slight deformation of the surface during the vibration will cause the speckle pattern to change, but the speckle mostly reflects the displacement of the image caused by the movement of the object. Hence, vibratory information can still be recovered in the case of expanded beam illumination. As we mentioned above, by using extended light illumination, we can increase the area covered by the laser, which allows us to find the location of the sound source in an image when it is not known where it is. InFigure 4.24 we give the experiment setup of picking up location of vibrating object with diffuse laser illumination, and in Figure 4.25 we show a scene with expanded laser illumination, where the laser beam is diffused by the optical lens, and the illumination range at 2m is an area with a diameter of 25cm. In this area, there are multiple objects covered by laser illumination, among which object 1 and object 2 are the vibrating sound sources, and object 3 is a common target.

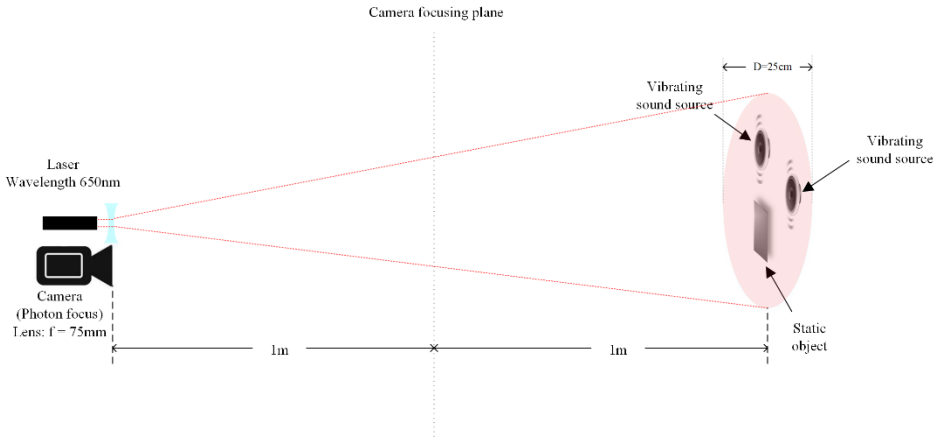

Figure 4.24 Experiment setup of picking up location of vibrating object with diffuse laser illumination

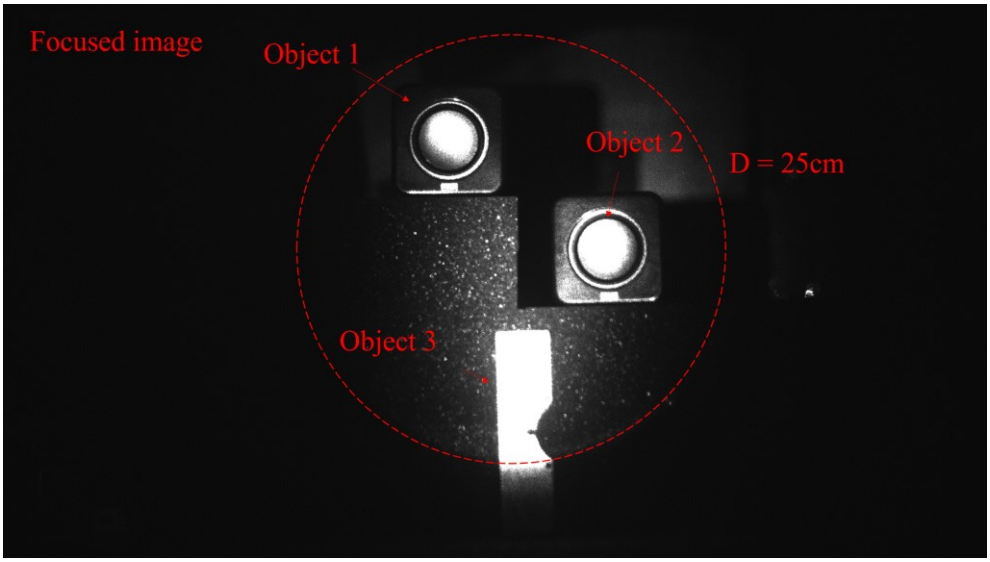

Figure 4.25 A example of expanded laser illumination

Since the laser is reflected by the surfaces of both these three objects, in the defocus camera, the camera captures three speckles, as shown in the Figure 4.26. Nevertheless, the speckle reflected by the surface of a sound source will appear to vibrate due to the vibration of the sound source, whereas the speckle reflected by a normal object will not. Therefore, we only need to calculate the optical flow field of the entire picture, find out where it exists, and then we can distinguish the sound source from ordinary objects. Figure 4.27 shows the binarized optical flow result of the entire image. Although there are three speckle patterns in the image, only two of them shows optical flow, which corresponding to the sound source. Moreover, the result also provides the centroid of each area, which allows us to switch the image to small window size, extracting the original audio signal of the sound source. In this case the audio signals of object 1 and object 2 are both single frequency waves. The frequency of object 1 is 80Hz and the frequency of object 2 is 50Hz. After obtaining the location of the two sound sources in the image, the image size is switched to 32 pixels with frame rate of 4000Hz to real-time restore the audio signal, and the restored waveform and their spectrogram are shown in Figure 4.28.

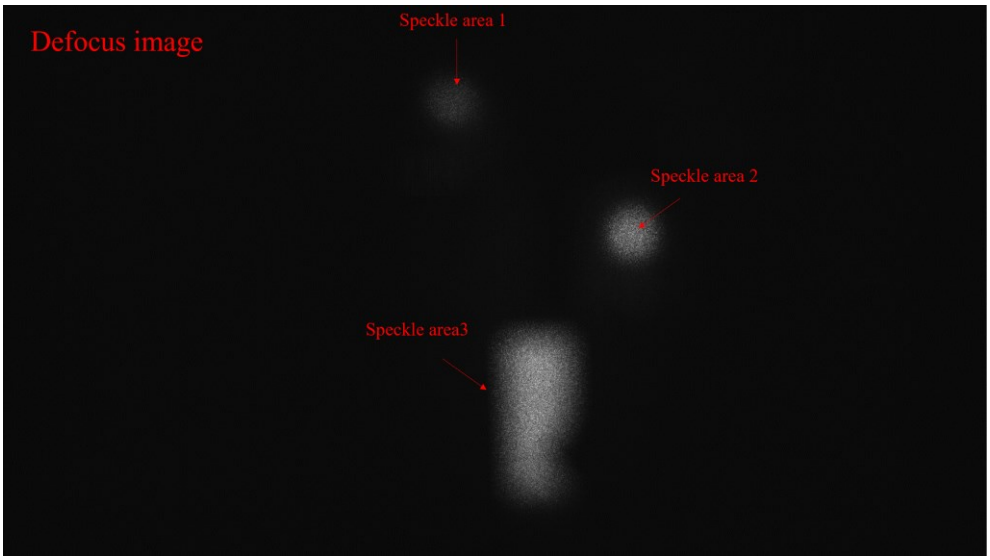

Figure 4.26 Defocused image with expanded laser illumination

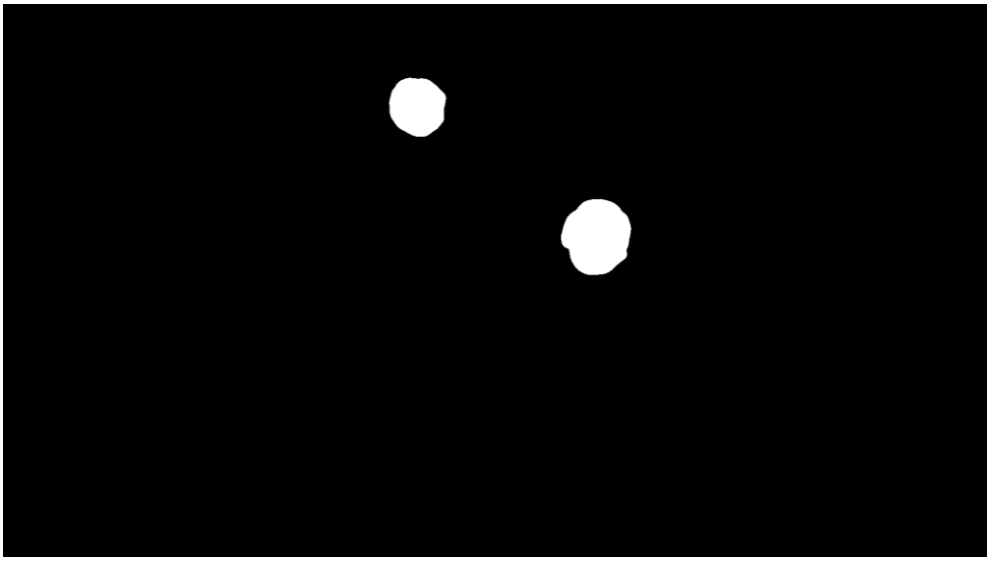

Figure 4.27 Binarized optical flow result of full-resolution images

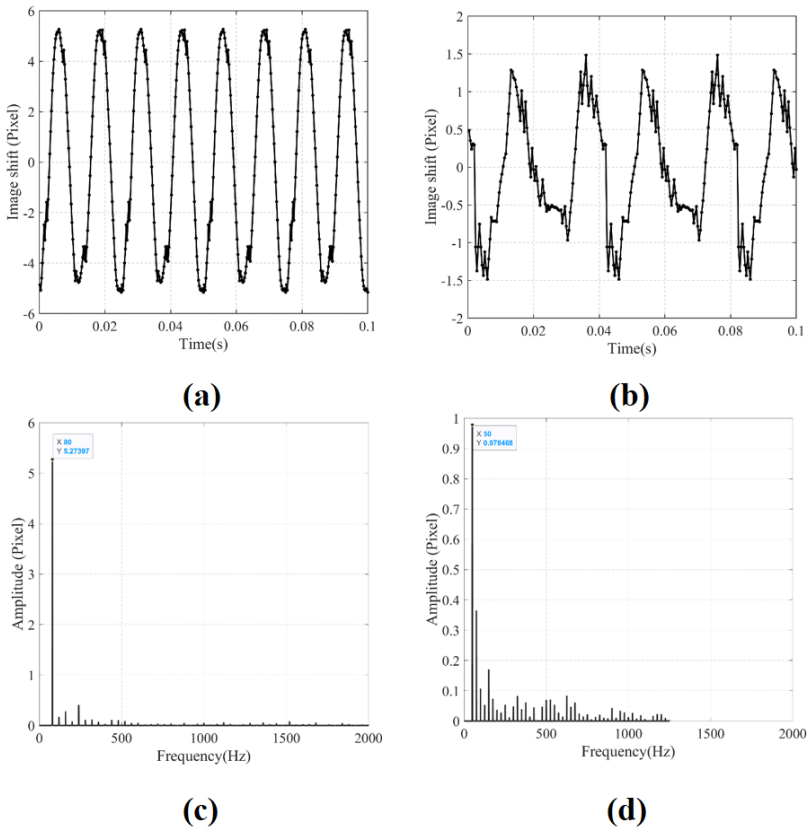

Figure 4.28 Result of restored audio signal with small image window in a full resolution image. (a) Recovered waveform of object 1. (b) Recovered waveform of object 2. (c) Spectrogram of object 1. (d) Spectrogram of object 2.

This experiment proves that using an expanded laser beam, our proposal can cover large range illumination, automatically finding the sound source within the illumination area using speckle's motion and then recovering the audio signal in local image window.

4.3.5 Extract audio signal with long distance

In the next experiment, we tried to use the proposed method to extract and restore audio vibrations at long distance. At the same time, our method was compared with normal electric microphone. We use the experiment result to prove the superiority of the proposed visual microphone in remote extraction of audio signal and extracting audio signal in noise

environment. First, we tried to use the electric microphone to record sound with long distance. As shown in Figure 4.29, the distance between sound source and microphone is 24m, and the target signal is a 100Hz single frequency wave. The result of microphone detection is shown in Figure 4.30. This result illustrates a shortcoming of the electric microphone: if the sound source is far away from microphone, the microphone cannot record the sound signal due to the energy loss in the propagation.

Figure 4.29 Experiment of extracting audio signal at long distance using electric microphone.

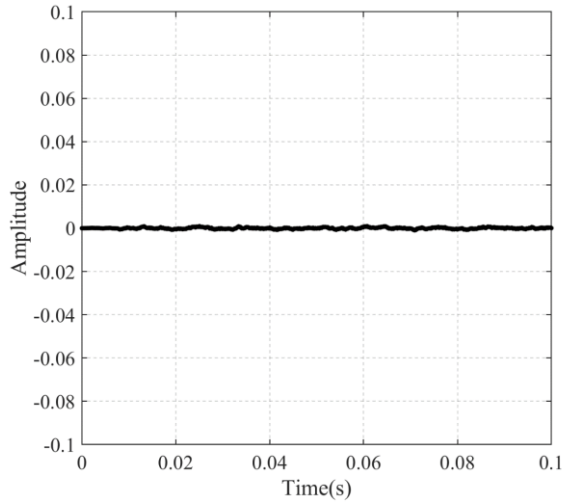

Figure 4.30 The result of the microphone recording the audio signal at a long distance

Another disadvantage of the microphone is that it can only passively perceive audio vibrations, and if there is noise in the environment, the influence of noise on detection cannot be avoided. As shown in Figure 4.31, in this experiment both target sound source and the noise sound source are located 2m away from the microphone. The frequency of the target signal is 100Hz while the noise signal is 200Hz. The result of microphone detection is shown in Figure 4.32. From the result we can see another shortcoming of the electric microphone: since the electric microphone cannot separate different sound sources, the target signal and the noise signal are mixed together in the result, which affects the extraction of the target signal

Figure 4.31 Experiment of extracting target signal in noise environment using microphone.

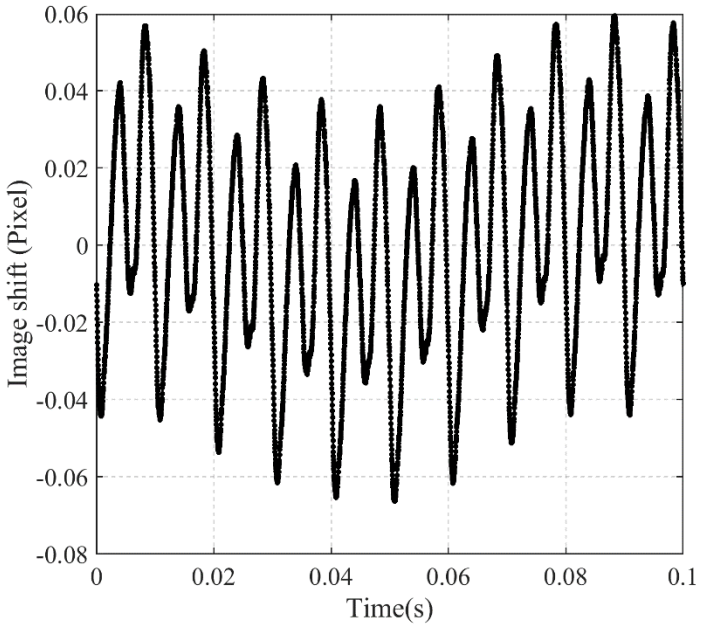

Figure 4.32 The result of the microphone recording the audio signal with noise

On the other hand, using our proposed technology to extract audio signals can avoid the problems associated with electric microphones. As shown in Figure 4.33, in this experiment, the target of interest is again located at 24m away from the detection device (camera), and we set a noise source as interference at 4m away from the detection device. The frequency of the target sound is 100Hz while the frequency of the noise is 200Hz. Here we use a laser pointer (5mW) to illuminate the surface of the vibrating target and use a camera with a $f = 300mm$ lens to capture the moving speckles, then recovering the audio signal from the speckle motion. The restored signal is shown in Figure 4.34. From the result we can see the advantages of detecting audio signal with optical means: the vibrations of distant objects can be visualized visually since the propagation process will not cause energy loss. In addition, because the camera can actively select the target to be detected in the image, it can actively shield the interference from other noises in the environment, making the measurement result more reliable.

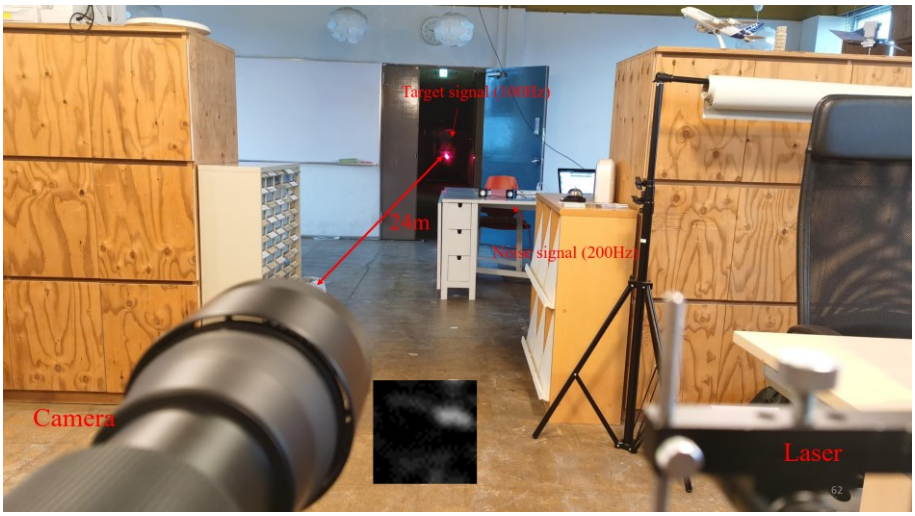

Figure 4.33 Experiment of extracting audio signal at long distance with noise in the environment using laser speckle images.

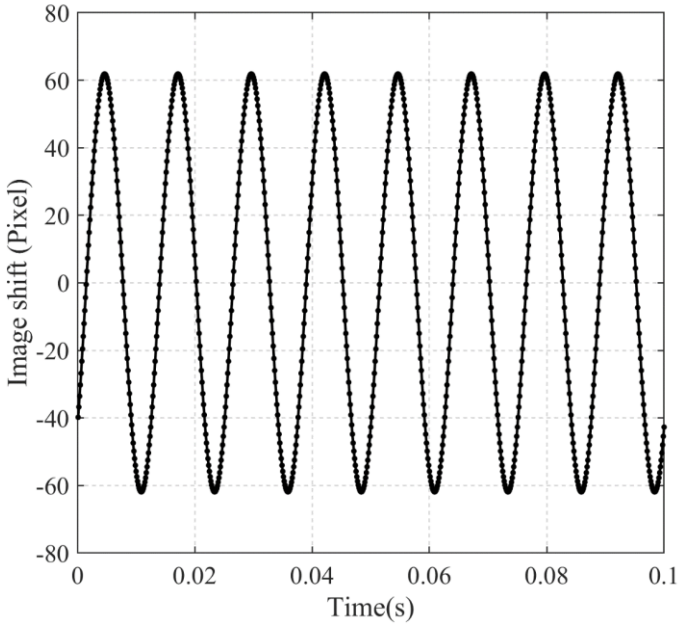

Figure 4.34 The result of the recovered audio signal using speckle images.

4.4 Chapter Conclusion

In this chapter a laser-speckle-based sound detection method has been proposed. In the proposed method, laser speckle image approach is adopted to detect the vibration of sound

source, and optical flow method, along with some denoising algorithms proposed by the authors are employed to fulfill the high accuracy and real time signal processing. The main advantages of the proposed method are real-time audio signal regeneration with high quality and the ability of audio signal regeneration of moving sound source. Since the method uses a camera as a sensing device, it can be combined with expanded laser illumination to detect and position the potential sound source targets in a large range, so that the device can extract audio signals from a large range of targets. In addition, compared to traditional electric microphones, the visual audio signal detection scheme can detect distant sound information, and will not be disturbed by noise in the environment. All these contributions make this technology more widely used in areas, such as laser microphone function. The results of experiments proved that our proposed method is an efficient way to regenerate audio signal under different situations.

The effective real-time sampling rate of the method is around 1kHz, and the result shows it can perform well in the low-frequency region. In the next part, using of faster imaging sensors and optimization of the algorithm will promote the sampling rate of the method so that the human speech can be regenerated with this method.

CHAPTER 5 FAST ACOUSTIC VIBRATION SENSING METHOD USING 1-D SPECKLE IMAGE

Audio signal acquisition using a laser speckle image is an appealing topic since it provides an accurate and non-contact solution for vibration measurement. However, due to the limitation of camera frame rate and image processing speed, previous research in this area could not achieve real time reconstruction of audio signal with high frequency. In this chapter, we use one-dimensional laser speckle image to measure the acoustic vibration of sound source and proposed a fast and sub-pixel accuracy algorithm to estimate the displacement of captured one-dimensional laser speckle images. Compared with previous research, the proposed method is faster and more accurate in displacement estimation. Owing to this, the frequency bandwidth and the robustness are significantly increased. Experiment results show that the proposed method can achieve 20k samples-per-second sampling rate, and the audio signal can be reconstructed with high quality in real time.

5.1 Chapter Introduction

Using optical means to measure the physical vibration caused by sound source is an appealing research topic. The optical vibrometers can accurately measure sub-wavelength mechanical vibrations without contacting objects. These characteristics allow the technology to have broad application prospects in various fields, such as optical microphone, surveillance of audio signal in rescue [78-80], visual accelerometer, and so on.

In our previous research, we used a small image size to improve frame speed of a conventional industrial camera and calculated the optical flow of the captured speckle images. The camera captures over 2300 frames per second with the image size of 32×32 pixels, and the average time consumption is 1 millisecond for optical flow calculation. Compared with the DIC method, we successfully realized the real time reconstruction of audio signals with the frequency up to 500 Hz. This approach has greatly improved the sampling speed of the laser-speckle measurement method. However, due to the limitations of both conventional area-scan camera's frame rate and image processing speed, it is still difficult to regenerate high frequency audio signals in real time. For example, the requirement for sampling human speech is at least 8 kHz. The real time sampling rate of previous methods cannot satisfy this requirement [81].

In this chapter, we propose using a line-scan camera to further improve the sampling speed of the sensing method. The line-scan camera usually can capture tens of thousands of 1D images per second, which greatly improves the detectable frequency range. Besides, using 1D images also brings the reduction of the computational cost of image processing. To estimate the motion of captured 1D images, a fast and accurate displacement estimation algorithm is proposed. The algorithm calculates 1D optical flow and estimates the global displacement of the captured speckle images. The results show that the proposed algorithm can achieve a microseconds level's calculation speed and an accuracy with the average error less than 0.03 pixels. Audio recovering experiment results show high frequency audio signals, such as human speech, can be detected and restored in real time with high quality by our method.

5.2 Materials and Methods

5.2.1 Speckle sensing model

The sketch of our proposed method is shown in Figure 5.1. An infrared laser beam ($\lambda = 908nm, power = 50mW$) illuminates on the surface of a speaker. Although the quantum efficiency of the image sensor is low at the wavelength of 980 nm, the laser light is also invisible to human eyes. The optical microphone using such a light source does not emit annoying light, which makes it suitable for commercial application. The scattered speckle pattern is captured by a line-scan camera (Photonfocus MV1-L2048-96-G2, Lachen, Switzerland) with a lens (Avenir TV zoom lens, $f =$ 75 mm, Tokyo, Japan). In our method, the camera lens is focused on the plane away from the object determined by distance L_1. The camera captures 1D laser speckle images with a frame rate of 20 kHz. The captured speckle images are processed by a common desktop PC (AMD Ryzen 5 3600 processor,3.60 GHz, Sunnyvale, America) to output the restored audio signal.

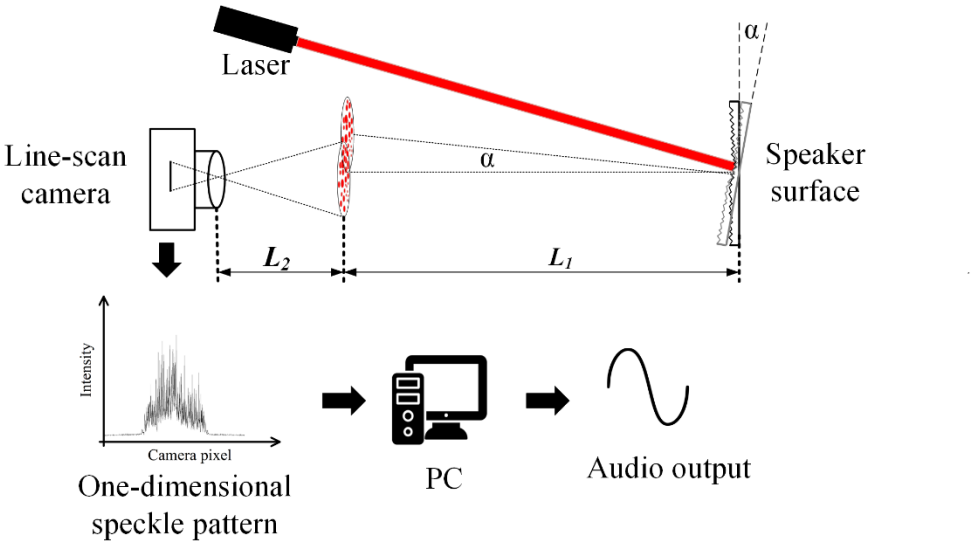

Figure 5.1 Schematic of the real time audio signal detection method using 1D laser speckle images

In the method, the vibration of the speaker surface can be regarded as a rigid body motion. Generally, the rigid body motion has three motion modes: in-plane transverse, axial motion, and tilting. The explanation in chapter 2 has proved that the displacement of the speckle pattern

is mainly determined by the tilting of the object if the camera is strongly defocused. Thus, there is a linear relationship between the image shift and the tilting angle of the object. When the object vibrates, the speckle pattern shows linear vibration at the image sensor. This means it is possible to use a line-scan camera to restore the vibration information of the object, such as the amplitude and the frequency of the vibration, by analyzing the motion of the captured speckle images.

5.2.2 Explanation of the 1D optical flow algorithm for displacement estimation

In our method, the camera's sampling rate is 20 kHz under line-scan mode. For such a high frame rate, there are two requirements for the real time motion estimation algorithm. First, the algorithm's calculation speed must be fast enough to catch up with the camera's frame rate. Besides, the displacement between two adjacent images is usually less than one pixel. The proposed algorithm must have sub-pixel accuracy.

Based on these considerations, a displacement estimation algorithm using 1D optical flow is proposed. Figure 5.2 shows the signal model of the 1D image, where image intensity is regarded as a quadratic function $f(x)$ of the pixel location. For example, for pixel x in image $f_1(x)$, the local signal model can be expressed by doing quadratic polynomial fitting using the neighborhood of pixel x, which is:

$$f_1(x) = a_1x^2 + b_1x + c_1 \qquad (5-1)$$

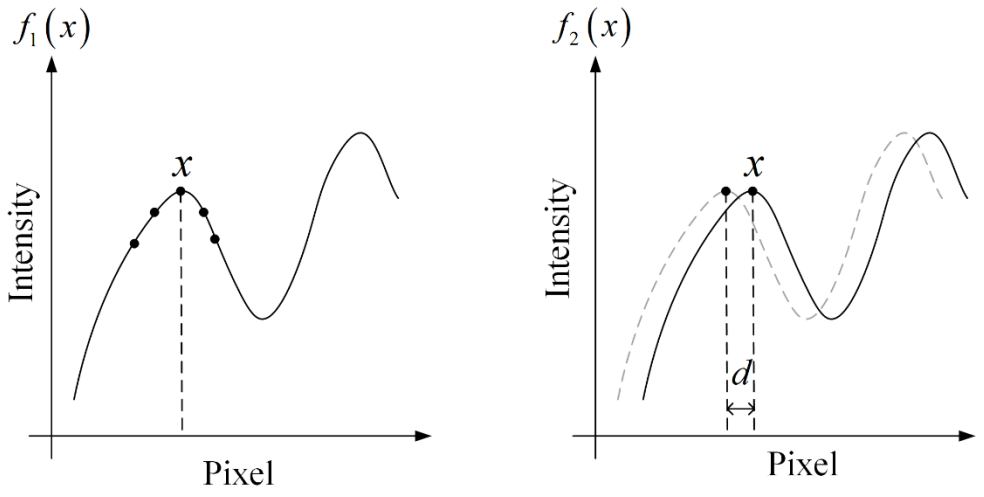

Figure 5.2 Local signal model of the 1D image undergoes a small shift.

Assuming the displacement is d, by doing polynomial fitting, we have the local signal model $f_2(x)$ at the corresponding pixel in the shifted image, which is:

$$\begin{aligned} f_2(x) &= f_1(x-d) \\ &= a_1(x-d)^2 + b_1(x-d) + c_1 \\ &= a_1x^2 + (b_1 - 2a_1d)x + a_1d^2 - b_1d + c_1 \end{aligned} \quad (5-2)$$

In this way, the polynomial coefficients of the shifted image are connected to the ones from the original image. Generally, the shift amount is very small between two captured frames because of the high sampling rate of the camera. Assuming the intensity is constant in two images, the coefficients in the two polynomials are equal:

$$a_1 = a_2 \quad (5-3)$$

$$b_1 - 2a_1d = b_2 \quad (5-4)$$

$$a_1d^2 - b_1d + c_1 = c_2 \quad (5-5)$$

By solving Equation (5-4), we can determine the value of d, which is:

$$d = -\frac{1}{2a}(b_2 - b_1) \qquad (5-6)$$

where $a = \frac{a_1+a_2}{2}$.

Repeating this process, the displacement of all pixels can be calculated. In other words, we can obtain a dense optical flow between two 1D images. However, due to the image noise and the deviation of quadratic fitting, pixelwise solutions of Equation (5-6) do not give good results. The values of displacement at each pixel are very different. On the other hand, a known fact is that, for rigid body, the small object motion does not change the speckle pattern but only shifts the pattern by a small amount [82, 83]. Each pixel's motion on the speckle should be the same. Therefore, in our proposed algorithm, we calculate the global displacement of the speckle images based on the obtained dense optical flow.

Rewriting Equation (5-6), let $\Delta b = -(b_2 - b_1)/2$. At each pixel, we have:

$$a \times d = \Delta b \qquad (5-7)$$

Assuming the captured image has N pixels, we can calculate the array of coefficients a_1, b_1 from $f_1(x)$ and a_2, b_2 from $f_2(x)$ for each pixel. In this way, we can build N equations as Equation (5-7). Here, we try to find d satisfying N equations as well as possible. In other words, our goal is minimizing the following function:

$$e = \sum_{i=1}^{N} |a_i \times d - \Delta b_i| \qquad (5-8)$$

In conclusion, the actual calculation process in the program is as follows. The two input images are preprocessed to reduce the high-frequency noise. After that, quadratic polynomial fitting is performed on the two images pointwise to obtain the coefficients. Then, the coefficients arrays $\mathbf{a}$ and $\Delta\mathbf{b}$ is obtained. Finally, the least square solution of displacement

d is determined, which can be expressed as:

$$d = \frac{\sum_{i=1}^{N} a_i \Delta b_i}{\sum_{i=1}^{N} a_i^2} \qquad (5-9)$$

5.3 Experiment Results

5.3.1 Performance of 1D optical flow algorithm

Before conducting experiments of audio signal detection, the computational accuracy, and the speed of the proposed 1D optical flow algorithm are investigated. In the experiment, the speaker was used as the test object. The material of the speaker surface was resin, and the speaker surface was diffusive to laser beam. Here, we illuminated the infrared laser on the center of the speaker surface. The diameter of the laser spot on the speaker surface was 6 mm. Due to the diffusive surface of the speaker, multiple lights were scattered randomly, and a high-contrast speckle pattern could be observed by camera because of the interferometry of the reflection lights. Figure 5.3(a) shows a captured 2D speckle pattern with the size of 512×512 pixels, whereas Figure 5.3(b) shows the captured 1D speckle pattern by the line-scan sensor. Here, we converted the 2D image to the frequency domain and added a linear phase to manually produce a known subpixel shift along the horizontal direction [73]. Then, an image template with a size of 1×500 pixels from the origin image and the shifted image were selected to do the displacement estimation. Since the intensity distribution of speckle pattern was stochastic, the average error on 200 random templates was carried out to improve the accuracy of error estimation.

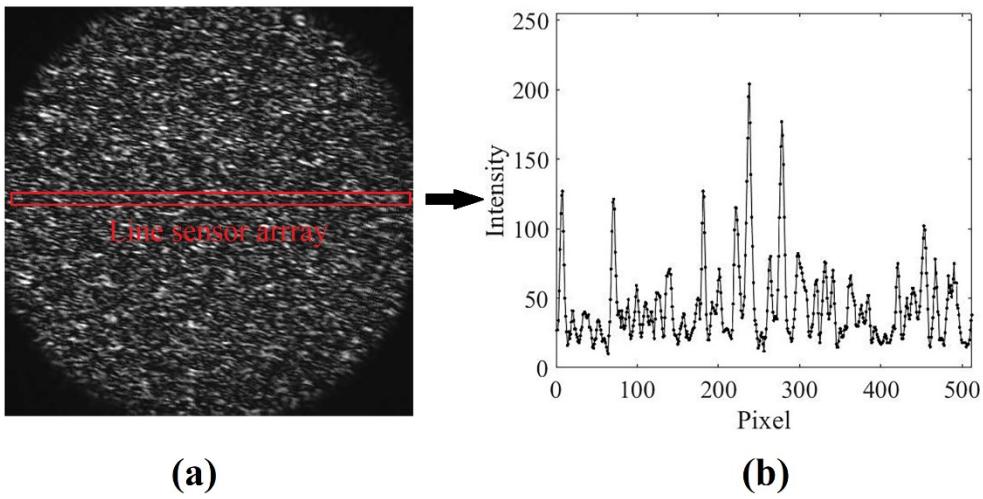

Figure 5.3 Captured speckle images; the shutter time was 20 μs. (a) Captured two-dimensional speckle pattern with the size of 512 × 512 pixels. (b) Captured 1D speckle pattern with the size of 1 × 500 pixels.

First, the relationship between the number of pixels n used in quadratic polynomial fitting and the accuracy of displacement estimation was investigated. The proposed algorithm performs quadratic polynomial fitting on the local 1D speckle image, thus the optimal number of pixels used in polynomial fitting is dependent on the average size of the speckle. If the number is too large or too small n reduces the fitting accuracy and results in bad results of displacement estimation. Here, we manually shifted the image from 0.05 pixels to 0.95 pixels with a step of 0.05 pixels. Figure 5.4 presents the dependence of the average error on shift amount for different polynomial kernel size n. The results showed that polynomial fitting using 5 pixels provided the best accuracy on displacement estimation. If the polynomial kernel size was small (e.g., $n = 3$), the result showed the largest error on displacement estimation. This is because the kernel size was too small to provide accurate polynomial fitting of a local signal. On the other hand, if the kernel size is larger than the speckle size, spatial aliasing of speckles

within the polynomial kernel could be an issue, which also affects the accuracy of polynomial fitting. One can see that, when the kernel size was larger than 5 pixels, the accuracy of displacement estimation decreased as the kernel size increased. Speckle size is determined by many factors, such as the wavelength and the diameter of the laser beam, the roughness of the surface, and the defocusing of the camera. In our situation, the 5 pixels polynomial fitting was adopted for further tests since it showed the best performance on displacement estimation.

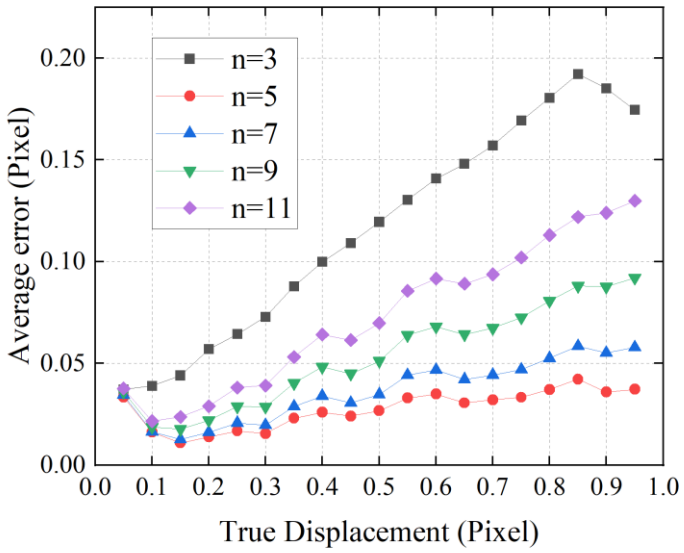

Figure 5.4 Dependence of the average error result on the shift amount for different quadratic polynomial fitting kernel size

In the next experiment, the displacement estimation accuracy using a different image template size was investigated. As a comparison, the performance of a well-used cross- correlation algorithm was also investigated. This algorithm uses cross-correlation in Fourier space and employs an up-sampled matrix-multiplication DFT to achieve arbitrary subpixel precision [84]. Here, we set the subpixel precision to 100, which meant the resolution of the cross-correlation algorithm was 0.01 pixel. Figure 5.5 shows dependence of the average error of displacement

estimation on the template size for the two algorithms. From the results, we can see that like the situation of 2D image, the 1D optical flow algorithm gave better results than the cross-correlation method. Moreover, it should also be noted that the error of cross-correlation algorithm gradually increased as the template size decreased. When the template size was 150 pixels, the accuracy of displacement estimation was not stable. On the other hand, the proposed 1D optical flow algorithm always gave good calculation accuracy under different template sizes.

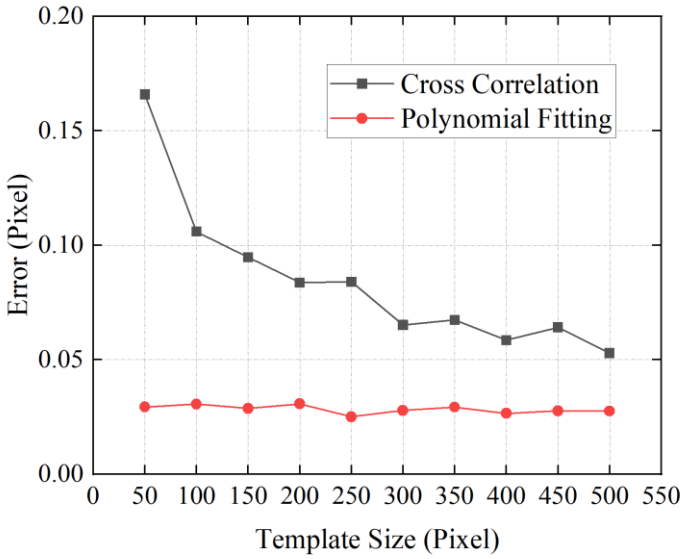

Figure 5.5 Dependence of the average error of displacement estimation on the template size for different algorithms

In our method, line sensor is used to observe the global shift of speckle image, and the direction of image vibration is supposed to correspond with the sensor array. However, in actual situations, it is difficult to perfectly match the directions of the line sensor and the speckle motion. The angle between the two directions may result in noise of the displacement estimation. Therefore, in the next experiment, we tested the performance of the two algorithms when the speckle shift was inconsistent with the line sensor array. We used the same method to

manually shift the image with a known amount along different directions. Then, we used a set of image templates with the size of 1×500 pixels to calculate the shift component in the horizontal direction and obtain the average error. Figure 5.6 presents the dependence of the average error on the angle for the two algorithms. It can be seen from the result that the inconsistency between the line sensor array and the vibration direction of the speckle image resulted in increasing errors in displacement estimation. Especially for the cross-correlation algorithm, the average error rapidly increased as the angle increased. However, the calculation by the 1D optical flow algorithm showed better robustness than the cross-correlation method. The average error was less than 0.05 pixels when the angle was less than 45°. In an actual situation, the proposed 1D optical flow algorithm is more robust in estimating the horizontal displacement when the image does not vibrate along the line sensor array.

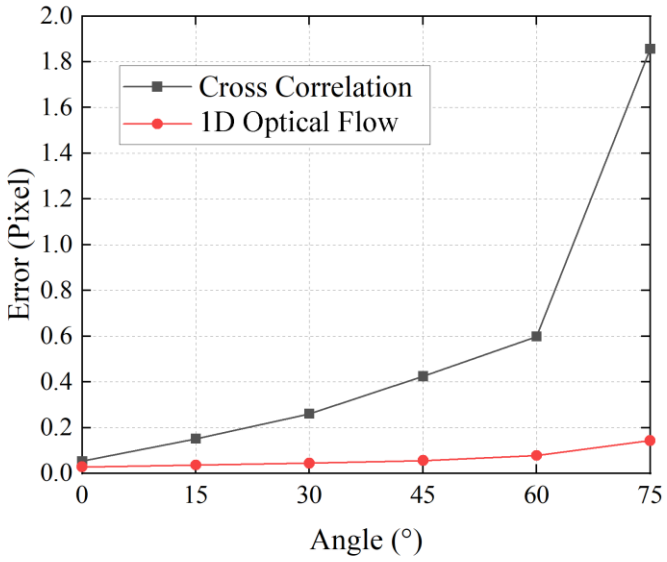

Figure 5.6 Dependence of the average error of displacement estimation on the angle for different algorithms

Last, we investigated the computational speed of the two algorithms. Here, 20,000 times

displacement calculations were conducted for both the 1D optical flow algorithm and the cross-correlation algorithm, and the average time consumption with different template size is shown in Figure 5.7(a). It can be seen from Figure 5.7(a) that using smaller size templates reduced the computational load and achieved a higher calculation speed. The test results also showed that the proposed 1D optical flow algorithm was much faster than the cross-correlation algorithm. Especially when the template size was 50 pixels, as shown in Figure 5.7(b), the 1D optical flow algorithm took only $41\mu s$ for displacement estimation, whereas the cross-correlation method required $205\mu s$. This meant that the calculation speed of the proposed algorithm matched the sampling speed of the camera to achieve a real time acquisition and processing rate of 20kHz. According to the Nyquist sampling theorem, our method can sample the signal frequency up to 10kHz in real time. This sampling speed can satisfy the requirement of human speech sampling.

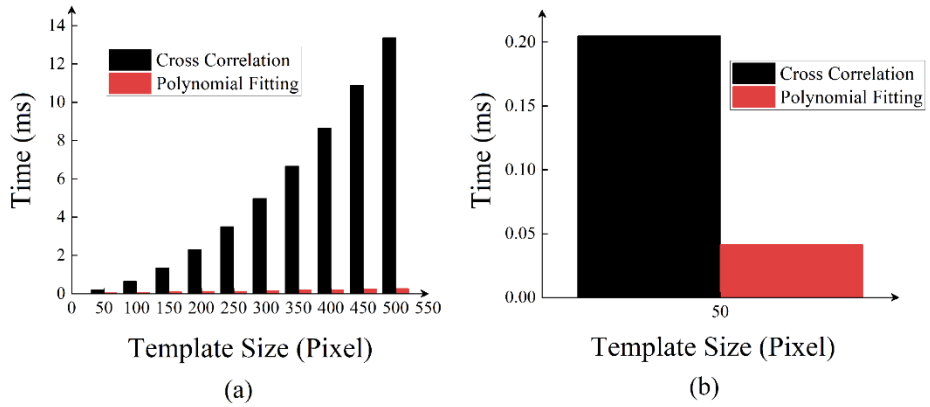

Figure 5.7 Comparison of average time consumption of two algorithms. (a) Dependence of time consumption on template size. (b) Time consumption of two algorithms with a template size of 1 × 50 pixels

5.3.2 Result of Real time audio signal extraction

In the next experiment, we tried to use our method to extract a single frequency audio signal

of a speaker. Before conducting the experiment, we first investigated the frequency response of the speaker. We made a sound-absorbing box using sound-absorbing materials and put the speaker and the microphone in it to make the result as accurate as possible. The speaker played an audio signal with increasing frequencies from 20 Hz to 20k Hz. The recorded signal was analyzed to obtain the efficiency of the speaker for different frequencies, as shown in Figure 5.8.

Here, we used a region of 1×50 pixels of the line-scan camera. The line-rate was 20,000 frames per second, and the shutter time was $20\mu s$. The speaker was positioned 2m away from the camera. The captured images were processed in real time, and the results were exported. First, we sent the signal with constant frequencies in the speaker. Figure 5.9(a) presents the first 20ms waveform of the reconstructed signal with the frequency of 100 Hz to 500 Hz. The waveform was the calculated shift of the observed speckle motion on the line sensor. Figure 5.9(b) shows the spectrogram of the results. The results matched the signal that was sent. It should be noted that the amplitude of the 100Hz regenerated waveform was smaller than other signals, because the frequency response of the speaker was low at 100Hz frequency. The performance of our method was also investigated using the signal-to-noise ratio (SNR) and the total harmonic distortion (THD). The results obtained with increasing frequencies from 100Hz to 1000Hz are shown in Figure 5.10, which prove that the sinusoidal audio signal could be restored with high quality using our method.

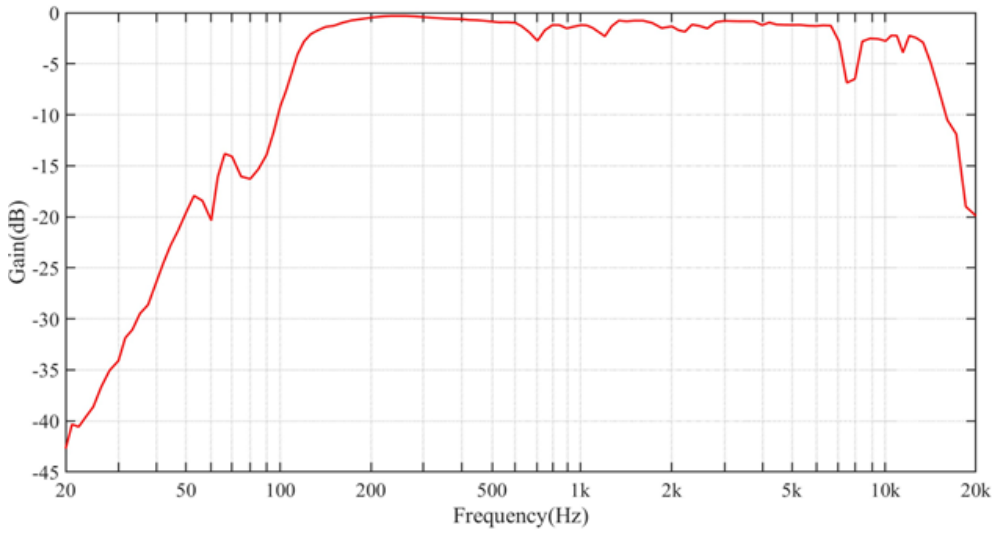

Figure 5.8 Frequency response of the speaker used in the experiments

In the next experiment, we tried to extract a signal with high frequency. This time, the 5kHz sinusoidal signal was played by the loudspeaker. We used the camera to take 10,000 frames and simultaneously analyzed the displacement between captured frames to restore the waveform of the signal. The result of the restored temporal signal is presented in Figure 5.11(a). In Figure 5.11(b), we present the spectrogram of the reconstructed signal. The results prove that the high-frequency vibration of the object could be observed with a speckle image of 1×50 pixels, and our algorithm restored the frequency and the amplitude information of the high-frequency vibration correctly in real time. The SNR was 12.64dB for the result. It should be noted that the restored high frequency signal had a lower signal- to-noise ratio compared with the low-frequency signal. This is because the amplitude was low when the speaker vibrated at high frequency, which resulted in the reduction of the SNR.

Finally, we present the result of human speech extraction with our method. Figure 5.12(a) shows the temporal signal of the original sound, which was the voice of a male counting from zero to nine in English. Figure 5.12(b) shows the temporal signal of the restored sound. Figure 5.12(c)(d) show the spectrograms of the original and the restored audio signals. The experiment

showed that the contents of human voice could be regenerated clearly in real time owing to the high sampling rate of the proposed method. For reference, the audio files of both original music and regenerated music are provided as the result of this experiment.

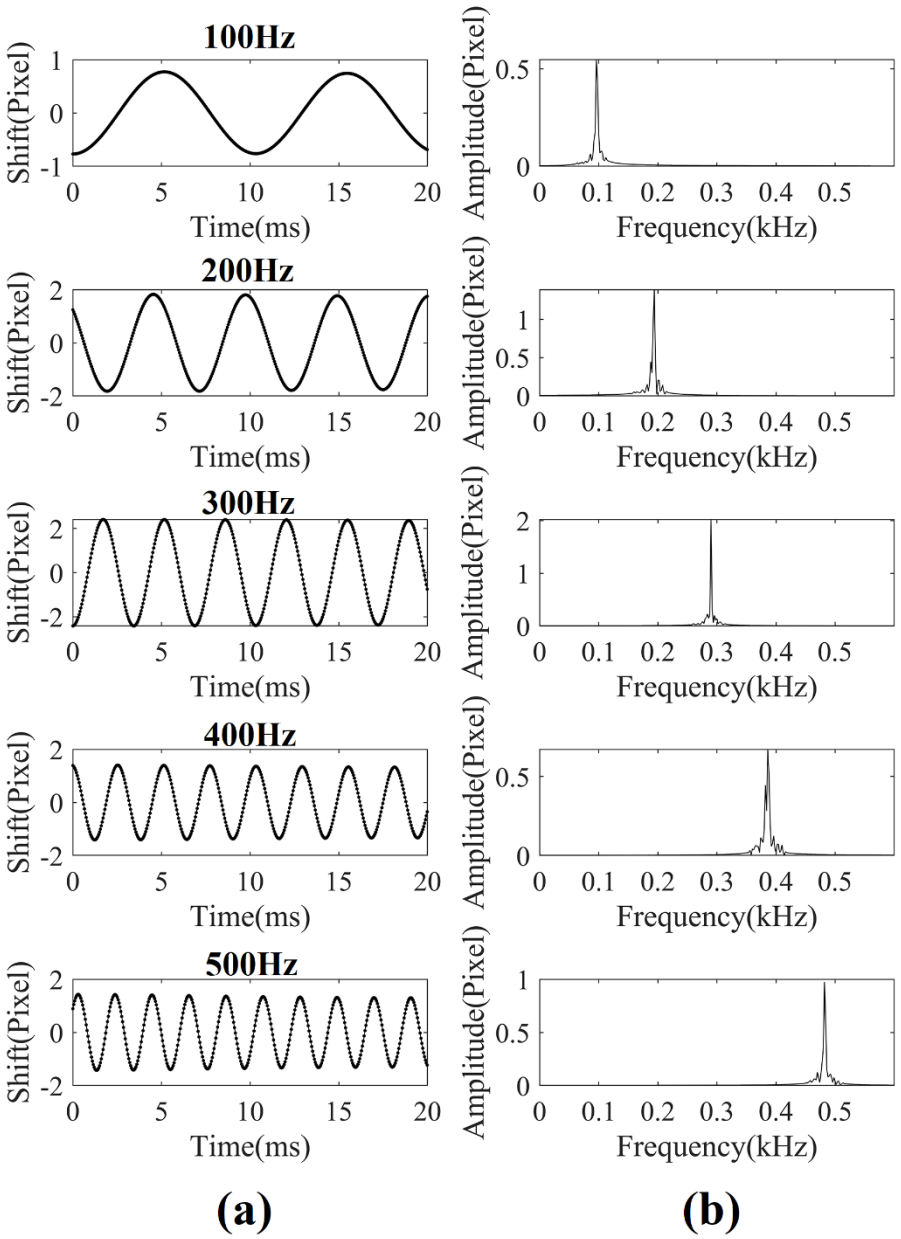

Figure 5.9 (a) Restored waveforms of audio signals at different frequencies. (b) Spectrogram of restored signals at different frequencies

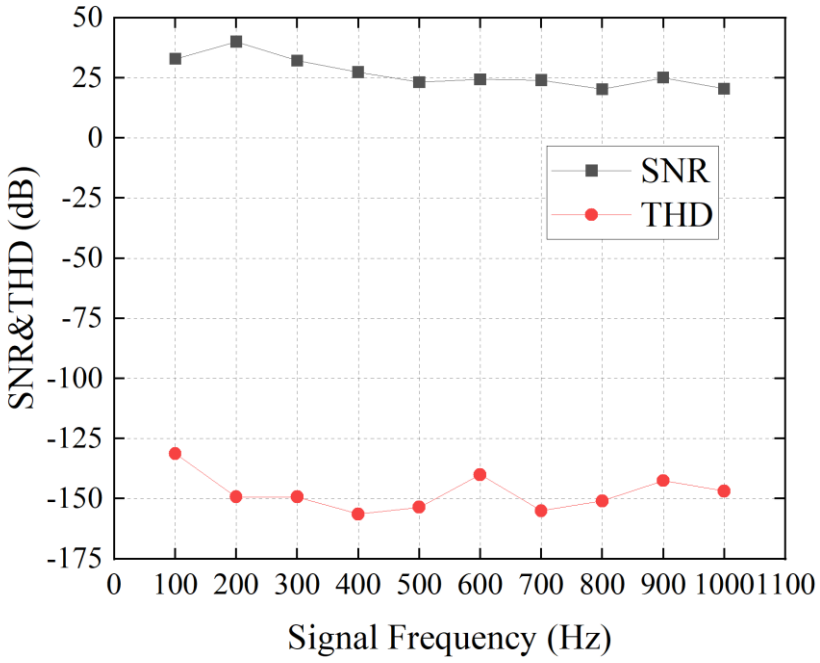

Figure 5.10 SNR and THD of the restored signal with different signal frequency

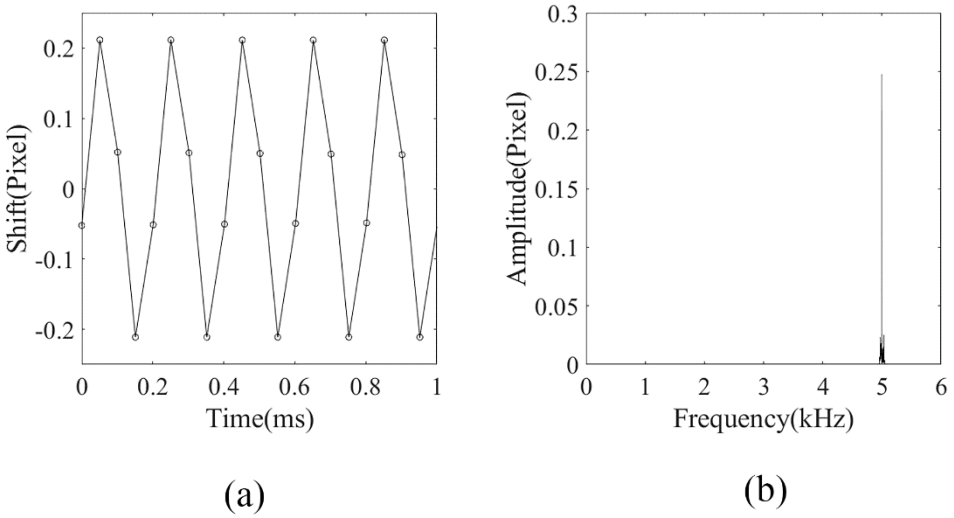

Figure 5.11 (a) The first 1ms of restored temporal signal; the original signal frequency was 5 kHz

(b) Spectrogram of the restored signal.

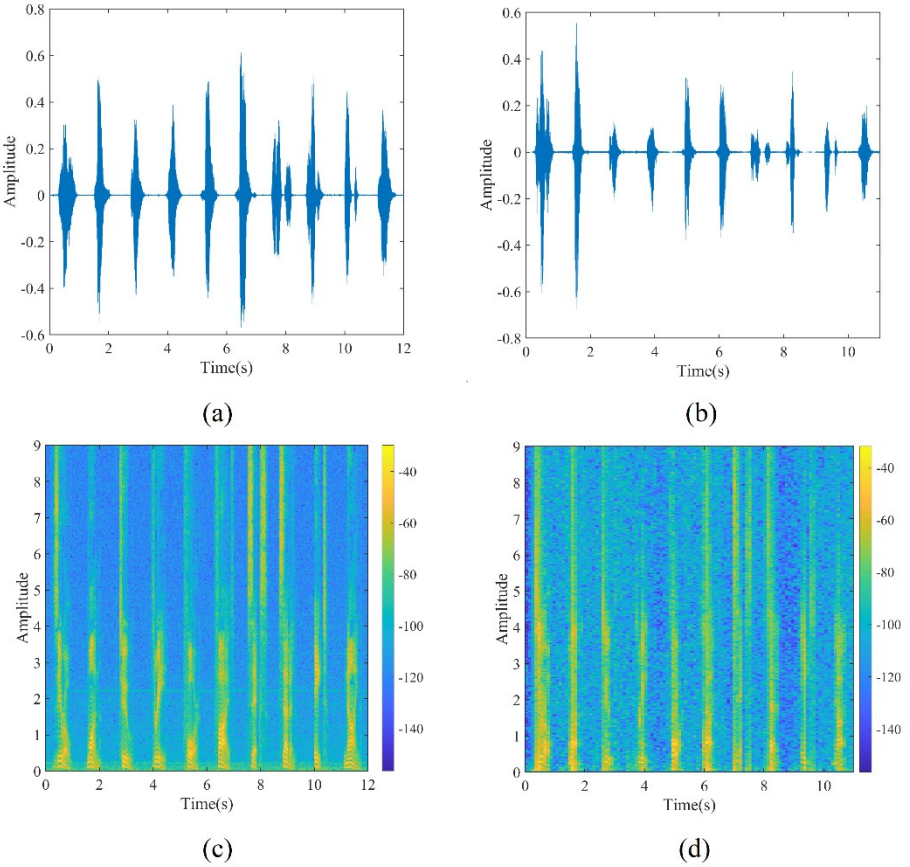

Figure 5.12 Experiment result of restoring human speech in real time. (a, b) Temporal signal of the original sound and the restored sound. (c, d) Spectrogram of the original and the restored sound.

5.3.3 Further Discussion

In 5.2, we provide an explanation of the laser speckle sensing model. In our method, a line-scan camera is adopted to sense the speckle motion. This brings increasing frequency bandwidth. However, one disadvantage of this approach is that a line-scan sensor can only observe speckle motion in one-dimensional format, whereas the speckle vibration is a two-dimensional motion. Usually, the speckle image vibration direction is not perfectly consistent with the line sensor array. Here, the experiment was conducted to investigate the performance of audio signal recovery using our algorithm when the speckle did not move along the line

sensor. For comparison, the performance of the cross-correlation algorithm was also investigated.

To obtain dynamic speckle patterns with different vibration directions, the speaker playing an audio signal with a frequency of 100 Hz was fixed on a rotation stage. By changing the angle of the rotation stage, we could control the vibration direction of the captured 2D speckle patterns. Figure 5.13 shows the trajectories of the captured 2D dynamic speckle patterns with different rotation angles. The image shift directions could be estimated from the trajectories, which were expressed as the absolute value of the angle (acute angle) between the speckle motion direction and the horizontal direction. Table 5.1 and Figure 5.14 show the relationship between stage angles and the 2D image shift directions. From the result, we can see that, when the stage angle was 60°, the image shift direction was almost consistent with the line-scan sensor's direction. As the angle of the rotation stage increased, the image shift direction also changed linearly. When the stage angle was 150°, the 2D image shift direction was almost orthogonal to the line sensor array.

Table 5.1 Relationship between rotation stage angle and the 2D image shift angle.

Stage angle (°)	60	75	90	105	120	135	150
2D image shift angle (°)	6.6577	23.7615	34.2741	49.9582	60.7933	72.3964	82.8111

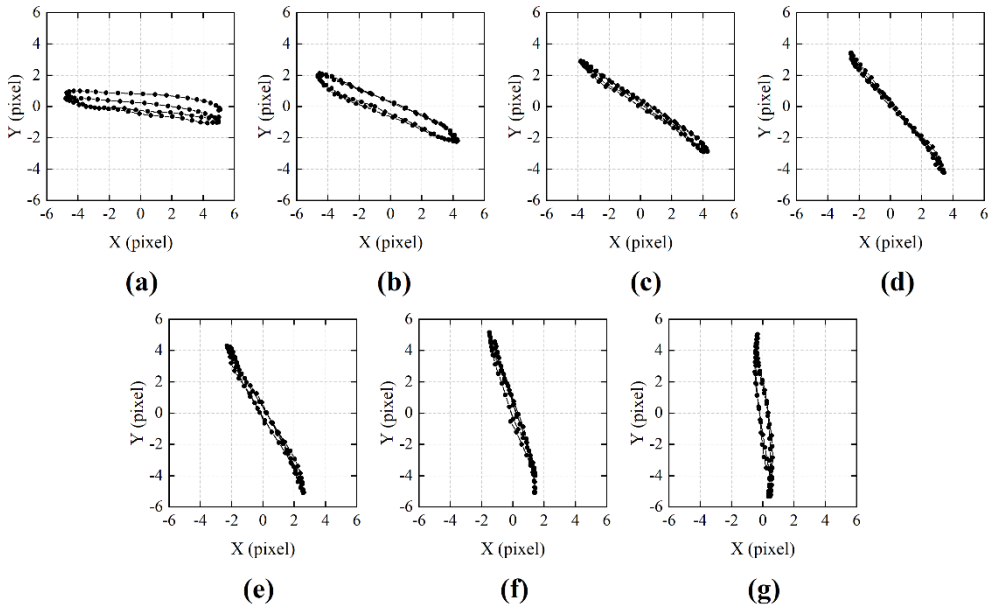

Figure 5.13 Trajectories of the 2D dynamic speckle patterns with different rotation stage angles. (a) Stage angle was 60°. (b) Stage angle was 75°. (c) Stage angle was 90°. (d) Stage angle was 105°. (e) Stage angle was 120°. (f) Stage angle was 135°. (g) Stage angle was 150°.

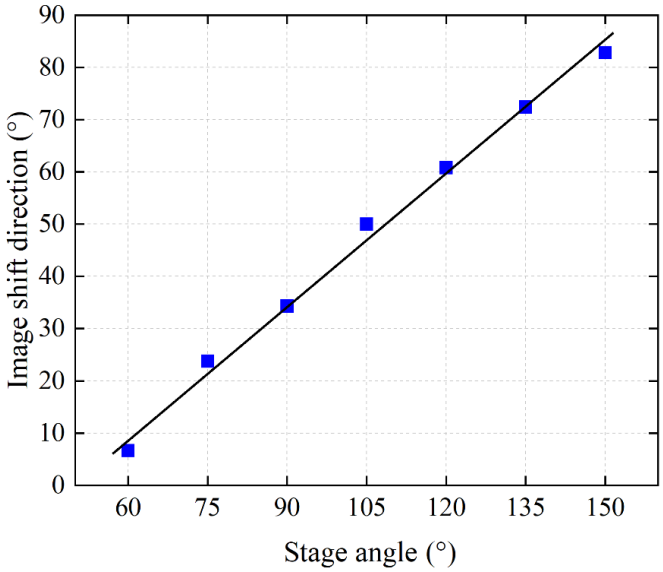

Figure 5.14 Dependence of the 2D speckle image shift angle and the rotation stage angle

Next, the camera was switched to line-scan mode to conduct audio signal recovering tests. The captured speckle images were analyzed both with our algorithm and the cross- correlation algorithm to restore the audio signal. Figure 5.15 shows the SNR results of both algorithms with different situations. Figure 5.16 shows the waveform of the restored signal. From the results, we can see that, when the angle between the 2D image movement direction and the line sensor was small, both algorithms performed well and restored the audio signal with high quality, as shown in Figure 5.16(a)(d). As the angle increased, the cross-correlation algorithm could not give stable results. The SNR dropped rapidly, and the restored waveform was distorted, as shown in Figure 5.16(e). On the other hand, our algorithm showed a better result than the cross-correlation method. As shown in Figure 5.16(b), when the image shift angle was 49.9582°, our algorithm still calculated the shift amount in the horizontal direction correctly and restored the information of the audio signal with high quality. When the image shift direction was orthogonal to the line sensor's direction, neither algorithm could restore the audio signal, as shown in Figure 5.16(c)(f). The experiment results showed that our algorithm had stronger robustness for the disadvantage that the line sensor can only observe the speckle movement in one dimension. Even when the speckle motion direction was not consistent with line sensor array, our algorithm could still restore high-quality audio signals within a certain range by calculating the horizontal motion component.

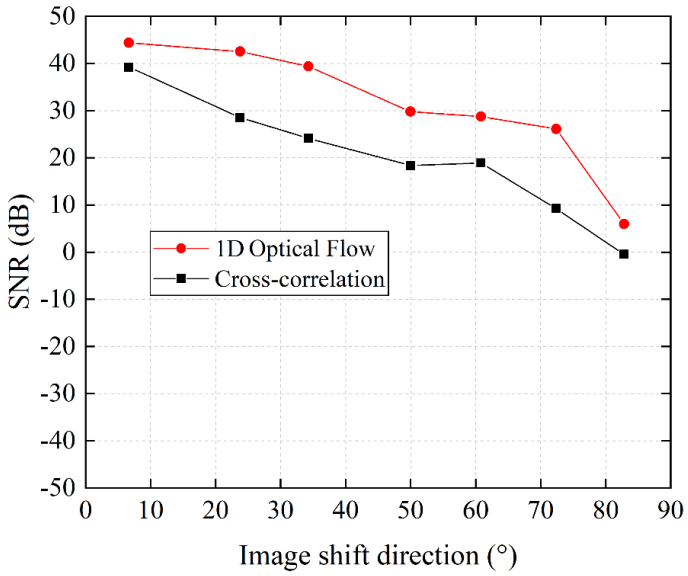

Figure 5.15 Dependence of the SNR of the restored audio signal on the 2D image shift angle for different algorithms.

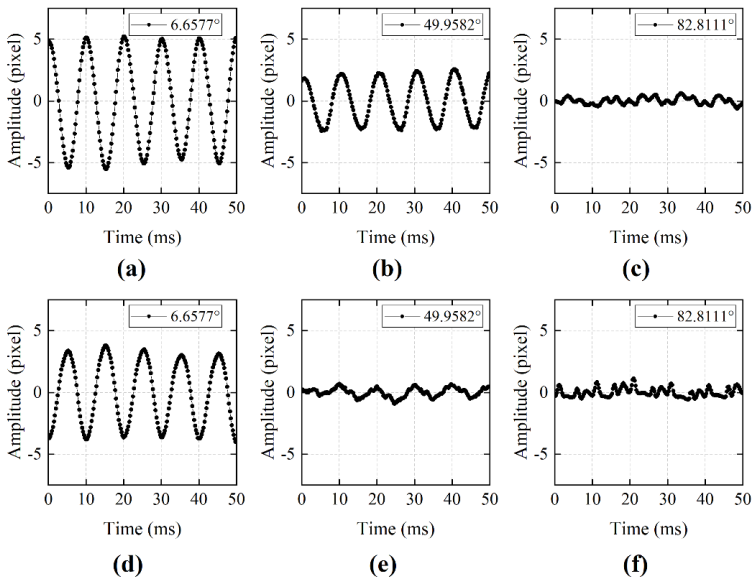

Figure 5.16 Restored waveform with different 2D image shift angle using different algorithms. (a–c) Restored waveform using 1D optical flow algorithm. (d–f) Restored waveform using cross-correlation algorithm

5.4 Chapter Conclusion

In this chapter, we proposed a fast motion estimation of a 1D laser speckle image and showed its application on a real time audio signal detection method. The major contribution of our work is the fast displacement estimation algorithm for 1D speckle images. Owing to this, the proposed method can achieve a 20k Hz real time sampling speed. Experimental results presented the capabilities of the proposed method for extraction of high frequency audio signals and human voice. The high speed, real time sampling method discussed in this chapter has broad application prospects, such as voice signal acquisition and high frequency vibration monitoring of industrial equipment.

CHAPTER 6 CONCLUSION

6.1 Thesis conclusion

In this thesis, we presented a real-time sound vibration detection and recovering method based on high-speed image processing of laser speckle images. In the proposed method, laser speckle images are used to measure micro vibrations of objects. Through high-speed image processing of the captured images, the method can achieve real-time regeneration of audio signals. The major contribution of this research is as follows:

1. Investigated the basic optical principles of speckle sensing method. Speckle measurements involve several parameters, such as laser parameters, camera parameters, which together drive various speckle image properties such as intensity, average pattern size, etc. A significant influence on the speckle image characteristics and motion characteristics is the degree of defocus of the imaging method. One work of this research is investigating relationship between those parameters so that the object vibration can be clearly observed through speckle motion

2. In order to restore the audio signal in real time, it is crucial to calculate the displacement of the speckle image at a high speed. The most direct and effective way to increase the image processing rate is to reduce the image size, but this could make image processing more challenging. This study also explores and analyzes different image processing algorithms. Based on analysis and experimental verification, it is determined that the optical flow method can be highly efficient and accurate for analyzing image displacement using only a small image size.

3. One original work of our research is proposing a 2D laser-speckle-based sound detection method. The main advantages of the proposed method are real-time audio signal regeneration with high quality and the ability of audio signal regeneration of moving sound source, which have never been realized before.

4. Another original work of the research is that we proposed a fast motion estimation of a 1D laser speckle image and showed its application on a real time audio signal detection method. The major contribution of our work is the fast displacement estimation algorithm for 1D speckle images. Owing to this, the proposed method can achieve a 20k Hz real time sampling speed. Experimental results presented the capabilities of the proposed method for extraction of high frequency audio signals and human voice. The high speed, real time sampling method has broad application prospects, such as voice signal acquisition and high frequency vibration monitoring of industrial equipment.

In this research, a real-time acoustic vibration detection and recovery method by performing high-speed image processing on laser speckle images has been proposed. The proposed method can real-time simultaneously detect sound vibration of object and recover the audio signal, making it suitable for applications such as visual microphones and vibration monitoring/reporting methods.

6.2 Future work

In this thesis, experiments have been carried out to prove that our proposal can solve the problems of electric microphones and have potential applications in situation such as conference and sports game broadcasting. Still, many work have to be done in the future.

For example, for conference application, we proved that using our proposal (expanded laser

+ camera), we can quickly find the location of sound source, extract audio signal from the sound source. But due to the low power density of the expanded laser beam, the quality of the recovered signal is not good. Increasing the laser power is one solution but this will also cause the laser safety problem. So, in the future, we will try to solve this problem by using other sensors that is more sensitive to the laser light.

For the long-distance sound extraction. We proved that using our proposal (laser pointer + camera), we can extract audio signal from target with long distance, without being affected by surrounding sound. Bur in the experiment our test object is a speaker. In the future we will try to recover sound from human skin. Actually, recovering sound from human skin is possible.

Here experiment has been conducted to prove this. In this experiment we tried to use our experiment setup to recover sound from human skin. Figure 6.1 shows the experiment parameters, the laser illuminating on the human skin and the waveform of the recovered sound.

In this experiment the target was speaking “one, two, three, four” in English, and from the resultant waveform we can see that the content of the human speech can be extracted with our experiment setup. But the currently the quality of the recovered sound is not as good as the experiment before, this is because 1) the human’s body is different from surface of a speaker, the material is different, this causes the speckle image reflected from human skin is noisy. 2) the human body cannot be perfect static. Those reason makes it difficult to capture a clear speckle motion compared with recovering sound from a speaker. If those problems can be solved, we can achieve our research target, make a hi-fi optical microphone in the future.

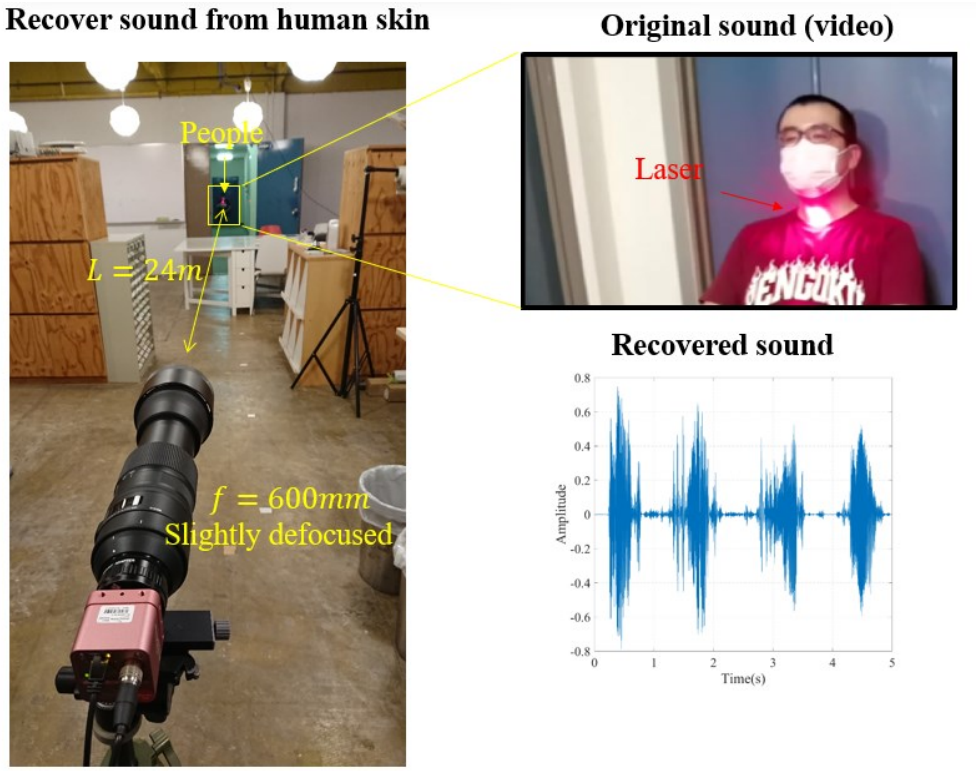

Figure 6.1 Experiment of extract speech from human skin with long distance

REFERENCE

[1] De Silva, C.W., *Vibration: fundamentals and practice*. 2006: CRC press.

[2] Martinez, D.R., T.D. Hinnerichs, and J.M. Redmond, *Vibration control for precision manufacturing using piezoelectric actuators.* Journal of intelligent material systems and structures, 1996. **7**(2): p. 182-191.

[3] Crispino, M. and M. D'Apuzzo, *Measurement and prediction of traffic-induced vibrations in a heritage building.* Journal of Sound and Vibration, 2001. **246**(2): p. 319-335.

[4] Hunaidi, O., *Traffic vibrations in buildings*. 2000: Citeseer.

[5] Norén-Cosgriff, K.M., et al., *Building damage due to vibration from rock blasting.* Soil Dynamics and Earthquake Engineering, 2020. **138**: p. 106331.

[6] Daga, A.P. and L. Garibaldi, *Machine vibration monitoring for diagnostics through hypothesis testing.* Information, 2019. **10**(6): p. 204.

[7] Goyal, D. and B.S. Pabla, *The vibration monitoring methods and signal processing techniques for structural health monitoring: a review.* Archives of Computational Methods in Engineering, 2016. **23**(4): p. 585-594.

[8] Tandon, N. and A. Choudhury, *A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings.* Tribology international, 1999. **32**(8): p. 469-480.

[9] Buzdugan, G., E. Mihăilescu, and M. Radeș, *Instrumentation for vibration measurement*, in *Vibration measurement*. 1986, Springer. p. 117-185.

[10] Okabe, T., K. Hirata, and T. Kumai, *Vibration measurements on a 32,000 ton dw super tanker.* International Shipbuilding Progress, 1956. **3**(24): p. 409-414.

[11] Sinha, D. and F. Feroz, *Obstacle Detection on Railway Tracks Using Vibration Sensors and Signal Filtering Using Bayesian Analysis.* IEEE Sensors Journal, 2016. **16**(3): p. 642-

649.

[12] Sue, C., et al. *Design and experimental analysis for innovative wide bandwidth vibration sensor*. in *2016 11th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)*. 2016.

[13] Bernstein, J., et al., *Low-noise MEMS vibration sensor for geophysical applications.* Journal of Microelectromechanical Systems, 1999. **8**(4): p. 433-438.

[14] *Contact type vibration measuring device*. Available from: http://www.musashi-in.co.jp/catalog_9/PDF/100.pdf.

[15] Hui, Z., P. Chao, and S. Xiaohan. *Vibration pattern recognition and classification in OTDR based distributed optical-fiber vibration sensing system*. in *Proc.SPIE*. 2014.

[16] Lutzmann, P., et al. *Laser vibration sensing: overview and applications*. in *Proc.SPIE*. 2011.

[17] Ding, Z., et al., *Long-range vibration sensor based on correlation analysis of optical frequency-domain reflectometry signals.* Optics Express, 2012. **20**(27): p. 28319-28329.

[18] *Free-space Single-point Laser Doppler Vibrometer*. Available from: http://www.holobright.com/free-space-single-point-laser-doppler-vibrometer/.

[19] Castellini, P., M. Martarelli, and E.P. Tomasini, *Laser Doppler Vibrometry: Development of advanced solutions answering to technology's needs.* Mechanical Systems and Signal Processing, 2006. **20**(6): p. 1265-1285.

[20] Sels, S., et al., *Three-dimensional full-field vibration measurements using a handheld single-point laser Doppler vibrometer.* Mechanical Systems and Signal Processing, 2019. **126**: p. 427-438.

[21] Morse, P.M., A. Acoustical Society of, and P. American Institute of, *Vibration and sound*. Vol. 2. 1948: McGraw-Hill New York.

[22] Rayleigh, J.W.S.B., *The theory of sound*. Vol. 2. 1896: Macmillan.

[23] Gabriela, P., et al. *Comparison of high speed imaging technique to laser vibrometry for detection of vibration information from objects*. in *Proc.SPIE*. 2015.

[24] Dashan, Z., et al., *Efficient subtle motion detection from high-speed video for sound recovery and vibration analysis using singular value decomposition-based approach.* Optical Engineering, 2017. **56**(9): p. 1-10.

[25] Zhang, D., et al., *Note: Sound recovery from video using SVD-based information extraction.* Review of Scientific Instruments, 2016. **87**(8): p. 086111.

[26] Zhaoyang, W., N. Hieu, and Q. Jason, *Audio extraction from silent high-speed video using an optical technique.* Optical Engineering, 2014. **53**(11): p. 1-3.

[27] Akutsu, M., Y. Oikawa, and Y. Yamasaki, *Extract voice information using high-speed camera.* Proceedings of Meetings on Acoustics, 2013. **19**(1): p. 055019.

[28] Davis, A., et al., *The visual microphone: Passive recovery of sound from video.* ACM Transactions on Graphics, 2014. **33**(4): p. 1-10.

[29] *Lamphone: Real-Time Passive Sound Recovery from Light Bulb Vibrations.*

[30] Dainty, J.C., *Laser speckle and related phenomena*. Vol. 9. 2013: Springer science & business Media.

[31] Zalevsky, Z., et al., *Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern.* Optics Express, 2009. **17**(24): p. 21566-21580.

[32] Chen, Z., et al., *Audio signal reconstruction based on adaptively selected seed points from laser speckle images.* Optics Communications, 2014. **331**: p. 6-13.

[33] Ge, Z., et al., *Sound recovery via intensity variations of speckle pattern pixels selected with variance-based method.* Optical Engineering, 2018. **57**(2): p. 1-9.

[34] Bianchi, S., *Vibration detection by observation of speckle patterns.* Applied Optics,

2014. **53**(5): p. 931-936.

[35] Newton, I., *Opticks*. 1730: Prabhat Prakashan.

[36] Langmuir, R.V., *Scattering of laser light.* Applied Physics Letters, 1963. **2**(2): p. 29-30.

[37] Oliver, B.M., *Sparkling spots and random diffraction.* Proceedings of the IEEE, 1963. **51**(1): p. 220-221.

[38] Goodman, J.W., *Some fundamental properties of speckle.* JOSA, 1976. **66**(11): p. 1145-1150.

[39] Goodman, J.W., *Speckle phenomena in optics: theory and applications*. 2007: Roberts and Company Publishers.

[40] Goodman, J.W., *Statistical properties of laser speckle patterns*, in *Laser speckle and related phenomena*. 1975, Springer. p. 9-75.

[41] Smythe, R. and R. Moore, *Instantaneous phase measuring interferometry.* Optical Engineering, 1984. **23**(4): p. 234361.

[42] Sprague, R.A., *Surface roughness measurement using white light speckle.* Applied Optics, 1972. **11**(12): p. 2811-2816.

[43] Léger, D., E. Mathieu, and J.C. Perrin, *Optical surface roughness determination using speckle correlation technique.* Applied optics, 1975. **14**(4): p. 872-877.

[44] Smith, B.M., et al., *CoLux: Multi-object 3d micro-motion analysis using speckle imaging.* ACM Transactions on Graphics, 2017. **36**(4): p. 1-12.

[45] Smith, B.M., M. O'Toole, and M. Gupta. *Tracking multiple objects outside the line of sight using speckle imaging*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[46] Jo, K., M. Gupta, and S.K. Nayar. *Spedo: 6 dof ego-motion sensor using speckle defocus imaging*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[47] Li, Z., et al., *Viaxl: A Solution of a Low-Cost Real-Time Visual Accelerometer Based on Laser Speckle Optical Flow Detection.* Sensors, 2020. **20**(24).

[48] Haken, H., *Laser theory*, in *Light and Matter Ic/Licht und Materie Ic*. 1970, Springer. p. 1-304.

[49] Shannon, R., *Applied Optics and Optical Engineering V8*. Vol. 8. 2012: Elsevier.

[50] Butters, J.N. and J.A. Leendertz, *Speckle pattern and holographic techniques in engineering metrology.* Optics Laser Technology, 1971. **3**(1): p. 26-30.

[51] Schnell, U., J. Piot, and R. Dändliker, *Detection of movement with laser speckle patterns: statistical properties.* JOSA A, 1998. **15**(1): p. 207-216.

[52] Houghton, A., G. Rees, and P. Ivey, *A method for processing laser speckle images to extract high-resolution motion.* Measurement Science and Technology, 1997. **8**(6): p. 611.

[53] Yamaguchi, I. and T. Fujita, *Linear and rotary encoders using eletronic speckle correlation.* Optical engineering, 1991. **30**(12): p. 1862-1868.

[54] Stetson, K.A. and W.R. Brohinsky, *Electrooptic holography and its application to hologram interferometry.* Applied optics, 1985. **24**(21): p. 3631-3637.

[55] Peters, W.H. and W.F. Ranson, *Digital imaging techniques in experimental stress analysis.* Optical engineering, 1982. **21**(3): p. 213427.

[56] Yamaguchi, I., *A laser-speckle strain gauge.* Journal of Physics E: Scientific Instruments, 1981. **14**(11): p. 1270.

[57] Castleman, K.R., *Digital image processing.* 1993.

[58] Anuta, P.E., *Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques.* IEEE transactions on Geoscience Electronics, 1970. **8**(4): p. 353-368.

[59] Tang, Z.-Z., et al., *Photogrammetry-based two-dimensional digital image correlation*

with nonperpendicular camera alignment. Optical Engineering, 2012. **51**(2): p. 023602.

[60] Pan, B., *Recent progress in digital image correlation.* Experimental mechanics, 2011. **51**(7): p. 1223-1235.

[61] Pan, B., H. Xie, and Z. Wang, *Equivalence of digital image correlation criteria for pattern matching.* Applied optics, 2010. **49**(28): p. 5501-5509.

[62] Pan, B., et al., *Study on subset size selection in digital image correlation for speckle patterns.* Optics express, 2008. **16**(10): p. 7037-7048.

[63] Dipl-phys, R.H., *Speckle pattern correlation by digital image processing.* Measurement, 1987. **5**(1): p. 30-33.

[64] Viswanathan, D.G. *Features from accelerated segment test (fast)*. in *Proceedings of the 10th workshop on Image Analysis for Multimedia Interactive Services*. 2009.

[65] Rublee, E., et al. *ORB: An efficient alternative to SIFT or SURF*. in *2011 International Conference on Computer Vision*. 2011.

[66] Bay, H., et al., *Speeded-up robust features (SURF).* Computer vision and image understanding, 2008. **110**(3): p. 346-359.

[67] Lindeberg, T., *Scale invariant feature transform.* 2012.

[68] Horn, B.K.P. and B.G. Schunck, *Determining optical flow.* Artificial intelligence, 1981. **17**(1-3): p. 185-203.

[69] Farnebäck, G. *Two-frame motion estimation based on polynomial expansion*. in *Scandinavian conference on Image analysis*. 2003. Springer.

[70] Farnebäck, G., *Polynomial expansion for orientation and motion estimation.* 2002.

[71] Thomas, O.H.C., K. Krzysztof, and P.T. Ralph. *Speckle tracking approaches in speckle sensing*. in *Proc.SPIE*. 2017.

[72] Charrett, T. and R. Tatam, *Performance and Analysis of Feature Tracking Approaches in*

Laser Speckle Instrumentation. Sensors, 2019. **19**(10).

[73] Reu, P.L., *Experimental and Numerical Methods for Exact Subpixel Shifting.* Experimental Mechanics, 2011. **51**(4): p. 443-452.

[74] Campbell, M., et al., *Review of LDA and PIV applied to the measurement of sound and acoustic streaming.* Optics & Laser Technology, 2000. **32**(7): p. 629-639.

[75] Zieger, C., A. Brutti, and P. Svaizer. *Acoustic based surveillance system for intrusion detection*. in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2009. IEEE.

[76] Ishikawa, K., et al., *Simultaneous imaging of flow and sound using high-speed parallel phase-shifting interferometry.* Optics Letters, 2018. **43**(5): p. 991-994.

[77] Matoba, O., et al., *Optical voice recorder by off-axis digital holography.* Optics Letters, 2014. **39**(22): p. 6549-6552.

[78] Lv, T., H.-y. Zhang, and C.-h. Yan, *Double mode surveillance system based on remote audio/video signals acquisition.* Applied Acoustics, 2018. **129**: p. 316-321.

[79] Han, X., et al., *A remote human activity detection system based on partial-fiber LDV and PTZ camera.* Optics & Laser Technology, 2019. **111**: p. 575-584.

[80] Chandrakala, S. and S. Jayalakshmi, *Environmental Audio Scene and Sound Event Recognition for Autonomous Surveillance: A Survey and Comparative Studies.* ACM Computing Surveys, 2019. **52**: p. 1-34.

[81] Barcellona, C., et al., *Remote recovery of audio signals from videos of optical speckle patterns: a comparative study of signal recovery algorithms.* Optics Express, 2020. **28**(6): p. 8716-8723.

[82] Gregory, D.A., *Basic physical principles of defocused speckle photography: a tilt topology inspection technique.* Optics & Laser Technology, 1976. **8**(5): p. 201-213.

[83] Tiziani, H.J., *A study of the use of laser speckle to measure small tilts of optically rough surfaces accurately.* Optics Communications, 1972. **5**(4): p. 271-276.

[84] Guizar-Sicairos, M., S.T. Thurman, and J.R. Fienup, *Efficient subpixel image registration algorithms.* Optics Letters, 2008. **33**(2): p. 156-158.

LIST OF PUBLICATIONS

Journal Publications

1. Wu, Nan, and S. Haruyama. "Real-time audio detection and regeneration of moving sound source based on optical flow algorithm of laser speckle images." *Optics Express* **28**.4 (2020): 4475-4488.

2. Wu, Nan, and Shinichiro Haruyama. "The 20k Samples-Per-Second Real Time Detection of Acoustic Vibration Based on Displacement Estimation of One-Dimensional Laser Speckle Images." *Sensors* **21**.9 (2021): 2938.

Full Length Conference Publications

3. Wu, Nan, and Shinichiro Haruyama. "Real-time Sound Detection and Regeneration Based on Optical Flow Algorithm of Laser Speckle Images." *2019 28th Wireless and Optical Communications Conference (WOCC)*. IEEE, 2019.

4. Wu, Nan, and Shinichiro Haruyama. "Fast motion estimation of one-dimensional laser speckle image and its application on real-time audio signal acquisition." *2020 the 6th International Conference on Communication and Information Processing (ICCIP)*. ACM, 2020.