

Digital Speckle Pattern Heterodyne Interferometry: Computational Focusing Compared with Optical Focusing

Okan Atalar
Stanford University
okan@stanford.edu

Abstract

For many applications, there is a need for non-contact detection of single frequency vibrations on a surface with wide-field of view and picometer level detection sensitivity. Standard approaches rely on scanning a small laser beam or having bulky and complex setups for producing multiple small beams in an optical interferometric configuration. These approaches are not scalable and have their own drawbacks. This paper investigates the use of a CMOS camera for detecting minute displacements on a surface by using a heterodyne interferometer with a multi-pixel CMOS camera as the detector. The measured surface can be spatially resolved by an imaging lens or computationally. Both approaches are investigated and compared. Picometer level detection sensitivity at a meter distance is demonstrated by the interferometric setup.

1. Introduction

Many applications rely on the non-contact detection of single frequency motion on a surface with subnanometer detection sensitivity. Some applications which rely on such a technique include the testing of a wide array of electromechanical systems, or non-contact imaging applications relying on the detection of ultrasound [1]. Non-contact imaging systems actively excite an opaque target and monitor the surface displacements caused by the dielectric contrast inside the material (if there is any). By solving the ultrasound propagation equation, hidden targets can be reconstructed by monitoring the surface displacement profile. Both applications usually require subnanometer level detection sensitivity.

The standard approach for non-contact detection of motion is to use a laser Doppler vibrometer (LDV) [2]. The LDV consists of a Mach-Zehnder interferometer and measures the displacement profile of a single point on a surface by sending a small laser beam, called the object beam, to the surface. The phase of the object beam is modulated by

the motion on the surface, which is detected by interfering the reflected photons from the surface with a controlled reference beam on a photodetector. The recorded intensity is a function of the phase difference between the two beams, and therefore the motion on the surface can be estimated by demodulating the recorded intensity as a function of time.

To extend the application space of the standard LDV to monitor multiple points (wide-field), the laser beam can be scanned over the surface. This approach, however, has a significant drawback in terms of the signal to noise ratio (SNR), and ultimately the detection sensitivity due to time multiplexing between different points on the surface. Another approach is to generate multiple laser beams with different temporal frequencies by using multiple frequency shifters, usually acousto optic modulators (AOMs) to spatially encode position in temporal frequency, similar to the operation of magnetic resonance imaging (MRI) [3,4]. This approach, however, results in overly bulky and complex systems if a reasonable spatial resolution is desired.

In this paper, the use of a standard CMOS camera as the detector for spatially resolving the motion on a surface is investigated, assuming that the vibration profile on the surface is single frequency. The spatial vibration profile of the monitored surface can be formed either by using an imaging lens in front of the CMOS sensor or through computational methods, relying on solving the diffraction equation governing the propagation of light. To detect high frequency vibrations, heterodyne interferometry is used for downconverting the high frequency to a measurable low frequency that falls within the temporal bandwidth of a CMOS camera. Heterodyne detection is analogous to the operation of the superheterodyne receiver for radio frequency (RF) circuits. In the rest of this paper, we introduce the interferometric setups to measure subnanometer motion in a wide-field of view with a CMOS camera, based on using a lens in front of the camera sensor and through computational methods. We compare the two modalities and experimentally demonstrate the capability of measuring picometer level displacements experimentally.

2. Related Work

The standard LDV consists of a Mach-Zehnder interferometer. A coherent light source (usually a laser) is split into two beams, reference and object beams, after passing through a beam splitter. The standard LDV is demonstrated in Figure 1. If $x(t)$ is the motion on a single point on the monitored surface, the instantaneous phase of the object beam is expressed as $\psi(t) = \frac{4\pi}{\lambda}x(t) + \phi$, where ϕ is the static phase difference between the two beams and λ the wavelength of laser light. The object beam can be expressed as $I_{obj}(t) = \sqrt{I_{obj}}e^{jw_L t}e^{j\psi(t)}$, where w_L is the temporal oscillation frequency corresponding to the laser light. By interfering the reference and object beams using another beam splitter in front of the photodetector, the recorded intensity becomes a function of the phase difference $\psi(t)$ between the two beams, which can be demodulated to estimate the motion ($x(t)$):

$$I_{tot}(t) = I_{ref} + I_{obj} + 2\sqrt{I_{ref}I_{obj}}\cos(\psi(t)) \quad (1)$$

To extend the application range of the standard LDV to monitor multiple points on a surface, the laser beam can be scanned. However, this results in lower SNR, leading to significantly lower detection sensitivity assuming photon shot-noise limited measurement scheme. Another approach is to spatially encode position in frequency by using multiple frequency shifters. This approach is not scalable if a reasonable spatial resolution is desired due to the many frequency shifters required. In this approach, there exists multiple object beams $I_{obj_k} = \sqrt{I_{obj_k}}e^{j(w_L+w_k)t}e^{j\psi_k(t)}$ for $k \in \{1, 2, \dots, n\}$. By spacing the temporal frequency differences w_k sufficiently apart, the spatial vibration profile of each point is encoded in different frequency bands, similar to the operation of frequency division multiplexing in communication systems.

There are also techniques relying on time of flight (ToF) cameras to record motion in a wide-field of view. The scene is actively illuminated with an amplitude modulated light source. The motion in the scene modulates the frequency of light linearly proportional to the velocity. By having an electronic demodulator in the ToF camera, the Doppler shift can be estimated per pixel to determine the motion in the scene in a wide-field of view [5]. These methods rely on the amplitude modulation frequency to detect the Doppler shift, which is equal to $\frac{2v(x,y)}{\lambda_s}$, where λ_s is the amplitude modulation frequency (usually on the order of Megahertz) and $v(x, y)$ the velocity of the scene corresponding to pixel location (x, y) . If coherent detection is performed via interferometry, λ is in the nanometer range whereas for Megahertz amplitude modulation frequencies λ_s is in the meters range. Although having a larger wavelength is advantageous for having a larger unambiguous range measurement,

the sensitivity of the measurement is proportional to sensitivity to phase changes: $\psi(t) = \frac{4\pi}{\lambda}x(t)$. Since interferometric techniques have a wavelength approximately six orders of magnitude smaller compared to ToF camera based techniques, they are approximately six orders of magnitude more sensitive to motion. Therefore, ToF cameras are not sensitive enough compared to interferometric techniques to measure subnanometer vibrations.

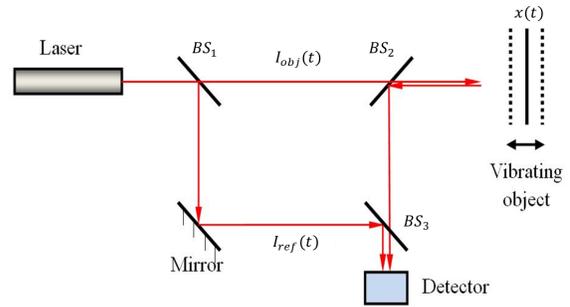


Figure 1. Standard Mach-Zehnder Interferometer. BS_1 , BS_2 , BS_3 : Beam Splitters.

3. Heterodyne Detection with a CMOS Camera

The frequency of vibrations of interest for most applications vastly exceed the temporal bandwidth of standard CMOS cameras. CMOS cameras have frame rates that go upto several hundred Hertz, whereas some applications require detection of megahertz (MHz) frequencies. To detect such high frequencies, heterodyne detection can be performed, which is similar in operation to the superheterodyne receiver in RF circuits. The information carrying carrier frequency is downconverted to a lower intermediate frequency (IF) for further processing. Similarly, by upshifting the frequency of the reference beam with a phase modulator $I_{ref}(t) = \sqrt{I_{ref}}e^{j(w_L+w_{AOM1}-w_{AOM2})t}$, usually an AOM, the beat frequency can be freely adjusted so that it falls within the temporal bandwidth of the camera. Usually two AOMs in series are required due to the high center frequency for AOMs of reasonable size. For the experimental demonstration, two AOMs with center frequencies of 80MHz with bandwidths of 25MHz were used.

4. Optical Configuration: Optical Focusing

The optical configuration for the digital speckle pattern interferometer with a focusing camera lens is based on a Mach-Zehnder interferometer. To operate in heterodyne measurement mode, the reference beam is upshifted in frequency by $w_{AOM1} - w_{AOM2}$ through using two AOMs,

AOM1 and AOM2 in series. The object beam I_{obj} is focused to the objects by using the lens L_1 . The variable attenuator VA allows the reference and object beam levels to be freely adjusted by also varying the laser power or camera exposure time. Lens L_2 is used to form a plane wave (constant intensity and phase) beam on the camera sensor after passing through the focusing lens of the camera L_3 . Lens L_3 also performs as a focusing lens to focus to piezoelectric actuators PZT_1 and PZT_2 . The optical configuration is demonstrated in Figure 2.

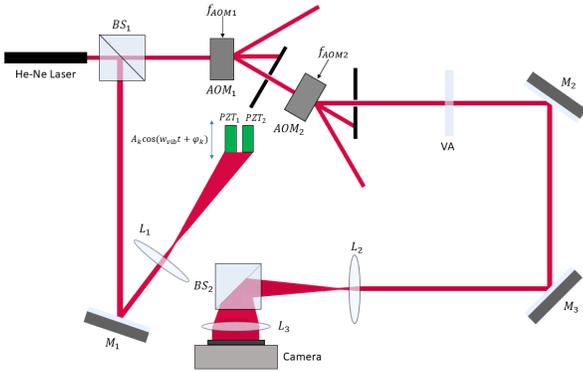


Figure 2. Optical Configuration. BS_1 , BS_2 : Beam Splitters. L_1 , L_2 , L_3 : Lenses. AOM_1 and AOM_2 : Bragg Cells. VA: Variable Attenuator. M_1 to M_3 : Mirrors. PZT_1 and PZT_2 : Piezoelectric Actuators.

5. Calibration Measurement for Amplitude Normalization and Speckle Phase Removal: Optical Focusing

The object beam forms a speckle pattern on the detector surface after reflecting from the optically rough object surface. The random amplitude for the speckle pattern is modeled through $\sqrt{I_{obj_{xy}}}$ and the speckle phase ϕ_{xy} as the two dimensional position on the image sensor (x, y) . Speckle pattern statistics has been studied extensively and it is well known that the phase ϕ_{xy} is uniformly distributed over the interval $[0, 2\pi)$ and intensity $I_{obj_{xy}}$ follows an exponential distribution. More details regarding the statistics of speckle patterns can be found in [6]. Turning our attention to the case of a sinusoidal motion of an object at a well defined temporal frequency with out of plane vibration profile $A\cos(w_{vib}t + \psi)$, the phase modulation on the object beam causes discrete set of frequencies to be observed, each equally spaced by the mechanical vibration frequency w_{vib} . Using the Jacobi-Anger expansion, the phase modulated object beam can be expressed as a discrete set of frequencies

scaled by the n^{th} Bessel function of the first kind:

$$I_{obj_{xy}}(t) = \sqrt{I_{obj_{xy}}} e^{j(w_L + \phi + \frac{4\pi}{\lambda} A \cos(w_{vib}t + \psi))} \quad (2)$$

$$= \sqrt{I_{obj_{xy}}} e^{jw_L t} \sum_{n=-\infty}^{\infty} j^n J_n(A) e^{j(n(w_{vib}t + \psi))} \quad (3)$$

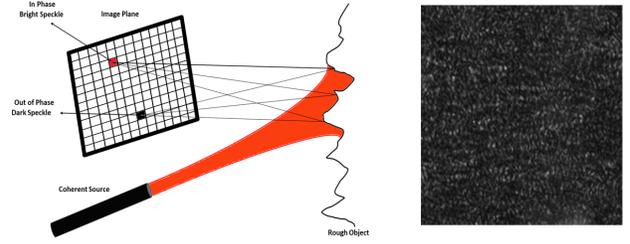


Figure 3. Speckle Pattern Formation (Left), Recorded Speckle Pattern on Camera (Right).

The limited temporal bandwidth of the camera, on the order of tens of Hertz (Hz), does not allow the capture of any of the discrete set of frequencies. Therefore, the reference beam is frequency shifted by two AOMs in series to perform heterodyne detection. Taking the $+1$ order from AOM1 and the -1 order from AOM2, the reference beam can be expressed as $I_{ref}(t) = \sqrt{I_{ref}} e^{j(w_L + w_{AOM1} - w_{AOM2})t}$. Heterodyne detection can now be performed to make the beat term $w_{beat1} = w_{AOM1} - w_{AOM2} - w_{obj}$ fall within the temporal bandwidth of the camera. The interference term $I_{motion_{xy}}(t) = |I_{ref}(t) + I_{obj_{xy}}(t)|^2$ can be expressed as:

$$I_{motion_{xy}}(t) = I_{ref} + I_{obj_{xy}} + 2\sqrt{I_{ref}I_{obj_{xy}}} J_1\left(\frac{4\pi}{\lambda} A\right) \cos((w_{beat}t + \phi_{xy} + \psi)) \quad (4)$$

The received signal is integrated over the exposure time T of the camera sensor with angular sampling rate w_F , leading to the discrete recorded signal:

$$I_{motion_{xy}}[n] = (I_{ref} + I_{obj_{xy}})T + 2\sqrt{I_{ref}I_{obj_{xy}}} J_1\left(\frac{4\pi}{\lambda} A\right) \text{sinc}\left(\pi \frac{w_{beat}}{w_F}\right) \cos\left(\frac{n2\pi}{w_F} w_{beat} + \phi_{xy} + \psi\right) \quad (5)$$

The amplitude A and phase ψ carry the amplitude and phase information on the monitored object. The detected signal, however, is scaled by a random factor $\sqrt{I_{obj_{xy}}}$ and

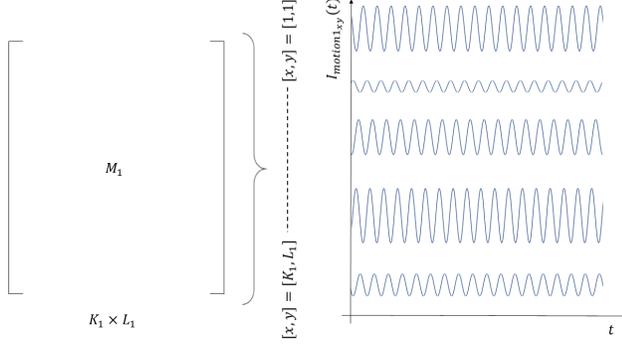


Figure 4. Random Amplitude and Phase for Recorded Pixel Signals as a Function of Time

the phase is corrupted by the speckle phase ϕ_{xy} . The random phase and amplitude of the recorded signal for different pixel locations are demonstrated in Figure 4. To remove these effects and have reliable readings on the amplitude and phase of concern, an additional calibration measurement will be made. This calibration measurement will measure the speckle phase matrix and normalization object constant of the pixels corresponding to the respective object in the scene.

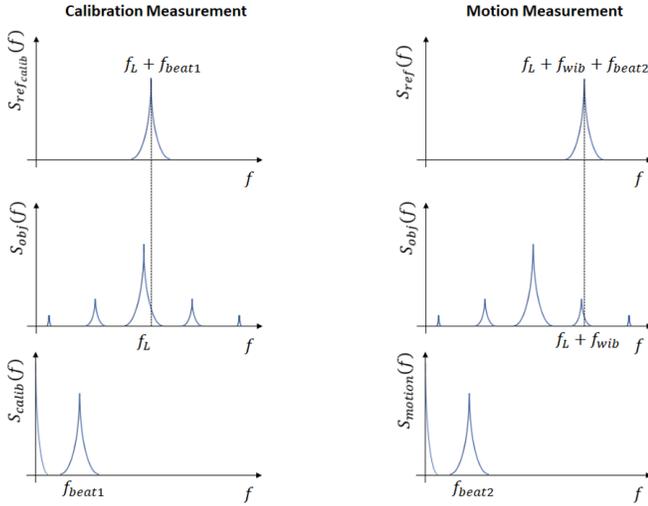


Figure 5. Temporal Power Spectrum of Calibration Measurement and Motion Measurement on Camera Sensor

We make the calibration measurement by changing the AOM frequencies such that $w_{beat2} = w_{AOM1} - w_{AOM2} \sim Hz$. The beat term that falls within the temporal bandwidth of the camera is now the zero order Bessel function of the first kind. Figure 5 shows the temporal spectrum measured by the camera for the calibration and motion measurements.

Based on the interference, the new recorded signal on each pixel of the camera can be expressed as:

$$I_{calib_{xy}}(t) = |I_{ref_{calib}}(t) + I_{obj_{xy}}(t)|^2 = I_{ref_{calib}} + I_{obj_{xy}} + 2\sqrt{I_{ref}I_{obj_{xy}}}J_0\left(\frac{4\pi}{\lambda}A\right)\cos((w_{beat2}t + \phi_{xy})) \quad (6)$$

$$I_{calib_{xy}}[n] = \int_{\frac{n2\pi}{w_F} - \frac{T}{2}}^{\frac{n2\pi}{w_F} + \frac{T}{2}} I_{calib_{xy}}(t)dt = (I_{ref_{calib}} + I_{obj_{xy}})T + 2\sqrt{I_{ref}I_{obj_{xy}}}J_0\left(\frac{4\pi}{\lambda}A\right)\text{sinc}\left(\pi\frac{w_{beat2}}{w_F}\right)\cos\left(\frac{n2\pi}{w_F}n + \phi_{xy}\right) \quad (7)$$

The amplitude and phase can now be recovered by quadrature amplitude (QAM) demodulation. Let N_1 represent the number of frames captured during the calibration measurement. The speckle phase matrix and the normalization constant are estimated as follows:

$$\phi_{xy}^* = \text{atan}\left\{\frac{\sum_{n=1}^{N_1} I_{calib_{xy}}[n]\sin\left(\frac{n2\pi}{w_F}n\right)}{\sum_{n=1}^{N_1} I_{calib_{xy}}[n]\cos\left(\frac{n2\pi}{w_F}n\right)}\right\} \quad (8)$$

$$S^* = 2\left|\sum_{n=1}^{N_1} I_{calib_{xy}}[n]e^{j\left(\frac{n2\pi}{w_F}n\right)}\right| \quad (9)$$

6. Recording and Processing of Frames: Optical Focusing

The calibration and motion measurements are carried out sequentially. The focusing lens in front of the camera allows to match the speckle pattern clusters formed on the image sensor with the respective objects in the object plane. Assuming two different objects with out of plane vibration profiles $A_1\cos(w_{vib}t + \psi_1)$ and $A_2\cos(w_{vib}t + \psi_2)$, we need to determine the normalization factors corresponding to the speckle patterns clusters formed for the two objects P_1 and P_2 , respectively. We also need to hold the speckle phase matrix corresponding to the speckle pattern formed on the image sensor. The calibration measurement allows the determination of the normalization constants P_1 , P_2 , and the speckle phase matrices M_1 and M_2 . Assuming K_1 by L_1 pixels on the image sensor capturing the speckle cluster formed by the first object, the speckle phase matrix elements are determined by applying the formula given in Eq. 8, yielding the speckle phase estimate ϕ_{xy}^* . The normalization constant $P_1 = \sum_{n=1}^{n=K_1L_1} S_{xy}^*$, where S_{xy}^* is estimated according to Eq. 9.

To start acquiring frames for the motion measurement, the AOM frequencies are changed so that $w_{beat2} \sim Hz$. To significantly improve the detection sensitivity, pixels corresponding to the same object can be spatially averaged. Assuming photon shot-noise limited regime, many pixels carry

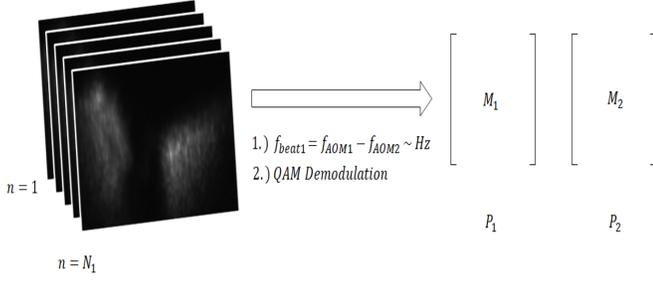


Figure 6. Calibration Measurement for Retrieving Normalization Constants P_1 , P_2 and Speckle Phase Matrices M_1 , M_2 for Two Different Piezoelectric Actuators

the same signal but have uncorrelated photon shot-noise. By performing coherent spatial averaging on these pixels after removing the speckle phases, the SNR can be improved on average by the square root of the number of pixels used for spatial averaging.

There are two different possible processing techniques for spatial coherent averaging of the pixels. To minimize the storage of pixels for storage limited conditions, a shift-add-discard method can be used. This method uses the speckle phase matrix M_1 to give the appropriate shift to the incoming values. For each object monitored in the field-of-view, a vector of size N_2 is initialized with all zero values, assuming N_2 frames are acquired in the motion measurement. The incoming values from each frame is added to the respective position based on the speckle phase matrix M_1 calculated as $n_{xy} = \text{round}(\frac{\phi_{xy}^*}{2\pi} \frac{w_F}{w_{beat2}})$. This processing pipeline allows the frames to be discarded once they are used. However, addition of the signals in time domain leads effectively to phase errors since the discrete shift is most often rounded. This effect can be modeled as phase noise with statistics $Unif(-\pi \frac{w_{beat1}}{w_F}, \pi \frac{w_{beat1}}{w_F})$.

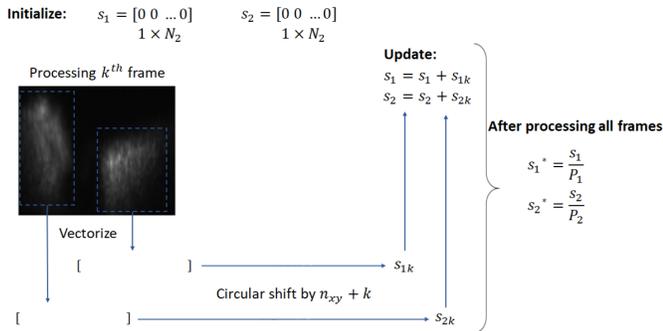


Figure 7. Real Time Processing of Frames by Shift-Add-Discard Method

The second method is to collect all the frames and then

to perform a temporal Fast Fourier Transform (FFT) to remove the speckle phases and perform the coherent addition in the frequency domain. This method eliminates the rounding error introduced by adding discrete signals in the time domain. However, all the frames need to be present to allow for taking an FFT for each pixel. Additionally, this method is computationally more costly, since $K_1 L_1$ many FFTs of length N_2 need to be performed, compared to only taking a single FFT of length N_2 for the first described method. Details of phase noise and the resulting SNR can be found in Appendix A.

7. Computational Focusing

The previous method used a focusing lens in front of the camera sensor to naturally form the image of the monitored objects on the camera sensor. An alternative method is to computationally form the image without using a lens in front of the camera sensor. If the distance of the monitored objects to the camera sensor z is significantly greater than the size of the monitored objects, then the Fresnel diffraction equation can be used to approximately model the propagation of light from the monitored objects to the image sensor.

$$F(k_x, k_y) = \frac{e^{jkz}}{j\lambda z} \int_{-\infty}^{\infty} f(x, y) e^{-j\frac{k}{2z}((k_x-x)^2+(k_y-y)^2)} dx dy \quad (10)$$

Implementation of the Fresnel Transform in its current form is a convolution with the free-space propagation kernel $h(x, y, z) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}(x^2+y^2)}$ under the Fresnel approximation. $F(k_x, k_y)$ is the light field incident on the camera sensor and $f(x, y)$ the reflected light from the monitored objects, $k = \frac{2\pi}{\lambda}$ is the wavenumber. This has computational complexity of n^2 , however, Fresnel transforms can also be implemented through FFTs, which have computational complexities on the order of $n \log(n)$ [6,7].

$$F(k_x, k_y) = \mathcal{F}_{2D}\{f(x, y) e^{j\frac{k}{2z}(x^2+y^2)}\} \quad (11)$$

The expression above excludes constant phase and scaling factors. The processing pipeline is very similar to the focusing lens method. After the speckle phases and normalization constants have been determined, these will be used for correcting each pixel in the motion measurement. The slight difference is between working in the Fourier space rather than the spatial domain. The speckle phases in this case will be removed by dividing the respective phasors.

The optical configuration for the computational focusing method is shown in Figure 8. Compared with Figure 2, the image forming lens is removed and the beam splitter in front of the image sensor is slightly tilted (introducing spatial carrier frequency to the reference beam) to allow

the separation of the light field from other components after performing numerical backpropagation with the Fresnel transform. The light field intensity incident on the camera sensor can be expressed as:

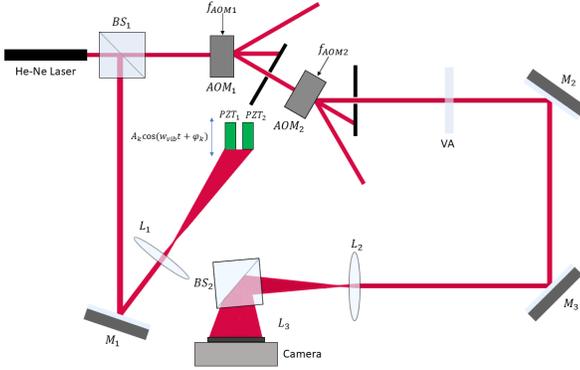


Figure 8. Optical Configuration for Computational Method. BS_1 , BS_2 : Beam Splitters. L_1 , L_2 : Lenses. AOM_1 and AOM_2 : Bragg Cells. VA: Variable Attenuator. M_1 to M_3 : Mirrors. PZT_1 and PZT_2 : Piezoelectric Actuators.

$$I_{tot_{xy}}(t) = I_{ref} + I_{obj_{xy}} + \sqrt{I_{ref}I_{obj_{xy}}} [e^{j\psi(t)} e^{j(ax+by)} + e^{-j\psi(t)} e^{-j(ax+by)}] \quad (12)$$

The constant intensity term $I_{ref} + I_{obj_{xy}}$ forms a "zero-order" term centered around zero frequency in the spatial domain after performing the Fresnel transform. The information carrying +1 and -1 orders are the terms $\sqrt{I_{ref}I_{obj_{xy}}} e^{j\psi(t)} e^{j(ax+by)}$ and $\sqrt{I_{ref}I_{obj_{xy}}} e^{-j\psi(t)} e^{-j(ax+by)}$. Choosing the +1 order to work with, this forms a computational image of the actual object after the Fresnel transform. The processing pipeline outlined for the optical focusing method could be performed on this image, treating it as if were the real image of the object.

A computer generated speckle pattern interfered with a reference beam with an appropriate carrier frequency ($ax + by$) is demonstrated in Figure 9. The encircled regions by the boxes indicate the +1 and -1 orders, whereas the center portion is the 0 order term.

8. Computational Focusing Compared with Optical Focusing

There are several fundamental differences between the two image formation methods. The focusing lens allows a smaller portion of the camera sensor to be used by focusing photons collected by the lens to a small area on the image sensor. The frame rate of most cameras is limited by the

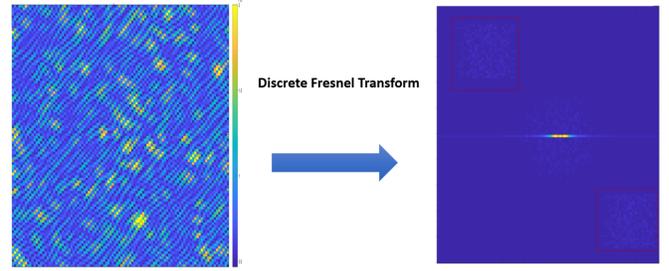


Figure 9. Simulated Interference of Speckle Pattern and Reference Beam with Spatial Carrier Frequency (Left), Corresponding Discrete Fresnel Transform (Right)

readout rate of each row. By accessing a smaller portion of the pixels, the frame rate can be increased. This can especially be useful if multi frequency measurements are taken, where orthogonal frequency division multiplexing (OFDM) [7] or similar modulation schemes requiring fine frequency resolution is required.

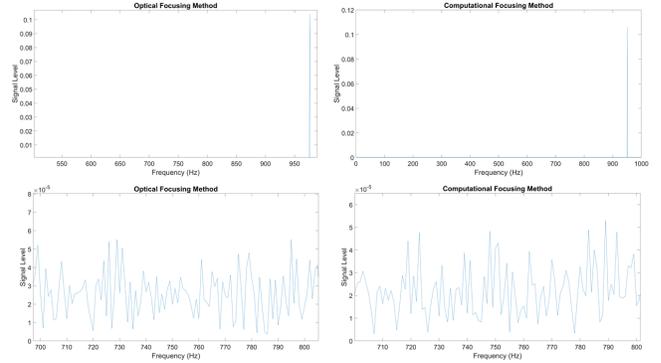


Figure 10. SNR Comparison of Optical and Computational Focusing from Simulated Data. The signal is at frequency 975Hz for Optical Focusing Method and 950Hz for Computational Focusing Method. The Signal Levels (Top 2 Plots) and Photon Shot-Noise Levels (Bottom 2 Plots) Measured in the Frequency Range (700-800)Hz are Approximately Equal for the Two Methods.

Another difference is based on the number of photons collected and the spatial resolution. The lens size could be chosen to allow more photons to be collected by the image sensor. The lens functions as the new aperture, and therefore allows more photons to be collected, leading to higher SNR. The larger lens also implies higher spatial resolution. Additionally, operating in the off-axis mode by using a spatial carrier frequency comes at the cost of reduced spatial resolution, since only some of the pixels are devoted for the +1 order beam.

The computational method, however, offers its own advantages as well. Computational focusing allows the flex-

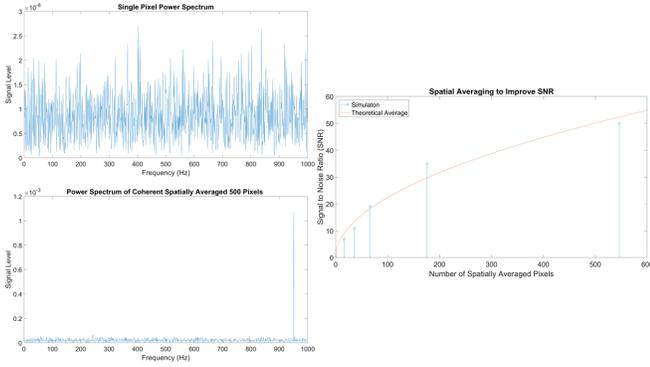


Figure 11. SNR as a Function of Number of Spatially Averaged Pixels from Simulated Data. Only After Spatial Averaging the 950Hz Signal Appears

ibility to arbitrarily focus to any desired plane, not limited by the focusing lens. The average speckle size on the sensor is proportional to $\frac{\lambda z}{D}$, where D is the size of the aperture. The size of speckle is the full width at half maximum of the autocorrelation function for the speckle patterns. Since the lens method has a larger aperture and a smaller distance between the lens (which acts as the new source) and the image sensor, the speckle sizes are significantly smaller. This is a disadvantage for measurement conditions under which there are undesired vibrations. Since the computational method has significantly larger speckle sizes, and therefore a more uniform autocorrelation function, the phases and amplitudes of nearby pixels are very similar. Therefore, even if random vibrations cause the speckles to randomly move on the image sensor, the measured intensities and phases will alter significantly less compared with an optical focusing method using a lens. Therefore, computational focusing allows robustness to the system to environmental noise in the form of vibrations. Finally, the SNR obtained for both methods are the same, if the larger aperture of the lens is neglected (leading to more photons to be captured). The detailed derivation is in Appendix B. Figure 10 shows the simulations results for the SNR obtained by the two methods assuming photon shot-noise limited regime and coherent spatial averaging over all the pixels. The equal SNR obtained for the two different methods agrees with the theoretical derivations in Appendix B.

9. Experimental Results

The optical configuration demonstrated in Figure 2 was used to test the detection capabilities of the system. Two piezos from Thorlabs (model AE0203D04F) were driven at 100kHz with amplitudes of 5nm and the camera was placed one meters away from the piezos.

Calibration and motion measurements were made to

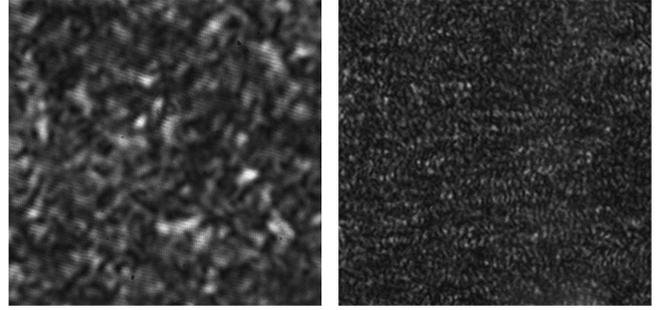


Figure 12. Recorded Speckle Patterns Without Imaging Lens (Left) and with an Imaging Lens with Focal Length 7.5cm. The Imaged Object is a Piezo at 1 meter Distance from Camera Sensor

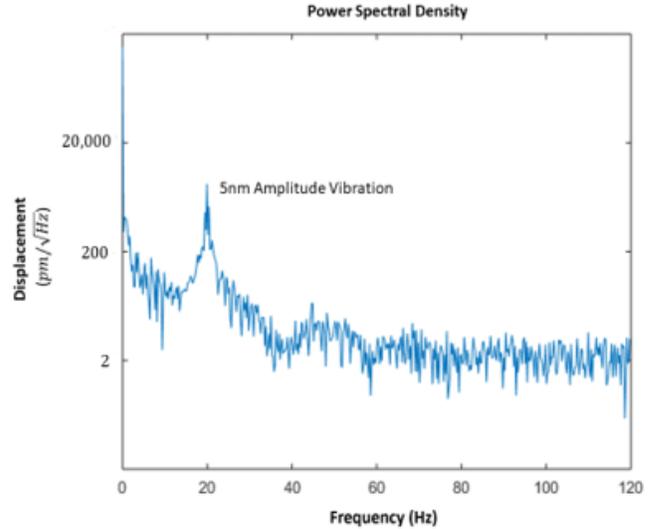


Figure 13. Power Spectral Density of Vibrating Piezo with Amplitude of 5nm at 100kHz. Downconverted to 20Hz by Heterodyne Detection

record the motion and the obtained spectrum after following the signal processing techniques outlined in the previous section (relying on time-domain coherent addition) is demonstrated in Figure 13. The 5nm amplitude vibration is clearly visible and the white noise floor corresponds to the photon shot-noise. This level is $2pm/\sqrt{Hz}$. Therefore, the detection system is capable of detecting picometer level displacements in a matter of seconds from a meter distance.

The computational method was also tested experimentally by using the setup shown in Figure 8. The computationally generated image of a monitored single Thorlabs piezo (model AE0203D04F) is shown in Figure 14.

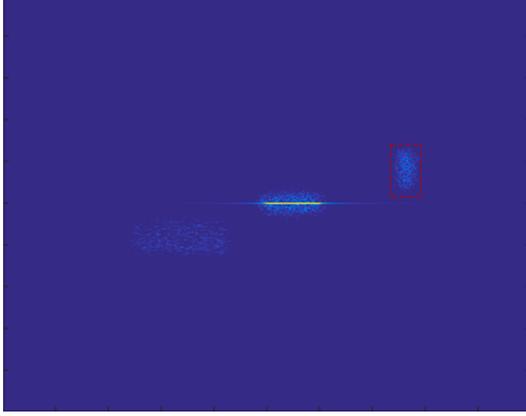


Figure 14. Fresnel Transform of Captured Image of Piezo at one meter Distance to Camera. The +1 Order Corresponding to the Computationally Generated Image of the Piezo is Highlighted with Red Rectangle

10. Conclusion and Future Directions

The interferometric configurations and signal processing algorithms for a heterodyne interferometer using a CMOS camera as the sensor for both optical focusing and computational focusing have been presented. The two configurations were experimentally shown to be sensitive to picometer level displacements in a wide-field of view even when operating a meter away from monitored objects. The two modalities have also been compared.

Future work could involve making the calibration and motion measurements simultaneously by driving the AOMs with two different frequencies. This would have advantages if there are random vibrations. The speckle phases measured would be useless if there are significant vibrations if the calibration and motion measurements are taken sequentially. By measuring both the speckle phases and the motion simultaneously, the vibrations could be removed in post-processing. Additionally, multiple frequencies could also be measured by driving the AOMs with multiple frequencies to downconvert each frequency to a different frequency in the temporal spectrum of the camera, similar to the operation of OFDM.

11. References

- [1] Aranchuk, V., Lal, A. K., Hess, C. F., & Sabatier, J. M. (2006). Multi-beam laser Doppler vibrometer for landmine detection. *Optical Engineering*, 45(10), 104302.
- [2] Rothberg, S. J., Allen, M. S., Castellini, P., Di Maio, D., Dirckx, J. J. J., Ewins, D. J., ... & Steger, H. (2017). An international review of laser Doppler vibrometry: Making light work of vibration measurement. *Optics*

and Lasers in Engineering, 99, 11-22.

- [3] Fu, Y., Guo, M., & Phua, P. B. (2011). Multipoint laser Doppler vibrometry with single detector: principles, implementations, and signal analyses. *Applied optics*, 50(10), 1280-1288.
- [4] Nishimura, D. G. (2010). Principles of magnetic resonance imaging. Stanford Univ.
- [5] Heide, F., Heidrich, W., Hullin, M., & Wetzstein, G. (2015). Doppler time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(4), 36.
- [6] Goodman, J. W. (2007). Speckle phenomena in optics: theory and applications. Roberts and Company Publishers.
- [7] Li, Y. G., & Stuber, G. L. (Eds.). (2006). Orthogonal frequency division multiplexing for wireless communications. Springer Science & Business Media.
- [8] Schnars, U., & Jüptner, W. P. (2002). Digital recording and numerical reconstruction of holograms. *Measurement science and technology*, 13(9), R85.
- [9] Verrier, N., & Atlan, M. (2011). Off-axis digital hologram reconstruction: some practical considerations. *Applied optics*, 50(34), H136-H146.

12. Appendix A

For sequential calibration and motion measurements and for a given fixed measurement time ($N_1 + N_2 = N^*$), how should N_1 and N_2 be chosen to maximize the SNR of the measurement, assuming shot-noise limited measurement. We now derive the optimal N_1 and N_2 for SNR maximization by evaluating the phase error caused by photon shot-noise. We begin by evaluating the phase error incurred when estimating the speckle phase ϕ_{xy} . Assuming photon shot-noise limited regime:

$$I_{calibxy}[n] = (I_{ref} + I_{objxy})T + 2T\sqrt{I_{ref}I_{objxy}}\cos(\phi_{xy} + \frac{n2\pi}{w_F}w_{beat1}) + X \quad (13)$$

Where $X \sim \mathcal{N}(0, (I_{ref} + I_{objxy})T)$ is the shot-noise (Poisson noise converges to a Gaussian noise for sufficiently large number of photons).

$$\phi_{xy}^* = \text{atan}\left\{\frac{\text{Im}\{D\} + X_1}{\text{Re}\{D\} + X_2}\right\} \quad (14)$$

Where $D = \sqrt{I_{ref}I_{objxy}}e^{j\phi_{xy}}TN_1$, $\text{Re}\{FFT\{X\}\} = X_1 \sim \mathcal{N}(0, (I_{ref} + I_{objxy})T\frac{N_1}{2})$ and $\text{Im}\{FFT\{X\}\} = X_2 \sim \mathcal{N}(0, (I_{ref} + I_{objxy})T\frac{N_1}{2})$, where X_1 and X_2 are

independent. We are interested in the variance of the estimated phase ϕ_k^* :

To simplify the derivation, we choose D to be purely real. This does not affect the result since we are at liberty in choosing the phase reference. This simplifies to (also assuming the argument of the atan function is small, since signal is very large compared to noise):

$$\text{var}(\text{atan}\{\frac{X_1}{\text{Re}\{D\} + X_2}\}) \approx \text{var}(\frac{X_1}{\text{Re}\{D\} + X_2}) \quad (15)$$

$$\begin{aligned} \text{var}(\frac{X_1}{\text{Re}\{D\} + X_2}) &= \mathbb{E}[(\frac{x_1}{D + X_2})^2] \approx \\ \mathbb{E}[(\frac{x_1}{D})^2] &= \frac{(I_{ref} + I_{obj})}{2TN_1 I_{ref} I_{obj}} \end{aligned} \quad (16)$$

Therefore, the estimated speckle phase is distributed as follows: $\phi_{xy}^* \sim \mathcal{N}(\phi_{xy}, \frac{(I_{ref} + I_{obj_{xy}})}{2TN_1 I_{ref} I_{obj_{xy}}})$

12.1. Effect of Rounding on Phase Error

Even if the speckle phase is estimated perfectly, there will be errors when actually shifting the discrete signal for coherent averaging in time-domain. This rounding operation will have its own error. This will be equivalent to a phase error with distribution $\psi_{xy}^* \sim \text{Unif}[\frac{-\pi w_{beat1}}{w_F}, \frac{\pi w_{beat1}}{w_F}]$. If $\frac{4\pi}{\lambda} A \ll 1$, $J_1(\frac{4\pi}{\lambda} A) \approx \frac{2\pi}{\lambda} A$.

$$I_{motion_{xy}}[n] \approx (I_{ref} + I_{obj_{xy}})T + \frac{4\pi}{\lambda} AT \text{sinc}(\frac{\pi w_{beat}}{w_F}) \sqrt{I_{ref} I_{obj_{xy}}} \cos(\frac{n2\pi}{w_F} w_{beat} + \phi_{xy} + \psi) + X \quad (17)$$

$$\begin{aligned} I_{eff}[n] &= \sum_{x,y=1,1}^{K,L} I_{motion_{xy}}[n - \text{round}(\frac{\phi_{xy}^*}{2\pi} \frac{w_F}{w_{beat1}})] = \\ &(KLI_{ref} + \sum_{x,y=1,1}^{K,L} I_{obj_{xy}})T + \frac{4\pi}{\lambda} AT \text{sinc}(\frac{\pi w_{beat}}{w_F}) \sqrt{I_{ref}} \\ &\cos(\frac{n2\pi}{w_F} w_{beat1} + \phi_{noise} + \psi) | \sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j(\phi_{xy} - \phi_{xy}^* + \psi_{xy})} | \end{aligned} \quad (18)$$

Due to the Gaussian distribution of the estimated phase ϕ_{xy}^* and the rounding error phase ψ_{xy}^* , there is an amplitude and a phase error. Next, the relevant statistics of the phase error ϕ_{noise} and the amplitude error will be determined. If the recorded signal is demodulated by an FFT to retrieve the amplitude of motion A and its phase

ψ , there will be variations about these mean values. Let $\phi_{xy}^- = \phi_{xy} - \phi_{xy}^* \sim \mathcal{N}(0, \frac{(I_{ref} + I_{obj_{xy}})}{2TN_1 I_{ref} I_{obj_{xy}}})$.

$$\phi_{noise} = \angle \sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j(\phi_{xy} - \phi_{xy}^* + \psi_{xy})} \quad (19)$$

$$\begin{aligned} \mathbb{E}[|FFT\{I_{eff}[n]\}|]_{w=w_{beat1}} &= \\ \mathbb{E}[N_2 \frac{2\pi}{\lambda} AT \text{sinc}(\frac{\pi w_{beat}}{w_F}) \sqrt{I_{ref}} | \sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j(\phi_{xy}^- + \psi_{xy})}] &= \\ N_2 \frac{2\pi}{\lambda} AT \text{sinc}(\frac{\pi w_{beat}}{w_F}) \sqrt{I_{ref}} K L \mathbb{E}[| \sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j(\phi_{xy}^- + \psi_{xy})} |] & \end{aligned} \quad (20)$$

$$\begin{aligned} \mathbb{E}[|\sqrt{I_{obj_{xy}}} e^{j(\phi_{xy}^- + \psi_{xy})}|] &= \\ \int_x \mathbb{E}[|\sqrt{I_{obj_{xy}}} e^{j(\phi_{xy}^- + \psi_{xy})}| | \sqrt{I_{obj_{xy}}} = x] \mathbb{P}\{\sqrt{I_{obj_{xy}}} = x\} dx & \end{aligned} \quad (21)$$

$$\begin{aligned} \int_x \mathbb{E}[|\sqrt{I_{obj_{xy}}} \phi_{xy}^- | \sqrt{I_{obj_{xy}}} = x] \mathbb{P}\{\sqrt{I_{obj_{xy}}} = x\} dx &= \\ \int_x x \mathbb{P}(\sqrt{I_{obj_{xy}}} = x) \frac{I_{ref} + x^2}{2TN_1 I_{ref} x^2} dx & \end{aligned} \quad (22)$$

$$\begin{aligned} \int_x x \mathbb{P}(\sqrt{I_{obj_{xy}}} = x) \frac{I_{ref} + x^2}{2TN_1 I_{ref} x^2} dx &= \\ \frac{1}{2TN_1 I_{ref}} \int_x \mathbb{P}(\sqrt{I_{obj_{xy}}} = x) \frac{I_{ref} + x^2}{x} dx & \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{1}{2TN_1 I_{ref}} \int_x \mathbb{P}(\sqrt{I_{obj_{xy}}} = x) \frac{I_{ref} + x^2}{x} dx &= \\ \frac{1}{2TN_1 I_{ref}} \{ \mathbb{E}[\sqrt{I_{obj_{xy}}}] + I_{ref} \mathbb{E}[\sqrt{\frac{1}{I_{obj_{xy}}}}] \} & \end{aligned} \quad (24)$$

$$\begin{aligned} \mathbb{E}[\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \sin(\psi) \sin(\phi_{xy}^- + \psi_{xy}^*)] &= \\ \sin(\psi) \mathbb{E}[\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} (\phi_{xy}^- + \psi_{xy}^*)] & \end{aligned} \quad (25)$$

$$\begin{aligned} \mathbb{E}[|FFT\{I_{eff}[n]\}|]_{w=w_{beat1}} &= \\ N_2 \frac{2\pi}{\lambda} AT K L \text{sinc}(\frac{\pi w_{beat1}}{w_F}) \sqrt{I_{ref}} \{ \mathbb{E}[\sqrt{I_{obj_{xy}}}] & \\ (1 - \frac{1}{4TN_1 I_{ref}} - \frac{\pi^2 w_{beat1}^2}{6w_F^2}) - \frac{1}{4TN_1} \mathbb{E}[\frac{1}{\sqrt{I_{obj_{xy}}}}] \} & \end{aligned} \quad (26)$$

If shift-add-discard method is not used, the expected value of the signal will be:

$$\begin{aligned} & \mathbb{E}[|FFT\{I_{eff}[n]\}|]_{w=w_{beat1}} = \\ & N_2 \frac{2\pi}{\lambda} ATKLSinc\left(\frac{\pi w_{beat1}}{w_F}\right) \sqrt{I_{ref}} \{\mathbb{E}[\sqrt{I_{obj_{xy}}}] \\ & \left(1 - \frac{1}{4TN_1 I_{ref}}\right) - \frac{1}{4TN_1} \mathbb{E}\left[\frac{1}{\sqrt{I_{obj_k}}}\right]\} \end{aligned} \quad (27)$$

12.2. Phase Error Statistics

The phase error ϕ_{noise} is random, and in this part the expected value and variance will be determined.

Eq. 5 to determine $I_{motion_{xy}}[n]$ can be viewed as a random walk in the two dimensional complex plane. With this in mind:

$$\phi_{noise} = \frac{Im\{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j\phi_{xy}}\}}{Re\{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j\phi_{xy}}\}} \quad (28)$$

Where we have again chosen the phase reference as zero for simplicity.

$$\begin{aligned} \mathbb{E}[\phi_{noise}] & \approx \mathbb{E}\left[\frac{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \sin(\phi_{xy})}{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}}}\right] \approx \\ & \frac{\sum_{x,y=1,1}^{K,L} \mathbb{E}[\sqrt{I_{obj_{xy}}} \sin(\phi_{xy})]}{KLE[\sqrt{I_{obj_{xy}}}] } = 0 \end{aligned} \quad (29)$$

$$\begin{aligned} Var(\phi_{noise}) & = \mathbb{E}\left[\left(\frac{Im\{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j\phi_{xy}}\}}{Re\{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} e^{j\phi_{xy}}\}}\right)^2\right] \approx \\ & \mathbb{E}\left[\left(\frac{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \phi_{xy}}{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}}}\right)^2\right] \end{aligned} \quad (30)$$

Using the approximation:

$$\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \approx KLE[\sqrt{I_{obj_{xy}}}] \quad (31)$$

$$\begin{aligned} & \mathbb{E}\left[\left(\frac{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \phi_{xy}}{\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}}}\right)^2\right] \frac{1}{(KLE[\sqrt{I_{obj_{xy}}}]^2)} \\ & \mathbb{E}\left[\left(\sum_{x,y=1,1}^{K,L} \sqrt{I_{obj_{xy}}} \phi_{xy}\right)^2\right] = \frac{1}{(KLE[\sqrt{I_{obj_{xy}}}]^2)} \\ & \mathbb{E}\left[\sum_{x,y=1,1}^{K,L} I_{obj_{xy}} \phi_{xy}^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{KL} \sqrt{I_{ref} I_{obj_{xy}}} \phi_i \phi_j\right] \end{aligned} \quad (32)$$

$$\begin{aligned} & \frac{1}{(KLE[\sqrt{I_{obj_{xy}}}]^2)} \mathbb{E}\left[\sum_{x,y=1,1}^{K,L} I_{obj_{xy}} \phi_{xy}^2\right] \\ & + 2 \sum_{i,j=1,1}^K \sqrt{I_{ref} I_{obj_i} \phi_i \phi_j} = \\ & \frac{KL}{(KLE[\sqrt{I_{obj_{xy}}}]^2)} \mathbb{E}[I_{obj_{xy}} \phi_{xy}^2] \end{aligned} \quad (33)$$

$$\begin{aligned} & \frac{1}{KL(\mathbb{E}[\sqrt{I_{obj_{xy}}}]^2)} \mathbb{E}[I_{obj_{xy}} \phi_{xy}^2] = \\ & \frac{1}{2TN_1 I_{ref} (\mathbb{E}[I_{obj_{xy}}])^2} (\mathbb{E}[I_{obj_{xy}}] + I_{ref} \mathbb{E}\left[\frac{1}{I_{obj_{xy}}}\right]) \end{aligned} \quad (34)$$

Therefore, the statistics for the phase noise ϕ_{noise} are:

$$\mathbb{E}[\phi_{noise}] = 0 \quad (35)$$

$$Var(\phi_{noise}) = \frac{1}{2TN_1 I_{ref} (\mathbb{E}[I_{obj_{xy}}])^2} (\mathbb{E}[I_{obj_{xy}}] + I_{ref} \mathbb{E}\left[\frac{1}{I_{obj_{xy}}}\right]) \quad (36)$$

12.3. Optimization of Calibration Measurement and Measurement Time

Given a fixed time for measurement ($N_1 + N_2$ is fixed), there is a constraint on N_1 and N_2 in the form $N_1 + N_2 = N^*$. Given N^* , I_{ref} and the distribution of $I_{obj_{xy}}$, we aim to find the optimal N_1 and N_2 to optimize $\mathbb{E}[|FFT\{I_{eff}[n]\}|]$. The optimization problem can be expressed as:

$$\begin{aligned} N_1^* & = \operatorname{argmax}_{N_1} (N^* - N_1) \frac{2\pi}{\lambda} ATKLSinc\left(\frac{\pi w_{beat1}}{w_F}\right) \sqrt{I_{ref}} \left\{\mathbb{E}\left[\sqrt{I_{obj_{xy}}}\right] \right. \\ & \left. \left(1 - \frac{1}{4TN_1 I_{ref}} - \frac{\pi^2 w_{beat1}^2}{6w_F^2}\right) - \frac{1}{4TN_1} \mathbb{E}\left[\frac{1}{\sqrt{I_{obj_{xy}}}}\right]\right\} \end{aligned} \quad (37)$$

Expanding the terms and making the assumptions $\frac{0.5}{N_1^3} \ll 1$ and $\frac{0.5}{\frac{N_1 N^*}{4}} \ll 1$:

$$N_1^* = \sqrt{\frac{N^*}{4T(1 - \frac{\pi^2 w_{beat1}^2}{6w_F^2})}} \sqrt{\frac{1}{I_{ref}} + \frac{\mathbb{E}\left[\frac{1}{\sqrt{I_{obj_k}}}\right]}{\mathbb{E}[\sqrt{I_{obj_k}}]}} \quad (38)$$

For the case of not using shift-add-discard method but coherent averaging in frequency domain:

$$N_1^* = \sqrt{\frac{N^*}{4T}} \sqrt{\frac{1}{I_{ref}} + \frac{\mathbb{E}[\frac{1}{\sqrt{I_{objk}}}]^2}{\mathbb{E}[\sqrt{I_{objk}}]}} \quad (39)$$

For small number of frames N^* , N_1^* can turn out to be smaller than 4. However, at least 4 samples are required for the calibration measurement since there are three unknowns: $I_{ref} + I_{obj_{xy}}$, $2\sqrt{I_{ref}I_{obj_{xy}}}$, and ϕ_{xy} (the phase is inside a cosine for the actual term, requiring two different phases for overcoming ambiguity). Eq. 39 is an expression for the optimal number of frames for the calibration measurement from an SNR perspective, without taking into account that at least 4 samples are required to solve to get the speckle phase ϕ_{xy} .

13. Appendix B

13.1. Computational Focusing SNR

$$I_{tot_{xy}} = I_{ref} + I_{obj_{xy}} + 2\sqrt{I_{ref}I_{obj_{xy}}}\cos(w_1t + \phi + \frac{4\pi}{\lambda}A\cos(w_2t + \Delta\phi)) + X \quad (40)$$

X is the shot-noise contribution and is modeled as:

$$X \sim \mathcal{N}(0, I_{ref_{xy}} + I_{obj_{xy}}) \quad (41)$$

$$\begin{aligned} \cos(w_1t + \phi + \frac{4\pi}{\lambda}A\cos(w_2t + \Delta\phi)) &= \\ \cos(w_1t + \phi)\cos(\frac{4\pi}{\lambda}A\cos(w_2t + \Delta\phi)) - & \\ \sin(w_1t + \phi)\sin(\frac{4\pi}{\lambda}A\cos(w_2t + \Delta\phi)) & \quad (42) \end{aligned}$$

Using Jacobi-Anger expansion:

$$\begin{aligned} \cos(w_1t + \phi)(J_0(\frac{4\pi}{\lambda}A) + 2\sum_{n=1}^{\infty}(-1)^n J_{2n}(\frac{4\pi}{\lambda}A)) & \\ \cos(2n(w_2t + \Delta\phi)) + \sin(w_1t + \phi)(2\sum_{n=1}^{\infty}(-1)^n & \\ J_{2n-1}(\frac{4\pi}{\lambda}A)\cos((2n-1)(w_2t + \Delta\phi)) & \\ \approx \cos(w_1t + \phi) - \sin(w_1t + \phi)\frac{4\pi}{\lambda}A\cos(w_2t + \Delta\phi) & \\ = \cos(w_1t + \phi) - \frac{4\pi}{\lambda}\frac{A}{2}[\sin(w_1t + w_2t + \phi + \Delta\phi) + & \\ \sin((w_1 - w_2)t + \phi + \Delta\phi)] & \quad (43) \end{aligned}$$

If we focus on a single pixel at a given time (neglecting time dependence for now):

$$I_{tot_{xy}} = I_{ref} + I_{obj_{xy}} + \frac{2\pi}{\lambda}A(\sqrt{I_{ref}I_{obj_{xy}}}e^{j(\psi(x,y)-ax-by)} + \sqrt{I_{ref_{xy}}I_{obj_{xy}}}e^{j(-\psi(x,y)+ax+by)}) + X \quad (44)$$

Using the Fresnel Transform, it can be shown that the spatial phase distribution $\psi(x, y)$ is band-limited and therefore can be expressed as the sum of a band-limited number of complex exponentials.

$$e^{j\psi(x,y)} = \int_{-k_{x_{lim}}}^{k_{x_{lim}}} \int_{-k_{y_{lim}}}^{k_{y_{lim}}} C(k_x, k_y)e^{j(k_x x + k_y y)} dk_x dk_y \quad (45)$$

$C(k_x, k_y)$ is the two dimensional spatial Fourier Transform coefficients of the phase variation across the image sensor $\psi(x, y)$. The Fourier coefficients have random amplitude and phase. From a k-space perspective, this is due to the fact that for a fixed sensor size (aperture), closely spaced frequencies cannot be distinguished. Since different frequencies have random phase due to the optical roughness of the monitored object, this leads to a random walk for determining each Fourier component $C(k_x, k_y)$. Assuming a uniform rectangular object:

$$\mathbb{E}[|C(k_x, k_y)|] = \begin{cases} C, & |k_x| < k_{x_{lim}}, |k_y| < k_{y_{lim}} \\ 0, & \text{otherwise} \end{cases} \quad (46)$$

Now we will calculate the value C by invoking Parseval's Theorem. Assuming we have M_1 pixels by M_2 pixels on the camera sensor:

$$M_1 M_2 = \frac{1}{M_1 M_2} 4C^2 k_{x_{lim}} k_{y_{lim}} \quad (47)$$

$$C = \frac{M_1 M_2}{2\sqrt{k_{x_{lim}} k_{y_{lim}}}} \quad (48)$$

After performing a 2D FFT to calculate the discrete Fresnel transform for reconstructing the discrete hologram from recorded interference pattern $I_{tot_{xy}}$ over all pixels of the camera sensor:

$$\mathbb{E}[FFT2\{\frac{2\pi}{\lambda}A\sqrt{I_{ref}I_{obj_{xy}}}e^{j(\psi(x,y)-ax-by)}\}] = \frac{M_1 M_2 \sqrt{I_{ref}I_{obj_{xy}}}}{2\sqrt{k_{x_{lim}} k_{y_{lim}}}} \frac{2\pi}{\lambda} A \quad (49)$$

$$X_1 = Re\{FFT2\{X\}\} \sim \mathcal{N}(0, \frac{M_1 M_2 (I_{ref} + I_{obj_{xy}})}{2}) \quad (50)$$

$$X_2 = \text{Im}\{FFT2\{X\}\} \sim \mathcal{N}(0, \frac{M_1 M_2 (I_{ref} + I_{obj_{xy}})}{2}) \quad (51)$$

We now calculate the expected noise level for the white noise coming from shot-noise:

$$\begin{aligned} \mathbb{E}[|X_1 + jX_2|] &\sim \\ \text{Rayleigh}(\frac{\sqrt{\pi M_1 M_2 (I_{ref} + I_{obj_{xy}})}}{2}, & \\ \frac{4 - \pi}{4} M_1 M_2 (I_{ref} + I_{obj_{xy}})) & \end{aligned} \quad (52)$$

The SNR is calculated by the ratio of the magnitude of the actual signal carrying phasor to the expected value of shot-noise contribution at that spatial frequency.

$$SNR = \frac{\sqrt{M_1 M_2 I_{ref} I_{obj_{xy}}} \frac{4\pi}{\sqrt{\pi\lambda}} A}{\sqrt{I_{ref} + I_{obj_{xy}}} 2\sqrt{k_{x_{lim}} k_{y_{lim}}}} \quad (53)$$

This SNR assumes that we have total of $4k_{x_{lim}} k_{y_{lim}}$ pixels available after discrete Fresnel transform to spatially average. For simplicity, lets assume that we only have a single pixel, so that $4k_{x_{lim}} k_{y_{lim}} = 1$. In this case, the total SNR we have from processing N frames (averaging over N frames) can be expressed as:

$$SNR_{total} = \frac{\sqrt{N M_1 M_2 I_{ref} I_{obj_{xy}}} \frac{4\pi}{\sqrt{\pi\lambda}} A}{\sqrt{I_{ref} + I_{obj_{xy}}}} \quad (54)$$

13.2. Optical Focusing SNR

$$\begin{aligned} I_{tot_{xy}} &= I_{ref} + I_{obj_{xy}} + \\ 2\sqrt{I_{ref} I_{obj_{xy}}} \cos(w_1 t + \phi) + \frac{4\pi}{\lambda} A \cos(w_2 t + \Delta\phi) &+ X \end{aligned} \quad (55)$$

X is the shot-noise contribution and is modeled as:

$$X \sim \mathcal{N}(0, I_{ref_{xy}} + I_{obj_{xy}}) \quad (56)$$

$$\begin{aligned} \cos(w_1 t + \phi) + \frac{4\pi}{\lambda} A \cos(w_2 t + \Delta\phi) &= \\ \cos(w_1 t + \phi) \cos(\frac{4\pi}{\lambda} A \cos(w_2 t + \Delta\phi)) & \\ - \sin(w_1 t + \phi) \sin(\frac{4\pi}{\lambda} A \cos(w_2 t + \Delta\phi)) & \end{aligned} \quad (57)$$

Using Jacobi-Anger expansion:

$$\begin{aligned} &\cos(w_1 t + \phi) (J_0(\frac{4\pi}{\lambda} A) + \\ &2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(\frac{4\pi}{\lambda} A) \cos(2n(w_2 t + \Delta\phi)) \\ &+ \sin(w_1 t + \phi) (2 \sum_{n=1}^{\infty} (-1)^n J_{2n-1}(\frac{4\pi}{\lambda} A) \cos((2n-1)(w_2 t + \Delta\phi)) \\ &\approx \cos(w_1 t + \phi) - \sin(w_1 t + \phi) \frac{4\pi}{\lambda} A \cos(w_2 t + \Delta\phi) \\ &= \cos(w_1 t + \phi) - \frac{4\pi}{\lambda} \frac{A}{2} [\sin(w_1 t + w_2 t + \phi + \Delta\phi) + \\ &\sin((w_1 - w_2)t + \phi + \Delta\phi)] \end{aligned} \quad (58)$$

For a single pixel, the recorded signal as a function of time is (assuming narrow temporal low pass filter which only passes the temporal frequency component $(w_1 - w_2)$):

$$I_{tot_{xy}}(t) = I_{ref} + I_{obj_{xy}} + \sqrt{I_{ref} I_{obj_{xy}}} \frac{4\pi}{\lambda} A \cos((w_1 - w_2)t + \phi + \Delta\phi) + X \quad (59)$$

Taking the FFT of this pixel signal with respect to time and averaging for N frames:

$$\mathbb{E}[FFT\{I_{tot_{xy}}(t)\}] = \sqrt{I_{ref} I_{obj_{xy}}} N \frac{2\pi}{\lambda} A \quad (60)$$

Noise Analysis:

$$X \sim \mathcal{N}(0, I_{ref_{xy}} + I_{obj_{xy}}) \quad (61)$$

After taking 1D FFT (temporal):

$$X_1 = \text{Re}\{FFT\{X\}\} \sim \mathcal{N}(0, \frac{N}{2} (I_{ref_{xy}} + I_{obj_{xy}})) \quad (62)$$

$$X_2 = \text{Im}\{FFT\{X\}\} \sim \mathcal{N}(0, \frac{N}{2} (I_{ref_{xy}} + I_{obj_{xy}})) \quad (63)$$

We now calculate the expected noise level for the white noise coming from shot-noise:

$$\mathbb{E}[|X_1 + jX_2|] \sim \text{Rayleigh}(\frac{\sqrt{\pi N (I_{ref} + I_{obj_{xy}})}}{2}, \frac{4 - \pi}{4} N (I_{ref} + I_{obj_{xy}})) \quad (64)$$

The SNR is calculated by the ratio of the magnitude of the actual signal carrying phasor to the expected value of shot-noise contribution at that spatial frequency.

$$SNR = \frac{\sqrt{N I_{ref} I_{obj_{xy}}} \frac{4\pi}{\sqrt{\pi\lambda}} A}{\sqrt{I_{ref} + I_{obj_{xy}}}} \quad (65)$$

Averaging spatially over the M_1 and M_2 pixels after correcting for the phases to do coherent averaging:

$$SNR_{total} = \frac{\sqrt{N M_1 M_2 I_{ref} I_{obj_{xy}}} \frac{4\pi}{\sqrt{\pi\lambda}} A}{\sqrt{I_{ref} + I_{obj_{xy}}}} \quad (66)$$

We finally see that the SNR obtained by the two different methods are equal. However, this analysis neglected shot-noise aliasing and saturation.