

Article

The 20k Samples-Per-Second Real Time Detection of Acoustic Vibration Based on Displacement Estimation of One-Dimensional Laser Speckle Images

Nan Wu *  and Shinichiro Haruyama

Graduate School of System Design and Management, Keio University, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8526, Japan; haruyama@sdm.keio.ac.jp

* Correspondence: wunan_sdm@keio.jp

Abstract: 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 could not achieve real time reconstruction of an audio signal. In this manuscript, we use a one-dimensional laser speckle image to measure the acoustic vibration of sound source and propose 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 system can achieve 20k samples-per-second sampling rate, and the audio signal can be reconstructed with high quality in real time.

Keywords: laser speckle image; line-scan sensor; image processing; real time system



Citation: Wu, N.; Haruyama, S. The 20k Samples-Per-Second Real Time Detection of Acoustic Vibration Based on Displacement Estimation of One-Dimensional Laser Speckle Images. *Sensors* **2021**, *21*, 2938. <https://doi.org/10.3390/s21092938>

Academic Editors: Min Yong Jeon and Paola Saccomandi

Received: 4 March 2021

Accepted: 20 April 2021

Published: 22 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. 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 [1–3], surveillance of audio signal in rescue [4–6], visual accelerometer [7], and so on.

Generally, optical vibrometers utilize light interference to measure the vibration of objects. The representative instrument is laser Doppler vibrometers (LDVs) [8–10]. In the last few decades, with the invention of digital cameras and the improvement of computer performance, there were many studies on vibration measurement which combine optical sensing and image processing technologies [11–13]. These studies can be divided into two categories according to the light source utilized. One way is using natural lighting illumination, and the representative research is the visual microphone proposed by researchers from MIT in 2014 [14]. In this research, the authors used a high-speed camera to record the video of a vibrated object and used a phase-based algorithm to extract the audio signal from the captured video. This approach does not require additional illumination, which makes system structure simple and easy to implement. However, processing the video containing the details of the object vibration requires a high computational cost. Slow image processing speed is a disadvantage of this method. Another way is using coherent light illumination. Specifically, when the laser beam illuminates an object of interest, the scattered light interferes in space, forming a speckle pattern. When the object vibrates, the speckle pattern presents a corresponding movement. The camera captures the dynamic speckle pattern rather than the object itself, and the vibration information of the object can be extracted by analyzing the motion of the captured speckle images. In the past, the digital image correlation (DIC) technique was well used for speckle displacement analysis [15–20].

However, DIC usually suffers from two disadvantages: (1) up-sampling of the correlation coefficients is required to achieve sub-pixel precision displacement estimation, which is very time consuming; (2) it requires a certain size of speckle pattern to provide a stable accuracy of displacement estimation. The slow computational speed makes it hard to fulfill real time reconstruction of an audio signal using the DIC method. Another technique extracts object vibration using the gray value variation of the speckle image [21–23]. This approach uses a special algorithm to filter out the appropriate seed points, obtaining the gray value variation of each point and carrying out data fusion to reconstruct the audio signal. However, this 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 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 [24,25]. The camera captures over 2000 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 system. 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 [26].

In this manuscript, we propose using a line-scan camera to further improve the sampling speed of the sensing system. 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 system.

The structure of this paper is as follows. The explanation of the speckle sensing model along with our displacement estimation algorithm are given in Section 2. Then, the experiment results of our system, including the performance of our proposed displacement estimation algorithm and the results of audio signal detection, are shown in Section 3. Finally, the conclusion of the paper is given in Section 4.

2. Materials and Methods

2.1. Speckle Sensing Model

The sketch of our proposed system is shown in Figure 1. An infrared laser beam ($\lambda = 980$ nm, $power = 50$ mW) 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 system 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 system, 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.

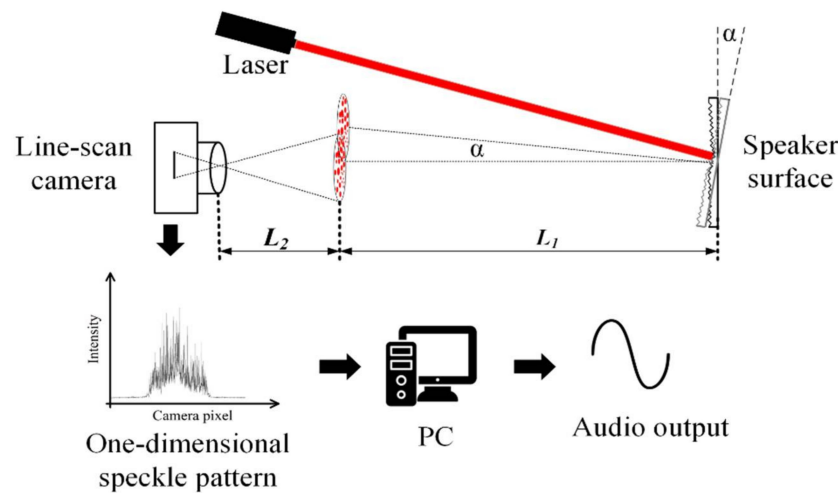


Figure 1. Schematic of the real time audio signal detection system using 1D laser speckle images.

In the system, 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. Previous research proved that the displacement of the speckle pattern is mainly determined by the tilting of the object if the imaging system is strongly defocused [15]. Assume the tilting angle of the object is α , then the shift amount D of the speckle pattern at the distance L_1 from the object can be expressed as:

$$D = L_1 \tan \alpha \quad (1)$$

The magnification of the imaging system is $M = \frac{L_2 - f}{f}$. Therefore, the image shift d at the image sensor is:

$$d = D \times \frac{1}{M} = \frac{L_1 f}{L_2 - f} \tan \alpha \quad (2)$$

Since the tilting angle of the object is usually very small, and L_2 is much larger than f , the expression of d can be simplified and approximately written as:

$$d = \frac{L_1}{L_2} f \alpha \quad (3)$$

It can be seen that 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.

2.2. Explanation of 1D Optical Flow Algorithm for Displacement Estimation

In our system, 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 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:

$$f_1(x) = a_1 x^2 + b_1 x + c_1 \quad (4)$$

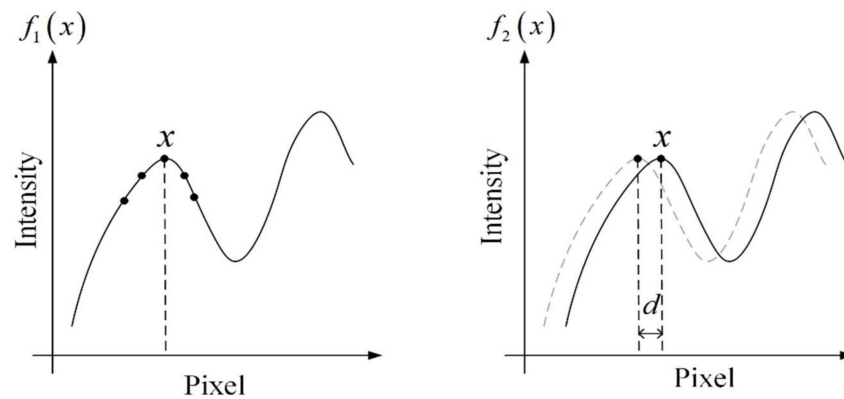


Figure 2. Local signal model of the 1D image undergoes a small shift.

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:

$$f_2(x) = a_2x^2 + b_2x + c_2 \quad (5)$$

It should be noted that $f_2(x)$ is created by shifting $f_1(x)$ with the amount of d . Therefore, $f_2(x)$ can be also expressed as $f_1(x - d)$. Expanding the expression, we have:

$$\begin{aligned} f_2(x) &= f_1(x - d) \\ &= a_1(x - d)^2 + b_1(x - d) + c_1 \\ &= a_1x^2 + (b_1 - 2a_1d)x + a_1d^2 - b_1d + c_1 \end{aligned} \quad (6)$$

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:

$$a_1 = a_2 \quad (7)$$

$$b_1 - 2a_1d = b_2 \quad (8)$$

$$a_1d^2 - b_1d + c_1 = c_2 \quad (9)$$

By solving Equation (8), we can determine the value of d , which is:

$$d = -\frac{1}{2a}(b_2 - b_1) \quad (10)$$

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 (10) 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 [27,28]. 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 (10), let $\Delta b = -\frac{1}{2}(b_2 - b_1)$. At each pixel, we have:

$$a \times d = \Delta b \quad (11)$$

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

such as Equation (11). Here, we try to find d satisfying N equations as well as possible. In other words, our goal is minimizing the following function:

$$e = \sum_{i=1}^N |a_i \times d - \Delta b_i| \quad (12)$$

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 a and Δb are obtained. Finally, the least square solution of displacement d is determined, which can be expressed as:

$$d = \frac{\sum_{i=1}^N a_i \Delta b_i}{\sum_{i=1}^N a_i^2} \quad (13)$$

3. Experiment Results

3.1. Performance Test 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 this manuscript, 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 3a shows a captured 2D speckle pattern with the size of 512×512 pixels, whereas Figure 3b 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 [29]. 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.

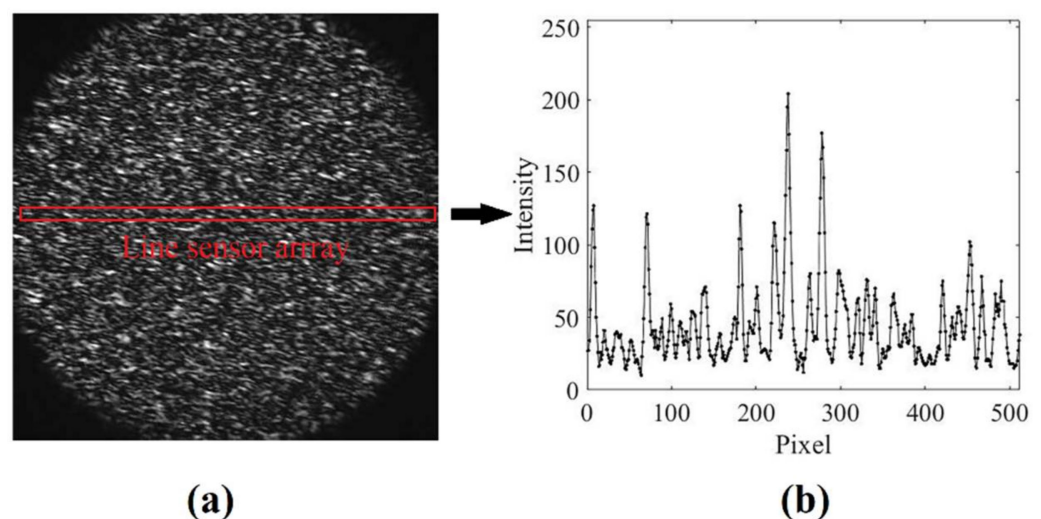


Figure 3. Captured speckle images; the shutter time was $20 \mu\text{s}$. (a) Captured two-dimensional speckle pattern with the size of 512×512 pixels. (b) Captured 1D speckle pattern with the size of 1×500 pixels.

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 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 was 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 imaging system. In our situation, the 5 pixels polynomial fitting was adopted for further tests since it showed the best performance on displacement estimation.

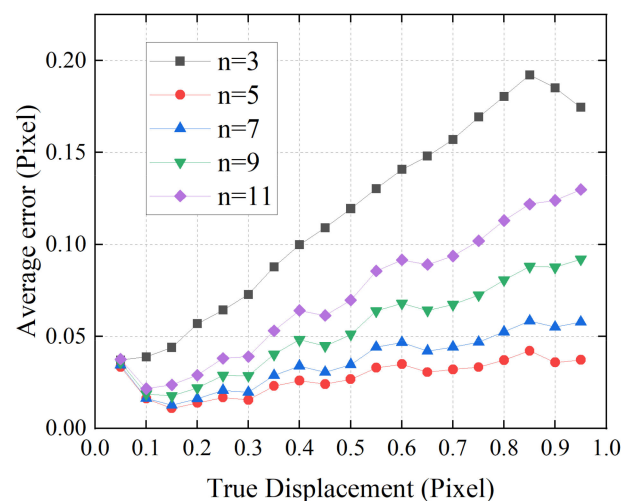


Figure 4. Dependence of the average error result on the shift amount for different quadratic polynomial fitting kernel size.

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 [30]. Here, we set the subpixel precision to 100, which meant the resolution of the cross-correlation algorithm was 0.01 pixel. Figure 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 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 1×50 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.

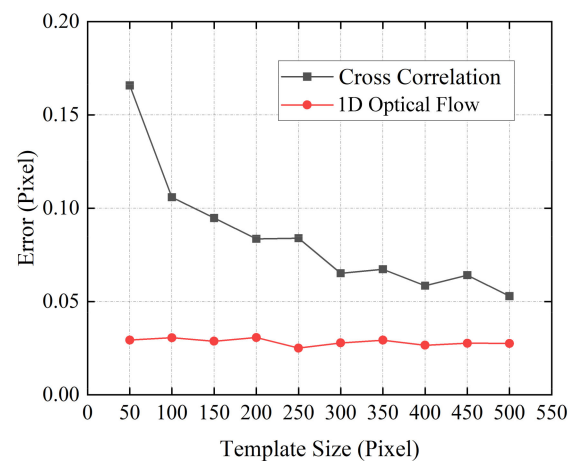


Figure 5. Dependence of the average error of displacement estimation on the template size for different algorithms.

In our system, 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 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.

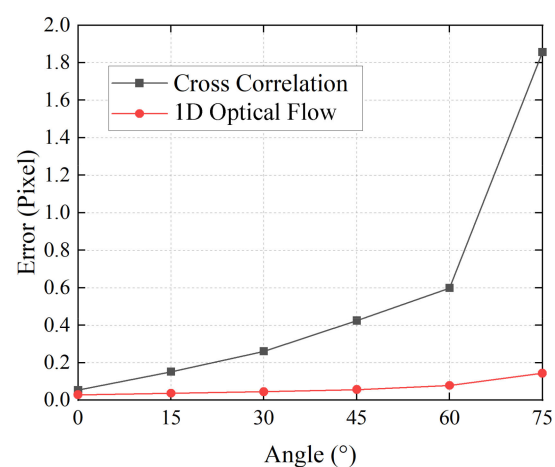


Figure 6. Dependence of the average error of displacement estimation on the angle for different algorithms.

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 7a. It can be seen from Figure 7a 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 7b, the 1D optical flow algorithm took only 41 μs for displacement estimation, whereas the cross-correlation method required 205 μs . 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 20 kHz. According to the Nyquist sampling theorem, our system can sample the signal frequency up to 10 kHz in real time. This sampling speed can satisfy the requirement of human speech sampling.

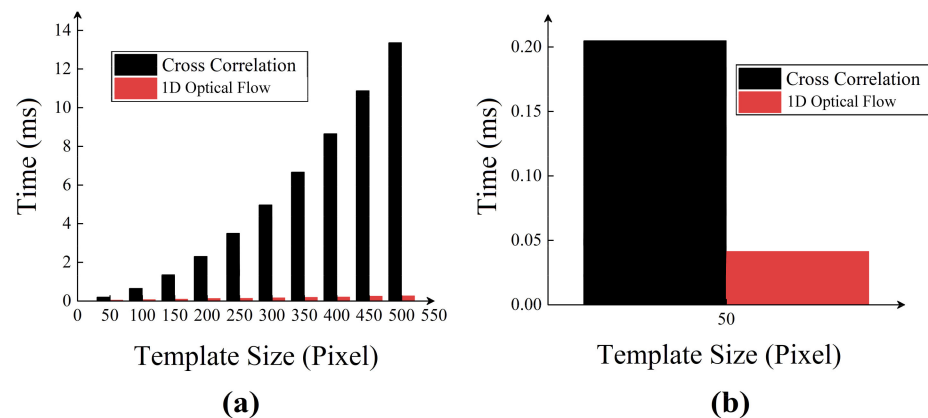


Figure 7. Comparison of average time consumption of two algorithms. (a) Dependence of time consumption on template size. (b) Time consumption of two algorithms with a template size of 1×50 pixels.

3.2. Result of Real Time Audio Signal Extraction

In the next experiment, we tried to use our system 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 8.

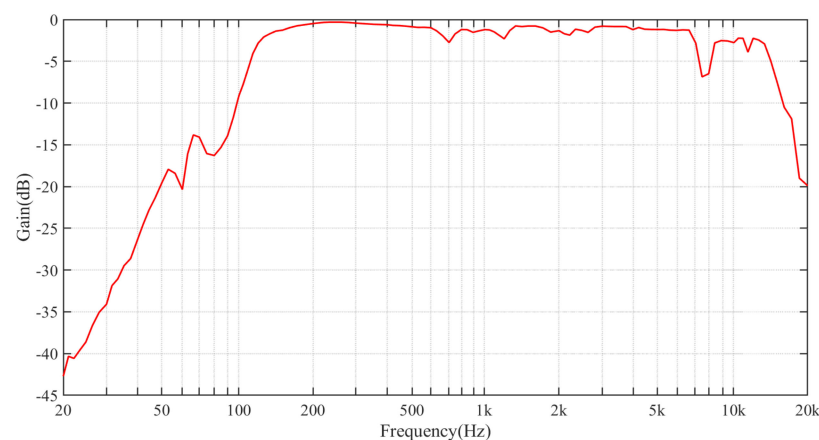


Figure 8. Frequency response of the speaker used in the experiments.

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 μs . The speaker was positioned 2 m 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 9a presents the first 20 ms 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 9b shows the spectrogram of the results. The results matched the signal that was sent. It should be noted that the amplitude of the 100 Hz regenerated waveform was smaller than other signals, because the frequency response of the speaker was low at 100 Hz frequency. The performance of our system was also investigated using the signal-to-noise ratio (SNR) and the total harmonic distortion (THD). SNR is defined as the ratio of signal power to noise power. A larger SNR means better signal quality. On the other hand, THD is defined as the ratio of the sum of the powers of all harmonic components to the power of the fundamental frequency. A lower THD means a more accurate reproduction of an audio signal. The results obtained with increasing frequencies from 100 Hz to 1000 Hz are shown in Figure 10, which prove that the sinusoidal audio signal could be restored with high quality using our system.

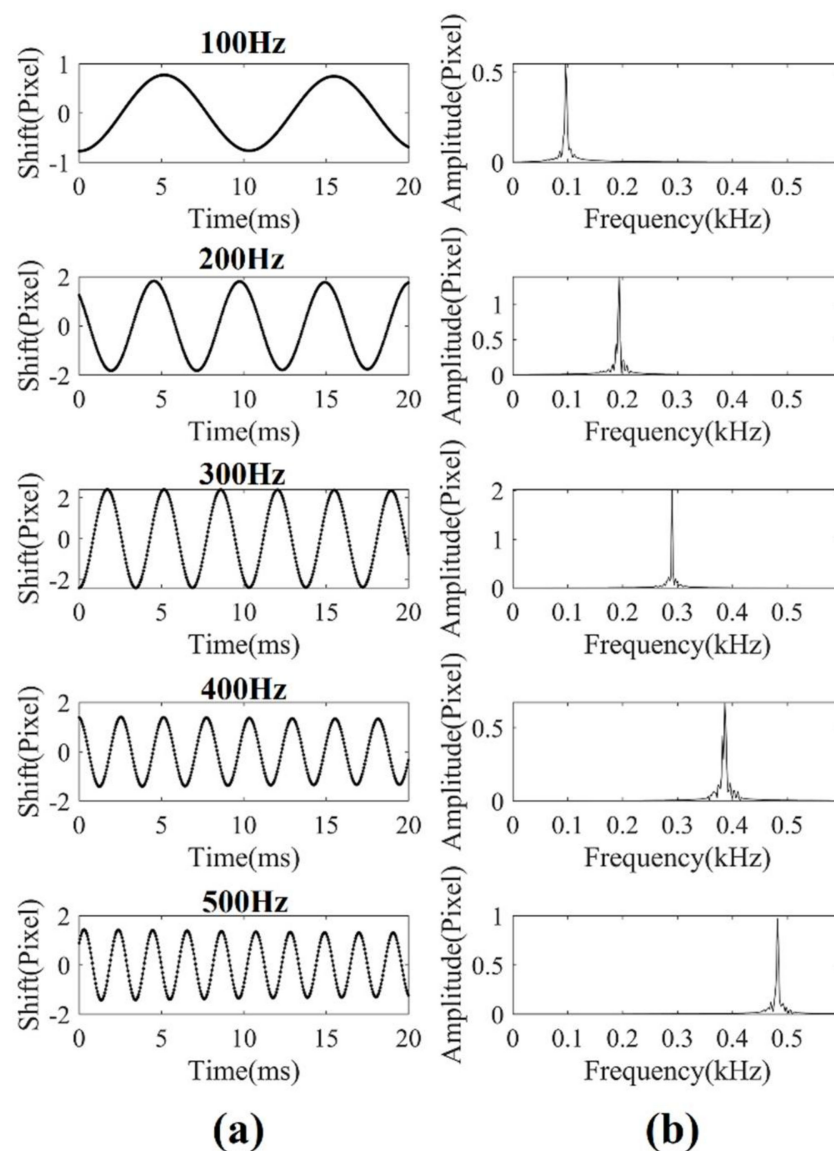


Figure 9. (a) Restored waveforms of audio signals at different frequencies. (b) Spectrogram of restored signals at different frequencies.

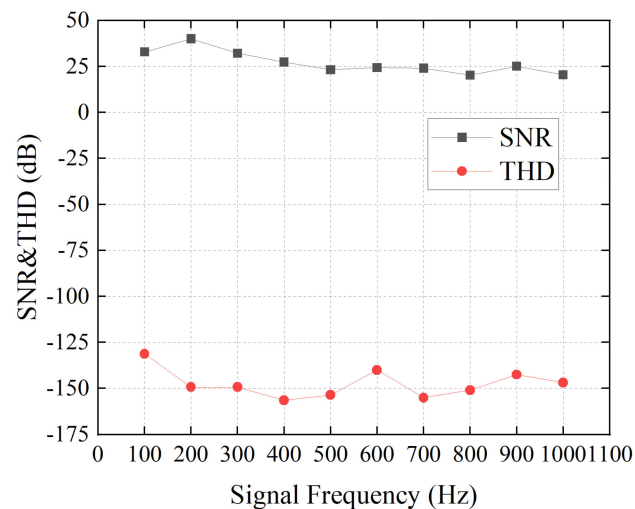


Figure 10. SNR and THD of the restored signal with different signal frequency.

In the next experiment, we tried to extract a signal with high frequency. This time, the 5 kHz 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 11a. In Figure 11b, 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 12.64 dB 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.

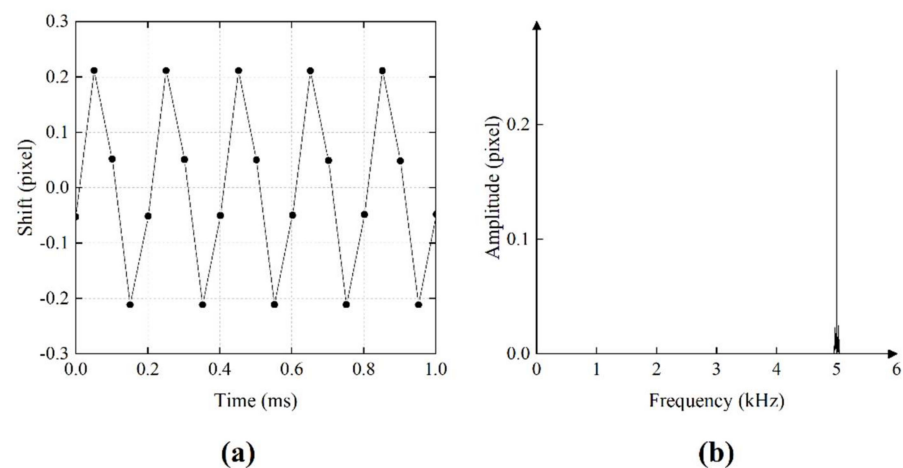


Figure 11. (a) The first 1 ms of restored temporal signal; the original signal frequency was 5 kHz. (b) Spectrogram of the restored audio signal.

Finally, we present the result of human speech extraction with our system. Figure 12a shows the temporal signal of the original sound, which was the voice of a male counting from zero to nine in English. Figure 12b shows the temporal signal of the restored sound. Figure 12c,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 system. For reference, the audio files of both original music (see Supplementary Materials Audio S1) and regenerated music (see Supplementary Materials Audio S2) are provided as the result of this experiment.

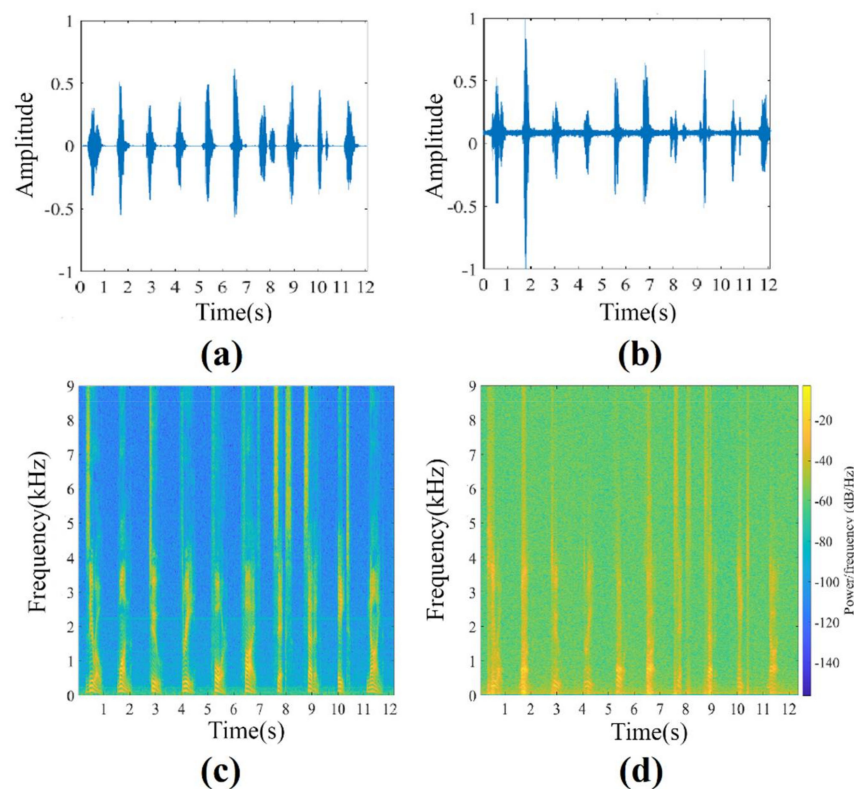


Figure 12. Experiment result of restoring human speech in real time. (a,b) Temporal signal of the original sound and the restored sound. (c,d) Spectrogram of the original and the restored sound. (see also Supplementary Materials at the end of the paper).

3.3. Further Discussion

In Section 2, we provide an explanation of the laser speckle sensing model. In our system, 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 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 1 and Figure 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.

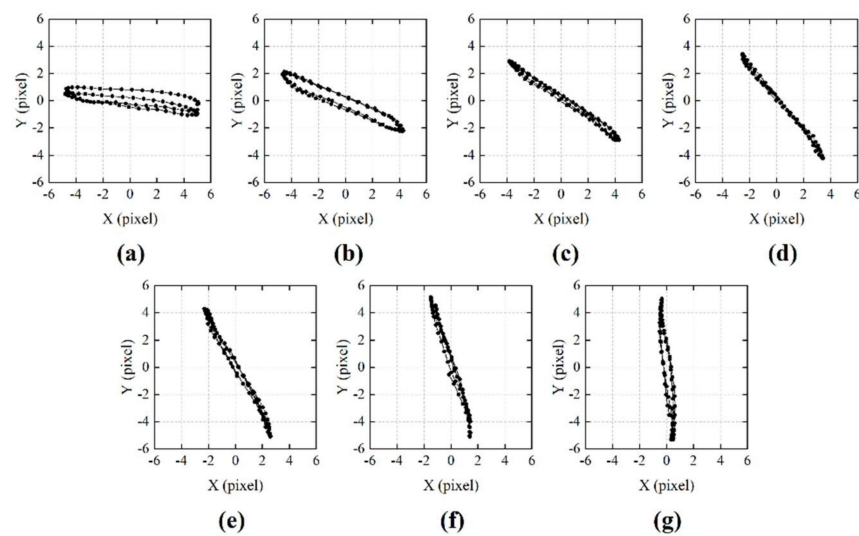


Figure 13. Trajectories of the 2D dynamic speckle patterns with different rotation stage angles. (a) Stage angle was 60° . (b) Stage angle was 75° . (c) Stage angle was 90° . (d) Stage angle was 105° . (e) Stage angle was 120° . (f) Stage angle was 135° . (g) Stage angle was 150° .

Table 1. Relationship between rotation stage angle and the 2D image shift angle.

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

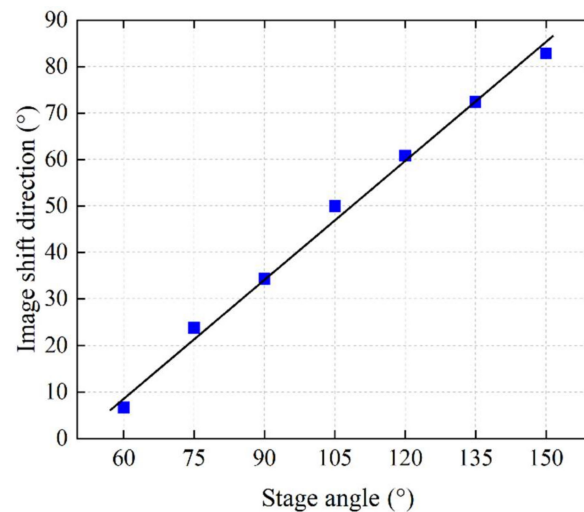


Figure 14. Dependence of the 2D speckle image shift angle and the rotation stage angle.

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 15 shows the SNR results of both algorithms with different situations. Figure 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 16a,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 16e. On the other hand, our algorithm showed a better result than the cross-correlation method. As shown in Figure 16b, 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 16c,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.

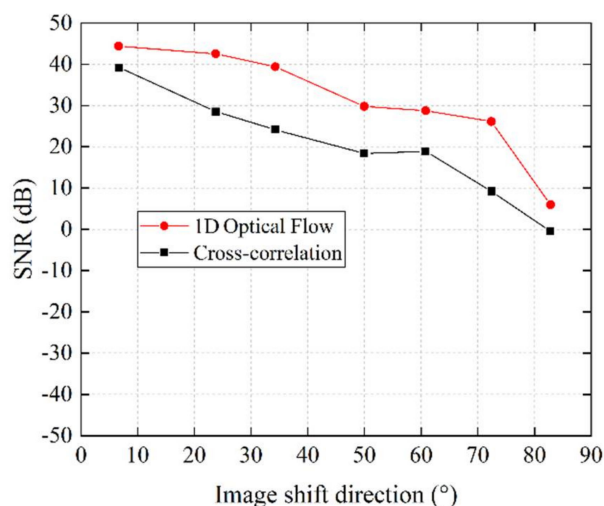


Figure 15. Dependence of the SNR of the restored audio signal on the 2D image shift angle for different algorithms.

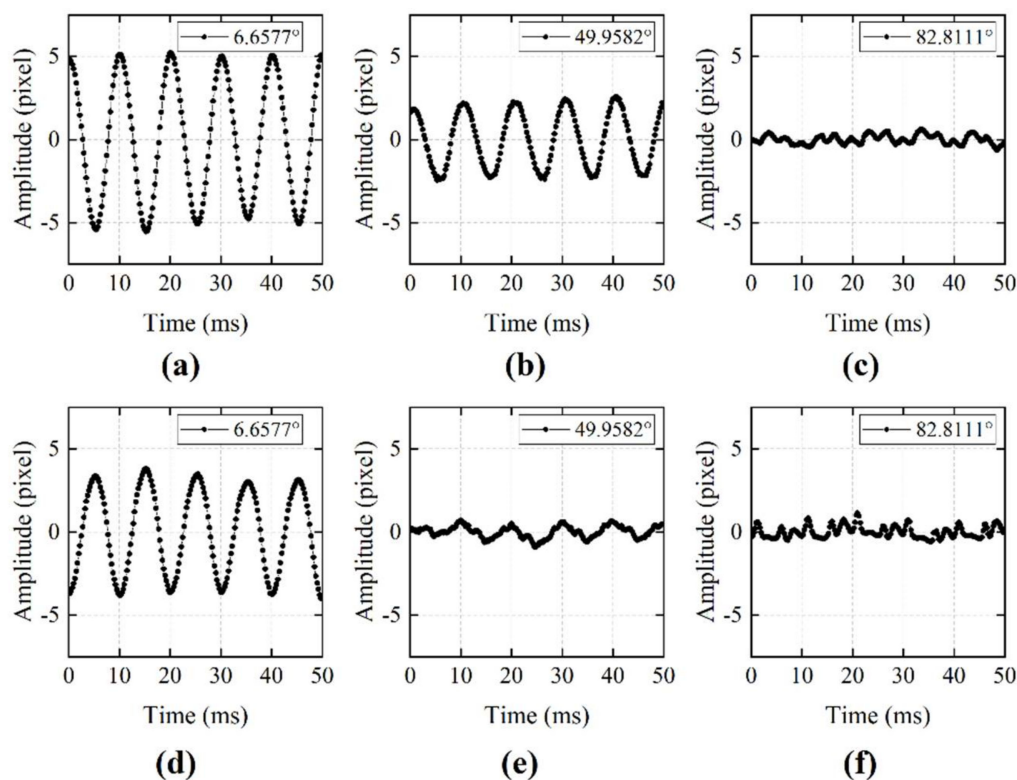


Figure 16. Restored waveform with different 2D image shift angle using different algorithms. (a–c) Restored waveform using 1D optical flow algorithm. (d–f) Restored waveform using cross-correlation algorithm.

4. Conclusions

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

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/s21092938/s1>, Audio S1: Original audio file, Audio S2: Recovered audio file.

Author Contributions: Conceptualization, N.W. and S.H.; methodology, N.W. and S.H.; software, N.W., validation, N.W., data curation: N.W., writing—original draft preparation, N.W.; writing—review and editing, S.H.; supervision, S.H. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Keio University doctorate student grant-in-aid program, graduate school recommendation by graduate school of System Design and Management, academic year 2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Wang, C.C.; Trivedi, S.; Jin, F.; Swaminathan, V.; Rodriguez, P.; Prasad, N.S. High sensitivity pulsed laser vibrometer and its application as a laser microphone. *Appl. Phys.* **2009**, *94*, 051112. [[CrossRef](#)]
2. Bakhom, E.G.; Cheng, M.H.M. Advanced Optical Microphone. *IEEE Sens. J.* **2014**, *14*, 7–14. [[CrossRef](#)]
3. Wan, Q.; Wang, C.; Xu, K.; Kang, J.; Wu, Y.; Trivedi, S.B.; Gehlbach, P.; Boctor, E. Ultrasound Signal Detection with Multi-bounce Laser Microphone. In Proceedings of the 2020 IEEE International Ultrasonics Symposium (IUS), Las Vegas, NV, USA, 7–11 September 2020; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2020; pp. 1–4. [[CrossRef](#)]
4. Lv, T.; Zhang, H.Y.; Yan, C.H. Double mode surveillance system based on remote audio/video signals acquisition. *Appl. Acoust.* **2009**, *129*, 316–321. [[CrossRef](#)]
5. Han, X.; Lv, T.; Wu, S.; Li, Y.; He, B. A remote human activity detection system based on partial-fiber LDV and PTZ camera. *Opt. Laser Technol.* **2019**, *111*, 575–584. [[CrossRef](#)]
6. Chandrakala, S.; Jayalakshmi, S.L. Environmental audio scene and sound event recognition for autonomous surveillance: A survey and comparative studies. *ACM Comput. Surv.* **2019**, *52*, 1–34. [[CrossRef](#)]
7. Li, Z.; Yang, W.; Xiong, X.; Wang, Z.; Zou, X. Viaxl: A Solution of a Low-Cost Real-Time Visual Accelerometer Based on Laser Speckle Optical Flow Detection. *Sensors* **2020**, *20*, 7033. [[CrossRef](#)]
8. Li, R.; Wang, T.; Zhu, Z.; Xiao, W. Vibration characteristics of various surfaces using an LDV for long-range voice acquisition. *IEEE Sens. J.* **2010**, *11*, 1415–1422. [[CrossRef](#)]
9. Li, R.; Madampoulos, N.; Zhu, Z.; Xie, L. Performance comparison of an all-fiber-based laser Doppler vibrometer for remote acoustical signal detection using short and long coherence length lasers. *Appl. Opt.* **2012**, *51*, 5011–5018. [[CrossRef](#)]
10. Wu, S.S.; Lv, T.; Han, X.Y.; Yan, C.H.; Zhang, H.Y. Remote audio signals detection using a partial-fiber laser Doppler vibrometer. *Appl. Acoust.* **2018**, *130*, 216–221. [[CrossRef](#)]
11. Matoba, O.; Inokuchi, H.; Nitta, K.; Awatsuji, Y. Optical voice recorder by off-axis digital holography. *Opt. Lett.* **2014**, *39*, 6549–6552. [[CrossRef](#)]
12. Ishikawa, K.; Tanigawa, R.; Yatabe, K.; Oikawa, Y.; Onuma, T.; Niwa, H. Simultaneous imaging of flow and sound using high-speed parallel phase-shifting interferometry. *Opt. Lett.* **2018**, *43*, 991–994. [[CrossRef](#)]
13. Bianchi, S. Vibration detection by observation of speckle patterns. *Appl. Opt.* **2014**, *53*, 931–936. [[CrossRef](#)]
14. Davis, A.; Rubinstein, M.; Wadhwa, N.; Mysore, G.J.; Durand, F.; Freeman, W.T. The visual microphone: Passive recovery of sound from video. *ACM Trans. Graph.* **2014**, *33*, 1–10. [[CrossRef](#)]
15. Zalevsky, Z.; Beiderman, Y.; Margalit, I.; Gingold, S.; Teicher, M.; Mico, V.; Garcia, J. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Opt. Express* **2009**, *17*, 21566–21580. [[CrossRef](#)] [[PubMed](#)]

16. Li, L.; Gubarev, F.A.; Klenovskii, M.S.; Bloshkina, A.I. Vibration measurement by means of digital speckle correlation. In Proceedings of the 2016 International Siberian Conference on Control and Communications (SIBCON), Moscow, Russia, 12–14 May 2016; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2016; pp. 1–5. [\[CrossRef\]](#)
17. Hu, W.; Miao, H. Sub-pixel displacement algorithm in temporal sequence digital image correlation based on correlation coefficient weighted fitting. *Opt. Lasers Eng.* **2018**, *110*, 410–414. [\[CrossRef\]](#)
18. Duadi, D.; Ozana, N.; Shabairou, N.; Wolf, M.; Zalevsky, Z.; Primov-Fever, A. Non-contact optical sensing of vocal fold vibrations by secondary speckle patterns. *Opt. Express* **2020**, *28*, 20040–20050. [\[CrossRef\]](#)
19. Liushnevskaya, Y.D.; Gubarev, F.A.; Li, L.; Nosarev, A.V.; Gusakova, V.S. Measurement of whole blood coagulation time by laser speckle pattern correlation. *Biomed. Eng.* **2020**, *54*, 262–266. [\[CrossRef\]](#)
20. Shen, L.; Li, D.G.; Luo, F. Study on laser speckle correlation method applied in triangulation displacement measurement. *J. Opt. Technol.* **2013**, *80*, 604–610. [\[CrossRef\]](#)
21. Chen, Z.; Wang, C.; Huang, C.; Fu, H.; Luo, H.; Wang, H. Audio signal reconstruction based on adaptively selected seed points from laser speckle images. *Opt. Commun.* **2014**, *331*, 6–13. [\[CrossRef\]](#)
22. Zhu, G.; Yao, X.R.; Qiu, P.; Mahmood, W.; Yu, W.K.; Sun, Z.B.; Zhai, G.J.; Zhao, Q. Sound recovery via intensity variations of speckle pattern pixels selected with variance-based method. *Opt. Eng.* **2018**, *57*, 026117. [\[CrossRef\]](#)
23. Zhu, G.; Yao, X.-R.; Sun, Z.-B.; Qiu, P.; Wang, C.; Zhai, G.-J.; Zhao, Q. A High-Speed Imaging Method Based on Compressive Sensing for Sound Extraction Using a Low-Speed Camera. *Sensors* **2018**, *18*, 1524. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Wu, N.; Haruyama, S. Real-time sound detection and regeneration based on optical flow algorithm of laser speckle images. In Proceedings of the 2019 28th Wireless and Optical Communications Conference (WOCC), Beijing, China, 9–10 May 2019; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2019; pp. 1–4. [\[CrossRef\]](#)
25. Wu, N.; Haruyama, S. Real-time audio detection and regeneration of moving sound source based on optical flow algorithm of laser speckle images. *Opt. Express* **2020**, *28*, 4475–4488. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Barcellona, C.; Halpaap, D.; Amil, P.; Buscarino, A.; Fortuna, L.; Tiana-Alsina, J.; Masoller, C. Remote recovery of audio signals from videos of optical speckle patterns: A comparative study of signal recovery algorithms. *Opt. Express* **2020**, *28*, 8716–8723. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Gregory, D.A. Basic physical principles of defocused speckle photography: A tilt topology inspection technique. *Opt. Laser Technol.* **1976**, *8*, 201–213. [\[CrossRef\]](#)
28. Tiziani, H.J. A study of the use of laser speckle to measure small tilts of optically rough surfaces accurately. *Opt. Commun.* **1972**, *5*, 271–276. [\[CrossRef\]](#)
29. Reu, P.L. Experimental and numerical methods for exact subpixel shifting. *Exp. Mech.* **2010**, *51*, 443–452. [\[CrossRef\]](#)
30. Guizar-Sicairos, M.; Thurman, S.T.; Fienup, J.R. Efficient subpixel image registration algorithms. *Opt. Lett.* **2008**, *33*, 156–158. [\[CrossRef\]](#)