Compressive super-localization

by

Yi Liu

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Signature redacted

Author ............................................................... ..............................
Department of Mechanical Engineering
Nov 6, 2015

Signature redacted

Certified by..... ..........................................................

George Barbastathis
Professor
Thesis Supervisor

Signature redacted

Accepted by ........

Rohan Abeyaratne
Quentin Berg Professor of Mechanics, Chairman, Department
Committee on Graduate Students
DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.
Compressive super-localization

by

Yi Liu

Submitted to the Department of Mechanical Engineering on Nov 6, 2015, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Localization accuracy is a fundamental quantity of an imaging system, since it often determines the performance (e.g. effective resolution, sensitivity) of the system. This thesis studies the influence on localization accuracy of finite sampling rate, number of samples, and noise in the data acquisition process. Two classes of super-localization problems will be investigated.

The first class of problem aims to improve the accuracy in localizing and tracking the physical position of an object of interest. For lateral localization problems, the key parameters to consider are the pixel size as compared to the size of the point spread function (PSF) of the imaging system, and to the amount of motion of the object of interest. If the pixel size is larger than PSF or object motion, accurate localization essentially becomes solving an under-sampling problem. In this thesis, a compressive holography algorithm is proposed to localize smaller than pixel size motion. 1/45 sub-pixel motion has been successfully detected.

For axial localization problems, the localization accuracy will be affected not only by the pixel size and number of pixels, but also by the imaging geometry. Traditional axial localization methods are fundamentally limited by the finite numerical aperture of the optical system. In this thesis, a class of compressive reconstruction method that exploits the “sparse” prior knowledge about the object in order to alleviate the missing angular information has been investigated and 1/16 depth of field (DOF) axial displacement was successfully extracted.

Successful implementation of the compressive holography based super-localization technique has been applied to image biomimetic sensors inspired by harbor sea seal whiskers for studying vortex-induced vibrations and wake-induced vibrations.

The second class of super-localization problem under investigation is to detect weak signals buried under strong background and noise. A compressive reconstruction method that is able to detect signals captured with extremely low signal-to-noise ratio (SNR) and signal-to-background ratio (SBR), by exploiting different “sparsities” in the respective signal and background subspaces has been demonstrated.
Thesis Supervisor: George Barbastathis
Title: Professor
Acknowledgments

The past five years have been a great journey to me. I would like to thank all the people that help make this trip wonderful.

I would like to thank, first and foremost, my Ph.D. advisor, Professor George Barbastathis. I came to MIT without any optics or electro-magnetics background. He has been warmly and patiently guiding me during the course of my research and giving me many helpful suggestions and encouragements. Furthermore, he has been a great friend to me, not only discussing research topics, but also sharing personal life experiences.

Also, I would like to thank my Ph.D. thesis committee meeting members: Professor Berthold Horn and Professor Michael S. Triantafyllou. Their suggestions at both committee meetings and individual meetings on the thesis document greatly improved this thesis. Thanks for being such great mentors to me.

I would also like to thank all my colleagues at the 3D Optical Systems group for all their help and many helpful discussions: Chih-Hung Max Hsieh, Justin Lee, Howard Yuanhao Huang, Hanhong Gao, Yuihui Zhu, Shuai Li, Nilu Zhao, Adam Pan, Wensheng Chen and so on. Especially, I would like to thank Max Hsieh, Justin Lee and Hanhong Gao for their many helps since I came here. They gave me much helpful guidance on optics, signal processing as well as my personal life.

To Heather Beem, my collaborator for one of the research projects. When we first collaborated on the whisker project, I knew nothing about Fluid Mechanics. Heather taught me a lot about how the vibration will act in fluid and helped me build the
optical experiments together. She has been a great collaborator and friend to me. Without her help, I do not think I can finish the whisker project.

Also, I would like to thank all my friends at MIT, especially Wen Zheng, Xue Feng, Xiao Fang, Lidan Wu, Qiong Ma and Xun Cai. I spent lots of memorable time with them and they are the people that I can call whenever I have a problem or want to have fun. Xun provided me lots of helpful suggestions in data processing.

Moreover, I would like to thank my dear husband: Lei Tian. We met because of optics. At the beginning, he was my lab mentor and then we fell in love. We have been together for almost five years and I can surely say that for every day that is spent with him, it is like Valentine’s day to me.

I would also like to thank my parents for their ever–lasting love, support, and encouragement. Thanks for always being there for me no matter what happens.

Finally, thanks to all the funding sources: the National Research Foundation (NRF) of Singapore through the Singapore-MIT Alliance for Research and Technology (SMART) Center for Environmental Sensing and Modeling (CENSAM), the Chevron Energy Technology Company and the Xerox company for offering me the Xerox MIT fellowship as well as the summer internship opportunity.
# Contents

1 Introduction 19

1.1 Super-localization problem ................................. 19

1.2 Classical sampling: Nyquist–Shannon theorem ............... 22

1.3 New paradigm of sampling: compressed sensing .............. 23

1.3.1 Sparsity .................................................. 25

1.3.2 Incoherence ............................................... 26

1.3.3 Examples of incoherent measurements ...................... 27

1.3.4 Number of required measurements for compressive reconstruction 29

1.3.5 Why $\ell_1$ minimization? ................................ 30

1.4 Traditional point-to-point imaging ............................ 31

1.5 Introduction to digital holography ............................ 32

1.6 Outline of the thesis ........................................ 35

2 One dimensional sub-pixel movement detection using compressive holography 37

2.1 Incoherence and sparsity in digital holography ............... 38

2.2 Theoretical model for compressive holography ............... 40
2.3 1D sub-pixel motion detection experimental setup

2.4 Experimental results for 1D sub-pixel motion detection

3 2D sub-pixel movement detection using spiral phase filtering and compressive holography

3.1 Spiral phase mask in edge extraction

3.2 Implementing the spiral phase mask to 2D compressive holography model

3.3 Experimental results

3.4 Discussion

4 Axial super-localization with sub-depth-of-field accuracy via compressive pattern matching

4.1 Classical limit of axial localization accuracy

4.2 Introduction to pattern matching

4.3 Axial super-localization via compressive pattern matching

4.4 Axial sub-DOF displacement tracking experiments

4.5 Result and discussion

5 Vortex and wake-induced vibration characterization using compressive holography

5.1 Introduction to whisker sensing in harbor seals

5.2 Whisker's vortex-induced vibration characterization using compressive holography

5.3 Vortex-induced vibration characterization results
5.3.1 VIV response comparison between whisker, ellipse and cylinder models .......................................................... 89
5.3.2 VIV response comparison between 3\times and 30\times-scale whisker models .................................................. 92
5.4 Whisker in vortex wake: wake–induced vibration characterization ................................................................. 94
  5.4.1 Wake–induce vibration characterization experiments ................................................................. 95
  5.4.2 Detected Wake–induced vibration among different models ....................................................... 96
6 Weak signal detection via null–space estimation and sparse reconstruction .......................................................... 101
  6.1 Introduction to weak signal detection .......................................................... 101
  6.2 Null–space based background–signal separation .......................................................... 103
  6.3 Acoustic crack detection for underwater platforms .......................................................... 108
  6.4 Fine crack detection simulation ........................................................................ 114
    6.4.1 Low SNR and SBR crack signal detection ........................................................................ 114
    6.4.2 Detection rate v.s. SNR ........................................................................ 116
7 Conclusion ........................................................................................................................................ 123
List of Figures

1-1 Traditional point-to-point imaging system. ........................................ 32

1-2 Schematic of a lensless in-line DH setup. ........................................ 33

2-1 (a) Vector representation of the 1D object; (b) The sparse derivative of the object. ........................................ 39

2-2 In-line DH geometry. The object was moved along $y$ direction with a uniform step size of 267nm(1/45-pixel) by a piezo-driven motion stage. 41

2-3 Experimental in-line DH setup. The pin object was moved along $y$ direction with a uniform step size of 267nm(1/45-pixel) by a piezo-driven motion stage. ........................................ 43

2-4 A sample 2D hologram. ......................................................... 44

2-5 A sample row vector from the 2D hologram in Figure 2-4. ............. 45

2-6 Real part of the compressive reconstruction of the 1D hologram in Figure 2-5. ........................................ 46
2-7 Compressive reconstructed positions of seven randomly chosen rows (in blue circles) and the "true" position (in red stair curve) at each step. The histogram on the bottom right combines the data of total 45 steps taken by the seven rows (45 x 7 = 315 data points total.)

2-8 The comparison between the average position (purple points) of each step and the "true" positions.

3-1 (a) A square object; (b) the derivative of the square along the horizontal direction; (c) the 2D derivative of the square.

3-2 Phase distribution of spiral phase mask

3-3 In-line DH geometry. The object was moved along x and y directions by a piezo-driven motion stage.

3-4 Subpixel notation. In this example, each camera pixel is divided into 4 x 4 pixels. \((n_x, n_y)\) denotes the camera pixel index and \((m_x, m_y)\) represents the pixel index after subpixel division.

3-5 (a) Real part of effective transfer function for edge-extraction holography; (b) Real part of transfer function for traditional in-line DH.

3-6 Summary of the computational reconstruction procedure, together with representative intermediate reconstruction results from an actual experiment with a star object.
3-7 (a) Microscopic image of a star shape binary amplitude object; (b) Hologram of the star object; (c) Compressive reconstruction of star hologram; (d) Traditional back-propagation of star hologram; (e) Image of a ring shape amplitude object; (f) Hologram of the ring; (g) Compressive reconstruction of the ring hologram; (h) Traditional back-propagation of the ring hologram; (i) “M” shape phase object; (j) Hologram of the “M”; (k) Compressive reconstruction of “M” hologram; (l) Traditional back-propagation of “M” hologram.

3-8 (a) Star trajectory characterization from compressive localization algorithm; (b) The histogram of estimated step size distribution using compressive reconstruction; (c) Star trajectory characterization using standard DIC approach; (d) The histogram of the estimated step size using standard DIC approach. In both (a) and (c), red dots indicate the localization results from the corresponding hologram using either method, and the blue “Z” pattern indicates the ground truth of the motion established using the piezo stage.

4-1 An example of single lens imaging system with object placed at $2f$.

4-2 4F imaging system. An object $g(x)$ is placed $\Delta z$ from the object plane and the imaging plane captures a defocused image of the object.

4-3 (a) Frequency modulation for a mild defocus system; (b) Frequency modulation for a strong defocus system.

4-4 A example of pattern matching.
4-5 In-line digital holography geometry. ........................................ 71
4-6 Digitally reconstructed images from the hologram shown in Fig. 4-5.
   (a) in-focus, (b) over-focus, (c) under-focus. ......................... 71
4-7 Algorithm. ................................................................. 73
4-8 Illustration of pattern matching applied to an axial stack of digitally
   reconstructed image. The reconstruction plane that generates the high-
   est cross-correlation peak value is considered as the object position. . 74
4-9 Experimental setup of in-line DH for axial localization demonstration.
   An object is placed at a motion stage, which is moved along the optical
   axis with step size equivalent to 1/16 of the DOF of the imaging system.
   A sample hologram is taken at each axial position. ................. 75
4-10 (a) A microscopic image of the star object used in the experiment; (b)
   a sample hologram of the star object taken in the in-line DH setup. . 76
4-11 Reconstruction of the hologram taken with star object at its initial
   position. The reconstruction at \( z = 112 \text{mm} \) has been chosen as the
   template for subsequent pattern matching. ......................... 76
4-12 Pattern matching result. Images on the second row are the reconstruc-
   tion results of the hologram taken after a single step displacement.
   Plots on the bottom show the normalized correlation results between
   the template and each reconstruction on different axial planes. ....... 77
4-13 The axial movement detection result by the proposed algorithm. .... 78
4-14 The axial movement detection result by conventional back propagation
   approach. ................................................................. 79
(a) A harbor seal displaying its whiskers; (b) SEM pictures of a whisker highlighting its undulatory, elliptical shape. Shed whisker specimen was provided through the courtesy of the New England Aquarium.

Image geometry of an in-line digital holography set up that is used to image whisker’s transverse vibration. The model was towed along \( z \) axis during the experiments.

Schematic of lab DH setup for imaging whisker’s VIV.

Images of the three plastic, 3D-printed models used in the tests. Each model was held by a piece of silica gel at the base and hanged vertically in the water. (a) 3D undulatory whisker model, with an elliptical cross-section; (b) 3D elliptical model with smooth edges; (c) 3D cylindrical model with smooth edges.

(a) A sample hologram of the bottom part of the plastic whisker model. (b) The compressive edge reconstruction of the sample whisker hologram shown in Fig. 5-5(a).

A sample plot of detected VIV response of the whisker model. The whisker was placed along 90 degree orientation and towed at a speed of 200mm/s.

The VIV responses comparison among all the three models, with 0 and 90 degree orientations.
5-8 (a) The measured VIV on 3x whisker model, using compressive holography approach. (b) The detected VIV response of the 30x whisker model, measured by strain gauges. Each model was positioned in its streamlined orientation (0 degree).

5-9 The 30x scaled-up whisker model used for VIV detection via strain measurement. Experiments were conducted by Dr. Heather Beem. The model whisker was mounted on a set of flexing plates that allow free vibration in in-line and crossflow directions. Bushings are used to reduce coupling between the axes.

5-10 Illustration of the relative positions between the upstream cylinder and the downstream whisker model in the WIV characterization experiments.

5-11 An image of laboratory WIV measurement setup.

5-12 Comparison of VIV and WIV responses among the three models. The results shown here are all with the models positioned along their streamlined orientation (0 degree).
5-13 Comparison of VIV and WIV responses measured on different scaled size models. For both models, the whisker vibrate with far higher amplitude in the wake of upstream body than it does on its own. (a) Detected vibration responses for the $3\times$-scale whisker model, measured by compressive holography approach. The whisker is within the wake of an upstream cylinder ($d_{cy} = 4d_w$) at a distance $d_x = 17.5d_w$, for WIV tests. (b) Characterized vibration responses of the $30\times$-scale model. The whisker is within the wake of an upstream cylinder ($d_{cy} = 2.5d_w$) at a distance $d_x = 8d_w$, for WIV tests.

5-14 The detected reduced WIV frequency for $3\times$ whisker model. When the wake of the upstream cylinder becomes strong, the peak frequency of the whisker’s vibration would synchronize with the frequency of the oncoming wake.

6-1 A lightening-shaped crack appears on the surface of a platform, whose surface is covered with a thick layer of marine growth.

6-2 The underwater crack detection setup using acoustic wave.

6-3 The autonomous crack detection algorithm flow.

6-4 An example of typical cracks on the platform.

6-5 An example of platform buried under marine growth. The whole image is divided into 170 sub-images and each of the sub-image can represent an example of platform background.

6-6 The estimated projection matrix $A$. 

17
6-7 The crack signal used for simulation

6-8 (a) The scattering measurement of the mixture signal. This image shows the scattering simulated at the acoustic sensor plane; (b) Plot of the remaining information after the projection. After the projection, most of background interference has been removed; (c) The compressively reconstructed crack signal.

6-9 (a) The compressively reconstructed signal when there is no crack in the simulation; (b) The reconstruction result from scattering measurement of the mixture signal using traditional back-propagation method.

6-10 (a) Plot of true positive detection rate v.s. SNR. Each detection rate is obtained from 500 simulations with randomly generated cracks; (b) Plot of true negative detection rate v.s. SNR. Each detection rate is obtained from 500 simulations without cracks.
Chapter 1

Introduction

1.1 Super-localization problem

Object localization is useful in applications, such as “super-resolution” microscopy [1, 2, 3, 4], particle tracking [5, 6, 7], pattern recognition [8, 9, 10], spectral imaging [11, 12, 13], etc. In these imaging applications, the localization accuracy is the fundamental quantity of interest, since it determines the performance (e.g. effective resolution, sensitivity) of the system [14, 15, 16, 17, 18]. As a result, it is of practical interest to study how the localization accuracy is affected by various parameters. In particular, this thesis will study the influence of the finite sampling rate (e.g. pixel size of the photo detector), the number of samples (e.g. number of pixels on the sensor), and noise in the data acquisition process (e.g. thermal and shot noises). Here, methods that can break the physical limit of localization accuracy set by the imaging optics will be referred to as “super-localization” methods. In this thesis, two classes of super-localization problems will be investigated.
The first class of problem I will investigate here aims to improve the accuracy in localizing and tracking the physical motion of an object of interest. This problem is encountered in both microscopy, e.g. for tracking molecules within a cell [19, 20], and macroscopic imaging, e.g. for tracking small vibrations of an object [21, 22].

In particular, I will investigate three-dimensional (3D) localization problems in 3D imaging applications. In general, the parameters that need to be considered for such problems along the lateral \((x, y)\) dimensions are different from those along the axial \((z)\) dimension.

For lateral localization problems, the key parameters to consider are the pixel size as compared to the size of the point spread function (PSF) of the imaging system, and to the amount of motion of the object of interest. If the pixel size is much smaller than both the PSF size and the amount of motion, the underlying problem is sufficiently densely-sampled. As a result, standard methods, such as curve fitting [4, 23], cross correlation [24, 25, 26] or statistical estimation method [27], can be used to achieve localization accuracy with sub-pixel accuracy. On the other hand, when the pixel size is larger than either quantities, the underlying problem is under-sampled, and thus accurate localization becomes a more challenging task. In this thesis, I will consider the signal models for the under-sampled imaging problem, and demonstrate compressive reconstruction methods that eliminate the dense sampling requirement.

For axial localization problems, the dimension of interest is perpendicular to the plane where the samples (e.g. images) are taken. As a result, the localization accuracy will be affected not only by the pixel size and number of pixels, but also by the imaging geometry. In particular, the depth of field (DOF) defines the amount of axial
displacement that an object can undergo without introducing a significant amount of intensity signal change in a standard imaging system [28]. Intuitively, this is because an imaging system can only allow scattered signals from a certain range of angles (spatial frequencies), set by the numerical aperture (NA) of the imaging system, the missing angular information results in ambiguities in determining the object's axial position. Traditional axial localization methods are fundamentally limited by this missing data problem. In this thesis, I will investigate a class of compressive reconstruction method that exploits the “sparse” prior knowledge about the object in order to alleviate the missing angular information.

The second class of super-localization problem under investigation is to detect weak signals buried under strong background and noise. This problem exists in many practical situations. For example, a critical yet challenging problem in thin film transistor (TFT) panel production process [29] is to detect unwanted particles. The underlying problem can be abstracted by a weak signal detection model since the signal of interest (scattered light field of micron-sized particles) is often orders of magnitude weaker than the background and noise (e.g. scattered light from other structures on the panel). Another example is imaging through muddy water. Signals from the object are often buried by the multiply scattered light from the particles suspended in the medium. In this thesis, I will demonstrate a compressive reconstruction method that is able to detect signals captured with extremely low signal-to-noise ratio (SNR) and signal-to-background ratio (SBR), by exploiting different “sparsitics” in the respective signal and background subspaces.
1.2 Classical sampling: Nyquist–Shannon theorem

Sampling is the process of converting a continuous signal (for example, a function of continuous time or space) into a discrete sequence signal (a function of discrete time or space). Taking a digital image is an example of samplings the intensity of the wave field at sensor plane.

To maintain the ability to reconstruct the original continuous signal from the sampled signal, the sampling strategy used must meet certain criteria. For example, considering the continuous–time signal \( \sin(\pi t) \), if we only take samples at integer numbers of \( t \), the resulting sampled signal will be all zeros, losing all the information about the original signal. This type of problem is known as under-sampling.

The classical Nyquist–Shannon sampling theorem, which is a fundamental theorem that describes the minimum sampling frequency that is required to perfectly reconstruct a signal without any prior information [30]. It states that if a function \( x(t) \) contains no frequencies higher than \( B \) Hz, then \( x(t) \) can be reconstructed from the discrete samples if the sampling spacing is \( 1/2B \) seconds or less. Equivalently, a band–limited analog signal with band–width of \( B \) can be perfectly reconstructed if we sample the signal with \( 2B \) Hz or higher sampling rate.

The theorem assumes an idealization of any real–world situation, as it only applies to signals that are infinitely long in the time space. However, in the real world all the signals are time–limited and so cannot be perfectly band-limited. However, many signals of practical interest meet both specifications approximately, i.e. no significant amounts of energy are present either outside the frequency band \( B \) or the
Recently, it has been shown that, for an under-sampled linear system (i.e. using sampling rate lower than the one dictated by Nyquist–Shannon theorem), the continuous analog signal can still be perfectly reconstructed if the input signal is “sparse” or “compressible”. This new sampling framework is known as “compressed sensing” or “compressive sampling” [31, 32, 33].

1.3 New paradigm of sampling: compressed sensing

The general statement of compressed sensing is that a sparse signal can be accurately reconstructed from a small number of linear measurements (e.g. take samples from a signal) that is much smaller than the one required by Nyquist–Shannon sampling theorem [31, 32, 33, 34].

It is much more convenient to elaborate the ideas using the following linear algebra perspective. The acquisition of a signal $x$ can be modeled as $M$ linear measurements:

$$y_k = \langle \phi_k, x \rangle, \quad k = 1, ..., M. \quad (1.1)$$

Here we assume that $x$ is a $N$-dimensional vector and that $\phi$ is a $M \times N$ matrix with row vectors $\phi_k$. Solving for $x$ from the measurements $y$ is a linear inverse problem. If $M = N$ and $\phi$ is a full rank matrix, meaning that the measurements are all linearly independent, $x$ can be solved directly by $x = \phi^{-1} \cdot y$, where $^{-1}$ denotes matrix inverse.
If $M > N$ and $\phi$ has a rank no smaller than $N$, the problem is over-determined. A solution can also be found depending on the criteria used. For example, the least square solution leads to the well-known pseudo inverse solution $(\phi^\dagger \phi)^{-1} \phi^\dagger y$, where $\dagger$ denotes the complex conjugate transpose (a.k.a. adjoint) of a matrix. If $M < N$, the problem becomes under-determined and no unique solution can be found unless constraints are introduced. In compressed sensing framework, the input signal is assumed to be sparse in some pre-determined basis. In addition, if the measurements are taken at random (more details on requirements about measurements in later sections), an accurate reconstruction of $x$ is still possible even if the number of measurements is much smaller than $N$.

When a signal $x$ is sparse, it means that under some known basis $\psi$, $x$ can be represented by a $N$-dimensional vector $a$ with only $s$ non-zero coefficients, where $s$ is much smaller than $N$. In this way, $x$ and $y$ can be expressed as

$$x = \psi \cdot a, \quad (1.2)$$

$$y = \phi \cdot \psi \cdot a. \quad (1.3)$$

In order to recover the signal $x$, one can take advantage of the prior knowledge about sparsity and reconstruct the sparse coefficients $a$. Given the sparsifying basis $\psi$, it is straightforward to calculate $x$ from $a$. Theoretical analyses have shown that $a$ can be perfectly reconstructed from noise-free measurements [31, 33, 34] by solving the
\ell_1\text{ minimization problem:}

\[ \hat{a} = \arg \min_a \|a\|_{\ell_1} \text{, such that } y = \phi \cdot \psi \cdot a. \] (1.4)

Compressed sensing can be highly successful in solving under-determined problems. To guarantee accurate and stable reconstruction, the theory also provides a guideline on the smallest number of measurements. The constraint here depends on the sparsity of the unknown signal and an "incoherence" parameter \( \mu \), which measures the correlation between the sensing/measurement matrix \( \phi \) and the sparsifying basis \( \psi \). Details about these two constraints are described in the next two sections.

1.3.1 Sparsity

If the signal has a sparse representation under the basis \( \psi \), one can discard some of the small coefficients without significant loss in reconstruction fidelity. Most image compression techniques already exploit this idea for compression. For example, in JPEG 2000 format, the image is transformed into a wavelet basis where the representation is sparser [35]. Comparing to the JPEG format [36] which discards the high frequency content, JPEG 2000 can keep the smooth regions on the original image with good fidelity while the edges become much sharper. Furthermore, the sparser the original signal \( a \) is, the fewer measurements are required to obtain accurate reconstruction.
1.3.2 Incoherence

The "incoherence" term measures how the information of the input signal, which has a sparse representation in $\psi$, spreads out in the sensing matrix $\phi$ [31, 34]. In other words, incoherence measures how the input signals mix in the measurements. The input signal is local, while the measurements are global as each of the measurements should contain some information about each component in the input signal. Let us take the Dirac function as an example. In the space domain, the Dirac function is just a spike, which is a very sparse signal, while in the Fourier domain, the signal turns to a very flat function with "1"s everywhere. This means that the information of the Dirac function spreads out onto the entire domain of the Fourier space.

With a sampling matrix that has been properly normalized by $A \cdot A^\dagger = NI$ or $\phi \cdot \psi \cdot (\phi \cdot \psi)^\dagger = NI$ where $I = \text{diag}(1)$ is the unit matrix, the coherence parameter can be computed by [31]

$$\mu(\phi, \psi) = \max_{1 \leq k, j \leq N} |\langle \phi_k, \psi_j \rangle|.$$ \hfill (1.5)

If we define $A \triangleq \phi \cdot \psi$ and $y = A \cdot a$, the coherence parameter can be also expressed as

$$\mu(A) = \max_{1 \leq k, j \leq N} |A_{kj}|.$$ \hfill (1.6)

From (1.5), we can see that the coherence parameter measures the correlation between the columns in $\phi$ and $\psi$. If some of columns in $\phi$ and $\psi$ are correlated or dependent, the coherence parameter will be large. On the contrary, if the columns in $\phi$ and $\psi$ are
totally independent, as is the case with Fourier sampling, the coherence parameter will reach its minimum value. Typically, $\mu$ is in the range of $[1, \sqrt{N}]$.

1.3.3 Examples of incoherent measurements

- Discrete Fourier Transform (DFT)

For $N$-dimensional DFT matrix $W$, the transformation matrix $w$ is of the form

$$W = \begin{bmatrix}
1 & 1 & 1 & 1 & \ldots & 1 \\
1 & w & w^2 & w^3 & \ldots & w^{N-1} \\
1 & w^2 & w^4 & w^6 & \ldots & w^{2(N-1)} \\
1 & w^3 & w^6 & w^9 & \ldots & w^{3(N-1)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & w^{N-1} & w^{2(N-1)} & w^{3(N-1)} & \ldots & w^{(N-1)(N-1)}
\end{bmatrix},$$

where $w = e^{-i2\pi/N}$. It is clear that the discrete Fourier transform matrix $W$ satisfies the normalization rule $W \cdot W^\dagger = NI$. Then we can randomly sample the Fourier coefficients of the input signal. By calculating the coherence parameter for this sensing matrix $W$, it is easy to see that $\max_{1 \leq k, j \leq N} |W_{kj}|$ is $1$. So for the complex Fourier sampling matrix, its coherence parameter $\mu$ is $1$ the lowest limit the coherence parameter may attain.

- Identity Matrix

When the sensing matrix is just a magnified identity matrix $\sqrt{NI}$, we can predict that this sensing system has the largest coherence parameter. Consid-
tering a signal $x$ in the space domain, if we randomly sample $x$ directly, then
$y = \sqrt{NI} \cdot x$. In this case, the signal is formatted in spatial coordinate, so the
sparsifying matrix is just an identity matrix. As a result, $\phi$ is the same as $\psi$, which
means these two matrices are coherent to the maximum extent. In the
language of compressive sensing, this kind of sampling is the least efficient sam-
pling. By calculating the coherence parameter from (1.6), we can also obtaine
$\mu$ as $\sqrt{N}$, which matches with our prediction.

- Random convolution

Convolution is the general implementation of a linear shift–invariant system.
Suppose $y = h \ast x$, using Fourier–domain multiplication instead of direct con-
volution, we obtain in matrix form

$$y = W^\dagger \cdot \hat{h} \cdot W \cdot x,$$  \hspace{1cm} (1.7)

where $W$ and $W^\dagger$ denote the Fourier transform and inverse Fourier transform
respectively, and $\hat{h}$ is the Fourier transform of $h$. In (1.7), the $h$ matrix is
diagonal and its entries along the diagonal are the Fourier coefficients of the
convolution kernel $h$.

If the magnitude of all the entries along the diagonal of $\hat{h}$ are the same, $\hat{h}$ can
be normalized by a scalar factor so that the sensing matrix in the Equation 1.7
satisfies the normalization rule $A \cdot A^\dagger = NI$. In this case, $A$ is just $W^\dagger \cdot \hat{h} \cdot W$. If
the magnitudes are not the same, it does not quite satisfy the pre-requiment
of compressive sensing. This case is beyond the scope of this work.

1.3.4 Number of required measurements for compressive reconstruction

Suppose that the fixed sparse representation of the signal $x$ in the orthogonal basis $\psi$ is $s$-sparse. Select $M$ random measurements from the $\phi$ domain. It is shown that [31] if

$$M \geq C \cdot \mu^2 \cdot S \cdot \log(N),$$

(1.8)

where $C$ is a small positive constant, the signal $x$ can be reconstructed exactly with high probability. It has been shown that the probability of successful reconstruction will exceed $1 - \delta$ when $M \geq C \cdot \mu^2 \cdot S \cdot \log(N/\delta)$.

From (1.8), it is clear that the number of measurements should be proportional to the coherence parameter and the degree of sparsity. As mentioned above, the more incoherent the sensing matrix is with respect to the orthogonal sparsifying matrix, the more mixing of the input signal the measurements will incur, which equivalently means fewer measurements are required to guarantee exact recovery. Another important point is that the more sparse the signal is, the fewer measurements it requires to reconstruct the signal $x$. These are the reasons why coherence and sparsity are two important factors for judging whether the sensing system and input signal are good to use compressive reconstruction.
1.3.5 Why $\ell_1$ minimization?

For a $N$-dimensional vector $u$, its $\ell_p$ norm can be expressed as

$$\|u\|_p = \sum_{j=1}^{j=N} |u_j|^p.$$  \hspace{1cm} (1.9)

In most works on signal recovery or signal denoising, $\ell_2$ minimization, also known as minimum square error approach is widely used. The $\ell_2$ norm corresponds to the total energy of a signal. It has been popular for denoising, in the sense of minimizing the energy in the difference between the recovered signal and the actual input signal. But why in compressive sensing do we choose $\ell_1$ norm, rather than $\ell_2$ norm?

Before thinking about this question in a mathematical way, let us solve a simple problem first\(^1\). In a farm, the farmer raised some chickens and sheep. The farmer told us that the total number of the animals' legs is 16. Now the question becomes how many chickens and sheep respectively are in the farm. This is definitely an under-determined problem as we are supposed to solve two unknowns with only one equation:

$$2x + 4y = 16,$$  \hspace{1cm} (1.10)

where $x$ denotes the number of chickens and $y$ denotes that of sheep. If the farmer tells us another constraint on the solution that the $\ell_1$ norm of the solution should be minimized, the solution can be immediately obtained as $x = 0, y = 4$, which is also a so-called sparse solution (only one entry of the solution is non-zero). However, if

\(^1\)This example was constructed by Justin W. Lee.
the solution's $\ell_2$ norm of the solution should be minimized, the answer will change to $x = 2, y = 3$. We can see that the solution that minimizes the $\ell_2$ norm are not sparse at all, instead, it decreases the value of $y$ at the cost of sparsity.

From the perspective of maths, in a three-dimensional Cartesian coordinate, all the sparse signals fall mostly on the axes. Points with the same $\ell_1$ norm form a diamond-shape surface with the diamond corner located on the axes, while those with the same $\ell_2$ norm form an ellipsoidal surface. When the size of the diamond or ellipsoid expands, the $\ell_1$ norm or the $\ell_2$ norm respectively become large. The minimum error solution is found on the point of intersection of the error surface with the surface expressing the solution constraint (e.g. $2x + 4y = 16$ in the farmer's example above). If the error surface is ellipse ($\ell_2$ case), it can intersect the solution constraint; whereas if the error surface is diamond-shape ($\ell_1$ case) then it is more directly to intersect the solution constraint on one of the axes, thus yielding a "sparse" solution. Thus, by solving the $\ell_1$ norm minimization problem, we are inclined to get a sparse solution.

Here we only compare the two methods, $\ell_1$ norm minimization and $\ell_2$ norm minimization. However, $\ell_1$ norm minimization is not the only way to reconstruct sparse signals; some other methods have also been proposed.

1.4 Traditional point-to-point imaging

Figure 1-1 shows the schematic of traditional point-to-point imaging system. The magnification of the object is decided by the focal length of the convex lens as well.
as the distance between the object and the lens. However, no matter how the object is magnified, when the range of motion is smaller than the pixel size (after taking magnification not account), most of the information on the image will remain the same except for that around the edges of the object. As a result, if we desire to detect the sub–pixel movement of the object by capturing images of the object before and after the displacement through this traditional method, most of the content on the images are not useful for detecting the tiny movement. In other words, it is not efficient to extract sub–pixel movement by using traditional imaging techniques.

We can also look at this imaging system from the perspective of compressive sensing. Since it is point–to–point imaging, the sensing matrix is basically an identity matrix, which is a very bad sensing matrix in CS as described above.

1.5 Introduction to digital holography

Unlike the traditional imaging, digital holography (DH), a famous technique for reconstructing a 3D profile of the object with only one single shot [37, 38, 39, 40, 41, 42],
records the diffraction pattern of an object. During the recording, the information of the object spreads out onto the whole camera sensor through diffraction. In this sensor, DH imaging system has a much larger coherence parameter compared to the traditional point-to-point imaging. This leads to a better sensing system and a higher chance to solve the super-localization problem.

Digital holography and digital holographic image processing have become more applicable due to advances in mega-pixel electronic sensors, e.g. CCD and CMOS, with high spatial resolution and high dynamic range. DH has been widely used in two-phase flow imaging, phase contrast imaging and particle image velocimetry (PIV) [41, 43].

There are mainly two different kinds of DH setup, in–line and off-axis. In our work, we only consider in–line lensless DH [41, 44, 45]. Figure 1-2 shows the schematic of a typical lensless in–line DH setup. Here, by “lensless”, we mean that there is no lens between the object and the digital detector.

When the plane wave illumination propagates onto the object, some of the light will be scattered by the object. In this simple in–line DH system, we assume that the
scattering is not severe; in other words, most of the illumination remains unscattered after it propagates through the object. The light field propagates according to free-space (Fresnel) propagation between the object and the detector. At the detector plane, the unscattered part will serve as a reference wave, and interfere with the propagating object field.

The intensity of the interference pattern $I(m\Delta, n\Delta, z_D)$ recorded on the 2D detector array can be written as

$$I(m\Delta, n\Delta, z_D) = |A_r + a(m\Delta, n\Delta, z_D)|^2$$

$$= A_r^2 + |a(m\Delta, n\Delta, z_D)|^2 + A_r a^*(m\Delta, n\Delta, z_D) + A_r^* a(m\Delta, n\Delta, z_D),$$

(1.11)

where $\Delta$ is the pixel size on the digital detector, $A_r$ is the amplitude of the illumination plane wave, $z_D$ is the distance between the object plane and detector plane, and $a(m\Delta, n\Delta, z_D)$ is the field of the scattered light at the recording plane. In Equation 1.11, the term $A_r^2$ is simply a constant and can be removed by eliminating the DC term when conducting Fourier transforming the interference pattern $I(m\Delta, n\Delta, z_D)$. Moreover, without loss of generality we may assume $A_r$ as 1. The second term may be dropped as negligible when assuming $|a(m\Delta, n\Delta, z_D)| \ll A_r$. In-line holography has an inherent limitation caused by the generation of overlapping twin images. Here the so-called twin image problem can be eliminated by using compressive reconstruction method as proposed in [46], which will be introduced in Chapter 2. According to the Huygens-Fresnel principle and the Fresnel approximation, the scattered field at
\( z = z_c \) can be expressed in the form

\[
a(m\Delta, n\Delta, z_D) = g(m\Delta, n\Delta) * h(m\Delta, n\Delta),
\]

(1.12)

where \(*\) denotes the convolution operator and \( h(m\Delta, n\Delta) \) is the free-space propagation point-spread-function \( \exp \left( \frac{ik}{\sqrt{m^2\Delta^2 + n^2\Delta^2 + z_D^2}} \right) \). Hence, by neglecting the halo and twin-image terms in (1.11), the information encoded on the hologram has a linear relationship with the field at the object plane as

\[
I(m\Delta, n\Delta, z_D) = g(m\Delta, n\Delta) * h(m\Delta, n\Delta) + e,
\]

(1.13)

where \( e \) includes the halo, twin image as well as other sources of noise that are uncorrelated to the object. Since there is a convolution operator in Equation 1.13, each measurement captured on the hologram contains the information of the whole object. In other words, all the measurements in this single shot during the holography experiment can contribute to detect sub-pixel movement, which means using digital holography can greatly increase the chance to detect the sub-pixel motion, as opposed to the traditional point-to-point imaging technique where the edge information is localized.

### 1.6 Outline of the thesis

In Chapter 2, the derivative operator is applied to sparsify the object function and a linear model is built for compressive holography. Experimental setup and results for
1D sub-pixel movement detection using compressive holography are presented. 1/45 pixel size movement can be successfully detected.

In Chapter 3, spiral phase mask is introduced to sparsify a general 2D object. Experimental setup and results for 2D sub-pixel movement detection are also shown. The 1/30 sub-pixel 2D motion has been successfully measured.

In Chapter 4, the theoretical limit for axial localization is subject to the system's depth of field. By applying compressive holography and template matching, axial localization with 1/16-depth of field accuracy is demonstrated.

In Chapter 5, the developed motion detection algorithm has been implemented onto vibration characterization. The vortex-induced vibration and wake-induced vibration of harbor sea seal whisker models are successfully detected using the compressive holography technique. To prove the validity of the characterized vibration, the detected vibrations are also compared with the experimental results obtained from large-scale models and data from different instrumentations.

In Chapter 6, compressive signal detection from low signal-to-noise ratio and low background-to-signal ratio measurements is explored. To remove the heavy background signal in the measurements, null space signal separation is introduced here. The success of the proposed technique is demonstrated for underwater crack detection using an acoustic imaging model.

In Chapter 7, conclusion of this thesis is presented and the main results are summarized.
Chapter 2

One dimensional sub–pixel movement detection using compressive holography

Optical real–time localization of objects enables progress in diverse scientific domains. Examples include: observation of Brownian motion [47], tracking of a single molecule [48], etc. Intuitively, the accuracy (smallest detectable displacement) should be limited by the finite pixel size of the digital camera. Methods to overcome this limitation typically involve scanning or using priors.

Scanning leads to dramatic accuracy improvements [49] at the cost of slower frame acquisition. Priors are utilized in techniques such as curve fitting [48], feature–based tracking methods such as digital image correlation [50], and gradient–based image registration [51, 52]. Compressed sensing, with implementing sparsity priors, has
been shown to be more robust to noise.

In this chapter, I will start my investigation of single-shot small (e.g. sub-pixel scale) motion detection without mechanical scanning from a compressive sampling perspective. Insights can be gained by considering the most simplified model involving only one dimensional (1D) motion, which will be focus of this chapter.

As shown in Chapter 1, in an in-line digital holography setup, the hologram is captured on a digital camera with finite pixel size, which limits the rate at which the intensity signal is sampled. In the real word, all objects are finite, which means that the frequency response of a object is infinite. In this way, it is impossible to apply Nyquist theorem to sample the signal and then perfectly reconstruct the input signal. Nyquist theorem, however, does not take into account any prior information about the object. It has been shown that using compressive sensing, highly accurate solutions of an under-sampled linear system can be obtained [33, 53]. In addition, the solution is robust to noise [54].

2.1 Incoherence and sparsity in digital holography

As mentioned in Chapter 1, successful implementation of compressive reconstruction is conditioned upon two requirements: sparsity and incoherence [34]. To enforce sparsity of the object signal, we consider a one-dimensional (1D) signal, as illustrated in Figure 2-1(a).
Let us define a coordinate vector

\[ \mathbf{x} = [x_0, x_1, \ldots, x_j, \ldots, x_{N-1}] \],

where \( x_n = nd; \quad n = 0, 1, \ldots, N - 1 \); and \( d \) denotes the desired motion accuracy. The pixel size is \( \Delta = Gd \), where \( G \) is our intended sub-pixel accuracy gain of motion detection. We will consider only a single opaque object surrounded by uniform intensity within the field of view; the intensity at the object plane may therefore be represented as a vector \( \mathbf{i}^{\text{obj}} \) of length \( N \) corresponding to the coordinate vector \( \mathbf{x} \), where

\[ \mathbf{i}^{\text{obj}} = [0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0] . \]\n
Fig. 2-1(a) shows \( \mathbf{i}^{\text{obj}} \) from one such possible object. The vector \( \mathbf{i}^{\text{obj}} \) is unknown in the sense that we do not know where the transitions between value 1 and value 0 occur. This vector is not sparse, but we can easily sparsify it by taking the spatial
derivative along $x$, which will produce two impulses at the edges of the object. Then, the new sparse derivative vector is in the form of

$$\Delta I^{\text{obj}} = [0, \ldots, 0, 1, 0, \ldots, 0, -1, 0 \ldots, 0], \quad (2.3)$$

and what is unknown is the locations of the impulses.

The second requirement about “incoherence” does not use the term according to the typical sense we assign in Statistical Optics; rather, it means that the information of the unknown vector $\Delta I^{\text{obj}}$ must be evenly spread over the set of basis vectors that describe it [34]. We utilize diffraction for that purpose, which is what motivated our use of Fresnel holography [46, 55, 56, 57, 58, 59, 60, 61, 62]. The spreading produced by the Fresnel propagator is not provably optimal, but it is extremely easy to attain by simple free-space propagation in the lab. To implement the optimal operator, one would require special-purpose phase masks placed at certain locations along the path [63, 64]; that is beyond the scope of the present work.

2.2 Theoretical model for compressive holography

Figure 2-2 is a schematic of a typical in-line digital holography setup. A linear model has been obtained in Equation 1.13 as

$$I(m\Delta, n\Delta, z_D) = g(m\Delta, n\Delta) \ast h(m\Delta, n\Delta) + \epsilon, \quad (2.4)$$
On-axis plane wave illumination

Stepping motion direction

z = 0

z = z_D

Figure 2-2: In–line DH geometry. The object was moved along y direction with a uniform step size of 267nm (1/45-pixel) by a piezo-driven motion stage.

where Δ is the pixel pitch on the digital detector and e includes the halo, the twin image as well as other sources of noise that are uncorrelated to the object.

We multiply the Fourier transform of the intensity by iu, where u is the spatial frequency variable, to obtain the derivative in the spatial domain. To upsample so that we can measure sub-pixel movement, the spectrum of the hologram is then zero-padded at both sides. Finally, a linear model relating the derivative of the object g' and the hologram I expressed in the Fourier domain can be obtained in the form of

\[ iu \cdot \mathcal{F} \cdot I = Q \cdot (H \cdot \mathcal{F} \cdot g' + iu \cdot e'), \tag{2.5} \]

where \( \mathcal{F} \) denotes the discrete Fourier transform matrix, \( H \) is a diagonal matrix with the Fourier transform of \( h \) at the diagonals, and \( e' \) is the Fourier-transformed noise. The mask \( Q \) in Equation 2.5 is a diagonal matrix with \([0, \ldots, 0, 1, 1, \ldots, 1, 0, \ldots, 0]\) along its diagonal and the total number of 1 in \( Q \)'s diagonal is \( M \), which is also the
number of pixels in the recorded hologram. Since the edges of the object are sparse, (2.5) can be inverted by enforcing the sparsity constraint using $\ell_1$-minimization. The edges (derivative) $\tilde{g}'$ of the object can be estimated by solving

$$\tilde{g}' = \arg\min_{g'} \|g'\|_1, \text{such that}$$

$$iu \cdot \mathcal{F} \cdot I = Q \cdot H \cdot \mathcal{F} \cdot g',$$

(2.6)

where $\| \cdot \|_1$ denotes the $\ell_1$-norm of a vector. We implemented $\ell_1$-minimization by adapting the Two-step Iterative Shrinkage/Thresholding (TwIST) algorithm [65].

2.3 1D sub-pixel motion detection experimental setup

Our lensless experimental setup for the 1D sub-pixel motion detection is shown in Figure 2-3. A collimated He-Ne laser of wavelength 632.8nm was used to illuminate the object. The input laser beam was expanded by a spatial filter, and collimated by a plano-convex lens. After propagating through the object, the resultant hologram was recorded by a Basler A504k camera with 1024×1024 pixels and 12\(\mu\)m pixel pitch. The object, a pin with an average diameter of 900\(\mu\)m, was placed 151.6mm away from the detector. A piezo-driven motion stage (Model number: Thorlabs Nanomax 312) with 20\(\mu\)m resolution was used to move the pin laterally along the $y$ direction step by step.

In the experiment, the pin object was modeled as a 1D rectangular function. This
Figure 2-3: Experimental in-line DH setup. The pin object was moved along $y$ direction with a uniform step size of $267\text{nm}(1/45\text{-pixel})$ by a piezo-driven motion stage.

is well justified since the width of most part of the pin is uniform. Using our simulation, we found that $C \approx 30$ in Equation 1.8 guarantees the reconstructed object position to be correct more than 50% of the time under our experiment conditions. Thus, the maximum sampling gain is found to be 60, which implies that the theoretically smallest detectable movement in our experimental arrangement is $1/60\text{-pixel}$ size ($200\text{nm}$).

We verified our theory by moving the pin object along the $y$-axis with a step size of $267\text{nm}$ (equivalently $1/45\text{-pixel}$ size or gain of 45). The pin was moved by 45 steps (1 pixel) in total. A hologram was captured after each step of movement; a sample hologram is shown in Figure 2-4. A row vector of length 1024 can be extracted from the 2D hologram in Figure 2-4 to form a 1D hologram, as shown in Figure 2-5.
2.4 Experimental results for 1D sub–pixel motion detection

The compressive reconstruction result of the 1D hologram in Figure 2-5 is shown in Figure 2-6. The edges of the original pin object were successfully reconstructed, free from artefacts due to twin image and other sources of noise.

To quantify the accuracy of our approach, we randomly chose seven rows (away from the pin’s tapered portion) from the first hologram (which defines the origin of pin’s movement), and then tracked the left edges of those seven rows for the following 45 steps. The compressive reconstruction was repeated for the same seven rows on the consecutive 45 holograms. Figure 2-7 shows the measured left edge positions of those seven rows from the total 46 holograms, compared with the “true” positions specified by the piezo stage. From the histogram shown on the bottom right in Figure 2-7, the
Figure 2-5: A sample row vector from the 2D hologram in Figure 2-4.

average step size and precision (standard deviation) were calculated as 269nm and 12nm, respectively. The fraction of correct position reconstructions, i.e. those falling precisely on the 267nm mark, is 58%. It is also encouraging that less than 7% of the data are off by more than 1 step compared to the "true" position.
Figure 2-6: Real part of the compressive reconstruction of the 1D hologram in Figure 2-5.

Figure 2-7: Compressive reconstructed positions of seven randomly chosen rows (in blue circles) and the “true” position (in red stair curve) at each step. The histogram on the bottom right combines the data of total 45 steps taken by the seven rows (45 × 7 = 315 data points total.)
Figure 2-8: The comparison between the average position (purple points) of each step and the "true" positions.
Chapter 3

2D sub–pixel movement detection using spiral phase filtering and compressive holography

With the success of implementing compressive holography into detecting 1D sub–pixel movements, developing an algorithm to detect the sub–pixel movement in 2D is naturally the next topic. However, it is not trivial to extend the 1D algorithm to 2D case, because the method we used to estimate the 1D derivative does not generalize to higher dimensions. Alternatively, we can think of the derivative as a Hilbert transform; which does not exist in 2D either [66].

Consider for example a square object (shown in Figure 3-1(a)). Its 1D derivative along the horizontal direction is a two-stripe function and the 2D derivative will be four points, corresponding to the four corners of the square. Apparently the four
points extracted from the square by taking its 2D derivative are not the edges of the square, thus we are deprived of much valuable information about the object, completely missing its edges and being left with information about its corners only. Clearly, that ought not to be enough to localize sub-pixel motion for this square object.

Figure 3-1: (a) A square object; (b) the derivative of the square along the horizontal direction; (c) the 2D derivative of the square.
3.1 Spiral phase mask in edge extraction

Spiral phase mask has been extensively used as a linear edge extraction operator in phase contrast microscopy [67, 68, 69, 70]. The mask is a Fourier-domain vortex shape filter in the form of

\[ S(\phi) = \exp(i\phi), \quad (3.1) \]

where \( \phi \) stands for the frequency polar coordinate. The spiral-like phase distribution on the spiral phase mask is shown in Fig. 3-2. Unlike the dark field microscopy [71], where the zeroth order of the illumination beam is blocked such that most of the intensity is lost, the spiral phase mask acts by redistributing the illumination field while preserving the edges of the structures in the specimen.

![Phase distribution of spiral phase mask](image)

Figure 3-2: Phase distribution of spiral phase mask

Spiral phase mask can also be treated as an approximation of a two-dimensional Hilbert transform [66]. The edge-extraction mechanism by the spiral phase filter is explained as follows. The Fourier transform of spiral phase mask can be expressed as \((1/r^2) \exp(i\varphi)\) [68], where \( r \) and \( \varphi \) are the polar coordinates in the object plane. When implementing such a spiral phase mask in the Fourier plane, the object field
is convolved with the point spread function (PSF) \((1/r^2) \exp(i\varphi)\). During the convolution, every point on the object field is multiplied with the PSF. If all the points have the same value, the summation over all the points will cancel out due to the \(\varphi\) dependence of the PSF. On the other hand, if the points are at the edge area where field distribution is varying, the convolution result will be a non-zero value. Conveniently, since the PSF is inversely proportional to the distance \(r\), the edge-extracting action of the spiral mask remains confined locally. We have shown preliminary experimental results of 2D subpixel movement detection using spiral phase filtering and compressive in-line holography [72]; since then, it has also been proven that spiral phase mask can provide edge contrast enhancement effect in incoherent imaging [73].

Traditionally in the spiral phase microscope, the spiral phase filtering is displayed at a high-resolution spatial light modulator where a blazed phase hologram is shown [67]. In some other cases, a static phase hologram or physical spiral phase plate was used in the imaging system [74, 75]. In our approach, we computationally implemented the spiral phase filtering in the spectral domain, followed by spectral domain zero-padding, and then we applied a sparsity constraint to the object edge signal. Finally, a simple cross-correlation operation [26, 76] between the compressively reconstructed images from consecutive frames revealed the subpixel displacement.
3.2 Implementing the spiral phase mask to 2D compressive holography model

We consider an in–line digital holography (DH) geometry, as illustrated in Fig. 3-3. The hologram recorded at the camera plane is the result of interference between the scattered optical field \(a(x, y; z_c)\) from the object and a reference wave \(A\), representing the unscattered part of the illumination. If the object is spatially sparse, the reference wave is approximately uniform. We discretize the signal according to \(a(m_xd, m_yd; z_c)\), where \(d\) denotes the desired localization accuracy, \(m_x, m_y\) index possible object positions within one pixel, and \(z_c\) is the distance between object and camera. Here, the number of pixels on the camera is \(N_x \times N_y\), the pitch is \(\Delta\) and each pixel on the camera is divided into \(G = (\Delta/d)^2\) subpixels, as illustrated in Fig. 3-4. \(G\) can be thought of as the gain of the subpixel localization algorithm. In the example shown in Fig. 3-4, \(G = 4^2\).

![Diagram of DH geometry](image)

Figure 3-3: In–line DH geometry. The object was moved along \(x\) and \(y\) directions by a piezo-driven motion stage.

Assuming a rectangular pixel function, the intensity measured by the hologram at the pixel \((n_x, n_y)\) is the integral of all the intensity within this pixel. In the discretized
Figure 3-4: Subpixel notation. In this example, each camera pixel is divided into $4 \times 4$ pixels. $(n_x, n_y)$ denotes the camera pixel index and $(m_x, m_y)$ represents the pixel index after subpixel division.

Figure 3-5: (a) Real part of effective transfer function for edge-extraction holography; (b) Real part of transfer function for traditional in-line DH.

In the form, the intensity $I(n_x \Delta, n_y \Delta; z_c)$ can be written as a sum of all the intensities at each subpixel as

$$I(n_x \Delta, n_y \Delta; z_c) = \sum_{m_x=(n_x-1)G+1}^{n_x G} \sum_{m_y=(n_y-1)G+1}^{n_y G} |A_r + a(m_x d, m_y d; z_c)|^2$$

$$= (\downarrow G) [A_r + a(m_x d, m_y d; z_c)]^2 \ast q_G]$$

$$= (\downarrow G) \{[A_r^2 + |a(m_x d, m_y d; z_c)|^2 + A_r \cdot a^*(m_x d, m_y d; z_c)$$

$$+ A_r^* \cdot a(m_x d, m_y d; z_c)] \ast q_G\}.$$

54
where \((\downarrow G)\) denotes a down sampling operator that computes the output by taking every \(G\)th entry of the input, \(*\) denotes convolution, and \(q_G\) is a \(G \times G\) matrix with all entries equal to 1. For computational convenience later, the \(q_G\) matrix is actually padded symmetrically by zeros up to size \(GN_x \times GN_y\), without altering the result in Eq. (3.2).

We assume that the object can be modeled as a 2D complex transmission function \(a(x, y; 0)\). In order to localize the object with accuracy \(d\), the object's transmission function is discretized with pixel pitch \(d\) and is denoted by a matrix \(a(m_xd, m_yd; 0)\). The scattered field recorded at the hologram place at \(z = z_c\) is related to \(a\) by a convolution with the point spread function \(h\) of free space (Fresnel kernel), as follows

\[
a(m_xd, m_yd; z_c) = a(m_xd, m_yd; 0) * h(m_xd, m_yd; z_c),
\]

where \(h(m_xd, m_yd; z_c) = \exp(jk\sqrt{(m_xd)^2 + (m_yd)^2 + z_c^2})\) and \(k\) is the wave number \([28]\). With the above assumptions, the intensity information encoded on the hologram is related to the optical field at the object plane linearly as

\[
I(n_xX, n_yY; z_c) = (\downarrow G)[a(m_xd, m_yd; 0) * h(m_xd, m_yd; z_c) * q_G] + e,
\]

where \(e\) includes the halo and twin image terms in Eq. (3.2), as well as all sources of noise.

The next step is to sparsify the object representation, i.e. perform edge extraction. As mentioned earlier, we achieve this by digitally applying a spiral phase mask pupil
to the Fourier transform of the Fresnel-propagated object (i.e., multiplying the object with the transfer function shown in Fig. 3-5(a) in the Fourier domain). The forward transfer function of a traditional DH imaging system is shown in Fig. 3-5(b) for comparison.

After sparsifying, we up-sample the measured hologram of size $N_x \times N_y$ and pitch $\Delta$ by a up-sampling operator $(\uparrow G)$ such that the up-sampled hologram is represented by a matrix of size with $GN_x \times GN_y$ with pitch $d = \Delta/G$. In the Fourier domain, the up-sampling operator is implemented by padding the spectrum of the original hologram symmetrically by zeros up to size of $GN_x \times GN_y$.

At this point, we are ready to compressively invert Eq. 3.4. Complying with standard compressed sensing notation, we express the image formation process as

$$(\uparrow G) \cdot S \cdot \mathcal{F} \cdot I = Q_G \cdot H \cdot \mathcal{F} \cdot a_{\text{edge}} + S \cdot E,$$  

(3.5)

where $a_{\text{edge}}$ denotes the object's edge signal that we desire to reconstruct; $S$ is the spirally phase mask in (3.1); $\mathcal{F}$ denotes the 2D Fourier transform operator; $Q_G$ is the Fourier transform of the $q_G$ matrix from (3.2); $H$ is the Fourier transform of the 2D PSF kernel $h$ from (3.3); and $E$ is the Fourier transform of the combined halo, twin image and noise term in (3.4).

The object edge signal $a_{\text{edge}}$ is obtained by minimizing $\ell_1$-norm $||a_{\text{edge}}||_1$ with (3.5) as a constraint. We implemented the $\ell_1$ minimization by adapting the Two-step Iterative Shrinkage/Thresholding (TwIST) algorithm [65].

The compressive reconstruction method described so far is repeated for each frame.
Figure 3-6: Summary of the computational reconstruction procedure, together with representative intermediate reconstruction results from an actual experiment with a star object.

during the course of object motion. To discover the relative motion of the object between successive frames \( t \) and \( t + 1 \), we cross-correlate \( a_{\text{edge}}(t) \) with \( a_{\text{edge}}(t + 1) \). The displacement of the correlation peak from the origin of the coordinate system equals the relative displacement of the object between the two frames. Since \( a_{\text{edge}} \) and the cross-correlations are sampled at pitch \( d \), the same pitch \( d \) also determines the accuracy of the displacement measurement. The entire computational reconstruction procedure is summarized in Fig. 4-7.

3.3 Experimental results

In the experimental in-line DH setup, we illuminated the object with a Helium-Neon laser of wavelength 633nm. Holograms were taken by a Basler A504k camera with 1024×1024 pixels and 12\( \mu \)m pixel pitch. To test the compressive edge reconstruction from our algorithm, we experimented with three different objects, shown in Figs. 3-7(a), 3-7(c) and 3-7(i). The first one is a star shape binary amplitude object with
lateral size of \(350 \mu m \times 350 \mu m\), fabricated on a transparent soda lime glass with 80nm-thick Chromium defined by the photolithography. The second one is a stainless steel shim, and it has an inner diameter of 2.24mm, an outer diameter of 3.88mm and an average thickness of 0.254mm. The shim was attached to an optical post by a pin of diameter 250\(\mu m\). The third one is an “M” phase object, which was etched onto a glass microscope slide \((n \approx 1.5)\) to a depth of approximately 225nm. Holograms of the three objects and their respective compressive edge reconstructions are shown in Figs. 3-7(b), 3-7(f), 3-7(j) and 3-7(c), 3-7(g), 3-7(k) respectively. Compared to the reconstruction results by traditional back-propagation methods (shown in Figs. 3-7(d), 3-7(h) and 3-7(l), respectively), not only can the compressive reconstruction algorithm successfully reconstruct the edges of original object, it can also eliminate the effects from background noise and twin image problems. The latter observation also holds without edge extraction, as was pointed out in [46].

Next we investigated the subpixel localization ability of our proposed algorithm. The star object was placed on a piezo–driven motion stage with 20nm accuracy (Thorlabs NanoMax 312). The distance between the object and the camera was calibrated by simple back-propagation reconstruction, and was computed to be 89.7mm. During the experiment, the star was moved by the piezo motion stage step by step along a 2D pattern resembling the letter “Z” pattern, with step size of 400nm. The total lateral motion was confined to be exactly equal to one camera pixel. At each step of motion, a hologram frame was captured and cropped to 256 \(\times\) 256 pixels such that the up-sampled hologram for gain \(G = 30 \times 30\) was of size of 7680 \(\times\) 7680. This was the maximum matrix size that our laptop computer (using a MacBook Pro with 2.66
Figure 3-7: (a) Microscopic image of a star shape binary amplitude object; (b) Hologram of the star object; (c) Compressive reconstruction of star hologram; (d) Traditional back-propagation of star hologram; (e) Image of a ring shape amplitude object; (f) Hologram of the ring; (g) Compressive reconstruction of the ring hologram; (h) Traditional back-propagation of the ring hologram; (i) “M” shape phase object; (j) Hologram of the “M”; (k) Compressive reconstruction of “M” hologram; (l) Traditional back-propagation of “M” hologram.
GHz Intel Core 2 Duo and 8 GB Memory) could accommodate. The algorithm in Fig. 4-7 was repeated for each step of motion, and the resulting estimates of relative displacement between captured frames were tabulated. The computation time for each hologram processing using the laptop computer was approximately 10 minutes. The computation time can be improved if processed by a server computer. On average, it takes 90 seconds to process one hologram on a computer using a Dell T7600 desktop computer with 2x E5-2660 (ES) 16 Core and 128GB Memory.

The result of tracking a star object by our compressive construction is shown in Fig. 3-8(a) with red dots indicating the detected locations at each step. It clearly indicates successful reconstruction of the ground-truth “Z” trajectory shape (marked by the blue “Z” pattern). A histogram of displacement estimates is shown in Fig. 3-8(b). 80% steps were located correctly and the average measured star trajectory step size is 403nm, also matching well with the ground truth. The standard deviation of measured step size is 183nm.

To compare, we also implemented the standard digital image correlation (DIC) method on the traditionally back-propagation (direct deconvolution) reconstruction results from the same data. Since the back-propagation method cannot remove the static background on the hologram, we used thresholding method to preprocess the reconstructed results and the thresholding method can remove more than 90% of the static background. The DIC analysis was performed on each frame of the background removed reconstructions using a MATLAB-based DIC solver [?]. The extracted motion pattern by DIC is shown in Fig. 3-8(c) with the red dots indicating the corresponding localization results, and the histogram of the estimated step size by DIC is
shown in Fig. 3-8(d). The corrected detection rate is 33.3%. With using thresholding and DIC, the average measured step size is 404nm, while the standard deviation is 545nm, which is three times as large as the results achieved from our compressive approach.

3.4 Discussion

The results of Fig. 3-8(a) show that errors tend to occur with higher frequency near the beginning of the trajectory, rather than later parts. We attribute this to imperfections in the mechanical design of the stage, since the reconstruction algorithm itself should be shift invariant (the object is always quite far from the edges of the camera, where the shift invariance assumption would break down.)

The cross–correlation algorithm is known to perform quite well for subpixel motion with gains in the order of $G \sim 10 \times 10$; its inferior performance displayed in Fig. 3-8(b) is because for even smaller motions edge effects become more pronounced and, as they aggregate over the entire object, they influence the accuracy. This is true for bright field imaging (where cross correlation is more typically applied) as well as digital holograms; however, the compressive algorithm used here is more effective at utilizing the sparsity prior to mitigate these artifacts. Thus, with our algorithm we can achieve better gain by a factor of approximately 3 in each dimension reliably, at the cost of more complex hardware with active illumination.

Our optimism about the compressive approach is further justified because the experiment reported here was limited by the piezo motion stage and computer we
had available. The motion stage was able to move the object only by a step size equal to an integral multiple of 20nm. Moreover, the up-sampled matrix size had to be limited in order to fit in the laptop’s 8GB memory, as mentioned earlier. Both of these constraints can be removed in future experiments, potentially beating the gain factor that we demonstrated here.

Finally, it is worth noting that, upon superficial examination, it might appear that the procedure described here somehow increases the amount of information extracted from the intensities measured by the camera, and that this is magically accomplished by the algorithm’s up-sampling step. Clearly, this cannot be so. What really happens is that the object’s position, with subpixel accuracy, is already encoded within the information captured by the camera’s $N_x \times N_y$ pixels, and that is despite undersampling of the hologram’s fringes, halo and twin contamination and noise. Therefore, a small change in object position is certainly not visible by naked eye, and back-propagation or other traditional DH processing techniques may still not be able to reveal it. The compressive algorithm succeeds only if the sparsity assumption holds. This was true in our experiments moving one object with few edges across one pixel, but it would probably break down if multiple or highly textured objects were presented to the system.
Figure 3-8: (a) Star trajectory characterization from compressive localization algorithm; (b) The histogram of estimated step size distribution using compressive reconstruction; (c) Star trajectory characterization using standard DIC approach; (d) The histogram of the estimated step size using standard DIC approach. In both (a) and (c), red dots indicate the localization results from the corresponding hologram using either method, and the blue "Z" pattern indicates the ground truth of the motion established using the piezo stage.
Chapter 4

Axial super–localization with sub–depth–of–field accuracy via compressive pattern matching

4.1 Classical limit of axial localization accuracy

In a lens-based imaging system, if the distance between the object plane and the lens, \( z_1 \), and the distance between imaging plane and lens, \( z_2 \), satisfy the imaging condition

\[
\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f'}, \tag{4.1}
\]

where \( f \) is the focus length of the lens, geometrical optics predicts that one can observe a shape image of the object at the imaging plane.

For example, in Fig. 4-1, an arrow–shaped object is placed at \( 2f \) away from the
Figure 4-1: An example of single lens imaging system with object placed at $2f$

...
bring an ambiguity in localizing the object.

Figure 4-2: 4F imaging system. An object $g(x)$ is placed $\Delta z$ from the object plane and the imaging plane captures a defocused image of the object.

Next, I will calculate this theoretical defocus tolerance by considering the 4F imaging system shown in Fig. 4-2. Without loss of generality, 1D case will be used here. Assume the object is described by $g(x)$, where $x$ represents the spatial coordinate on the object plane. If the object is defocused by $\Delta z$, the optical field on the object plane is $g(x) \exp\left(i\frac{\pi x^2}{\lambda \Delta z}\right)$. Accordingly, the optical field at the imaging plane will be $g(x') \exp\left(i\frac{\pi x'^2}{\lambda \Delta z}\right)$, where $x'$ denotes the spatial coordinate on the imaging plane. To examine how the defocus distance $\Delta z$ will affect the recorded signal, here we will consider the problem in the spatial frequency domain. If we take the Fourier transform of the defocused signal, $g(x) \exp\left(i\frac{\pi x^2}{\lambda \Delta z}\right)$, its Fourier spectrum is $G(u) \exp\left[-i\pi \frac{\lambda \Delta z u^2}{\lambda f^2}\right]$, where $u$ stands for the spatial frequency coordinate. To
better interpret the Fourier spectrum, we can replace $u$ with spatial coordinate at the pupil plane, $x''/\lambda f$, and rewrite the spectrum as $G(x''/\lambda f) \cdot \exp[-i\pi \Delta z x''^2 / \lambda f^2]$. Comparing this with the in-focus spectrum $G(u)$, we can see that because of defocus, the spectrum is modulated by a quadratic phase term $\exp[-i\pi \Delta z x''^2 / \lambda f^2]$. The defocus-insensitivity can be studied by looking at the real part of this term.

Figure 4-3 shows the real part of defocus modulation functions $\exp[-i\pi \Delta z x''^2 / \lambda f^2]$ at two defocus distances $\Delta z$. If the modulation function is an all-"1" function, it will not change the object's spectrum, which can only be achieved when $\Delta z$ is 0. Figure 4-3(a) shows the modulation function with a small $\Delta z$ (small defocus) and Fig. 4-3(b) a large $\Delta z$ (large defocus). From these plots, one can see that as long as the spatial frequencies passing through the optical system is confined within the flat regions of the modulation functions, the object's spectrum will not be changed much, thus leaving the system defocus-insensitive. The larger $\Delta z$ is, the narrower the flat region will be. This requirement poses a constraint on largest defocus distance allowed, according to

$$\frac{x''}{\lambda f} \leq \frac{1}{\sqrt{2\lambda \Delta z}}. \quad (4.2)$$

It implies that as long as the defocus distance is smaller than $\lambda/2 \left(\frac{Z_{\text{max}}}{f}\right)^2$, the image signal received at the imaging plane is indistinguishable. This defocus distance is known as the Depth of Field (DOF) of the optical system; $Z_{\text{max}}/f$ defines the numerical aperture (NA) of the imaging system, which characterizes the maximum angle allowed.

When it comes to reconstructing the object signal, naturally the axial loca-
Figure 4-3: (a) Frequency modulation for a mild defocus system; (b) Frequency modulation for a strong defocus system.

of the reconstructed object will have a $2 \times \text{DOF}$ uncertainty due to the double–side defocus–insensitivity. This can be a huge problem in applications requiring precious knowledge about the object’s axial position. Next, we will demonstrate that by using compressive holography and pattern matching techniques, we could break the theoretical limit and achieve sub–DOF axial localization.

### 4.2 Introduction to pattern matching

Pattern matching [82, 83, 84] is a technique in digital imaging processing for finding small parts of an image which match a pre–defined template pattern. An example is shown in Fig. 4-4. In this case, our goal is to find the letter “N” among the 16-grid pattern. Image pattern matching has found applications in image pattern recognition [85], quality control in industrial manufacturing [86], robot navigation [87].

The most standard method to perform pattern matching is cross–correlation. Take Fig.4-4 as an example, in order to search for letter “N”, one can cross–correlate the
Figure 4-1: A example of pattern matching.

template pattern with each grid on the 16-letter image. The ones that have the highest correlation peak will correspond to the image patch that contains the desired letter "N". Since the relative value of the correlation results matter, this method will only work if the intensity of all the candidate images are approximately on the same order. Otherwise, false correlation peak can be produced if a candidate image has a significantly higher intensity than the rest.

To avoid this problem, one can use the normalized cross-correlation [25, 88, 89]. For a template pattern \( t(x, y) \) and a candidate image \( f(x, y) \), the normalized cross-correlation is defined as

\[
\text{Corr}(f, t) = \frac{1}{n} \sum_{x,y} \frac{(f(x, y) - \bar{f})(t(x, y) - \bar{t})}{\sigma_f \sigma_t},
\]

where \( n \) is the number of pixels in \( f(x, y) \) and \( t(x, y) \), \( \bar{f} \) and \( \bar{t} \) denote the average pixel values of \( f \) and \( t \), respectively, and \( \sigma \) represents standard deviation of the associated image.
4.3 Axial super-localization via compressive pattern matching

In the previous chapters, I have described how digital holography (DH) is different from traditional imaging techniques. Holograms recorded in a DH system encode object's axial information. In this chapter, I will investigate techniques that allows super-localization using a modified compressive holography technique combined with pattern matching.

![Figure 4-5: In-line digital holography geometry.](image)

Figure 4-6: Digitally reconstructed images from the hologram shown in Fig. 4-5. (a) in-focus, (b) over-focus, (c) under-focus.

A typical in-line digital holography setup is shown in Fig. 4-5. In this setup, a ring-shaped object $g(x, y)$ is placed at $z_0$ distance from the digital camera. The
simulated hologram is shown on the right. The information recorded on hologram \( a(x, y) \) can be considered as

\[
a(x, y) \approx g(x, y) \exp \left( i \frac{\pi}{\lambda z_0} (x^2 + y^2) \right). \tag{4.4}
\]

Ideally, during the digital reconstruction, if the back-propagation distance is \( z_0 \), the reconstructed image would be a focused image, as shown in Fig. 4-6(a). If the reconstruction plane is different from the original object location, the reconstructed image is defocused. Example reconstruction images over focus and under focus are shown in Fig. 4-6(b) and (c), respectively.

By taking advantage of the intensity differences between in-focus and defocused images, pattern matching can be applied to each reconstructed images. The reconstruction distance that produces the highest cross-correlation coefficient appear at the expected object’s axial location.

The whole sub-DOF localization algorithm flow is demonstrated in Fig. 4-7. After acquiring a hologram of the object at a specific position, the hologram is reconstructed at a series of distances using the compressive holography model as in Chapter 3. The template that will be used for subsequent experiment is taken as the reconstruction having the most sparse representations.

Since it is difficult to know the exact axial position of an object, the accuracy of the proposed method is quantified by tracking an object moving along the axial direction and tracking the relative displacements and comparing them to the ground truth. For doing so, the compressive reconstruction is repeated for all the other
holograms recorded while the object is moving. For each hologram, a stack of images at various axial distances are reconstructed. For each stack of reconstructed images, normalized cross-correlation with template image is applied to each image to find the estimated axial position corresponding to the image that has the highest correlation peak value.

A simulated example is shown in Fig. 4-8. A template is chosen and then pattern matching is performed between this template and all the axial reconstructions from a different hologram. The bottom shows the normalized cross-correlation at each axial position. The plane that has the highest peak value (marked with orange dash line)
corresponds to the estimated object position.

![Diagram of pattern matching](image)

Figure 4-8: Illustration of pattern matching applied to an axial stack of digitally reconstructed image. The reconstruction plane that generates the highest cross-correlation peak value is considered as the object position.

4.4 Axial sub-DOF displacement tracking experiments

To explore how the proposed method will perform in practice, an in-line DH setup is used (details same as the one in Chapter 3). The illumination light is produced by a Helium–Neon laser of wavelength 633nm and holograms were taken by a Basler A504k camera with 1024x1024 pixels and 12μm pixel size. During the reconstruction, each hologram is cropped to 256x256. As shown Fig. 4-9 an object is placed on a single axis motorized translation stage (Thorlabs LTS300). A microscopic image of the
object is shown in Fig. 4-10(a). The initial distance between the object and camera is about 110mm. The NA of the optical system is approximately 0.014. According to Eq. 4.2, the DOF is about 3.2mm.

A sample hologram is shown in Fig. 4-10(b), which is captured at the initial location of the object. Later, during the post-processing, this initial hologram will be used to construct the template for pattern matching. In order to test whether object localization within the DOF can be achieved, the object is moved using the motion stage along optical axis and holograms are taken during the movement. The designed step size is 200μm, which is ~1/16 of the DOF; the object is moved for 50 consecutive steps.

Figure 4-9: Experimental setup of in-line DH for axial localization demonstration. An object is placed at a motion stage, which is moved along the optical axis with step size equivalent to 1/16 of the DOF of the imaging system. A sample hologram is taken at each axial position.
4.5 Result and discussion

During the post-processing, the first step is to compressively reconstruct the initial hologram. The first hologram is reconstructed between $z = 106\text{mm}$ and $z = 118\text{mm}$. Sample reconstructions are shown in Fig. 4-11. By inspection, the reconstruction at $z = 112\text{mm}$ preserves the edge of the star object most well, and is chosen as the template for subsequent pattern matching step.

![Figure 4-11: Reconstruction of the hologram taken with star object at its initial position. The reconstruction at $z = 112\text{mm}$ has been chosen as the template for subsequent pattern matching.](image)

Next, the holograms recorded at each axial localization are reconstructed using the same compressive algorithm at a series of defocus distances. Sample reconstructed
images at 6 axial planes with 200μm spacing are shown in Fig. 4-12. Although they look very similar, by calculating the normalized cross-correlations for each image, the one that is closest to the template can be found (shown in the bottom of Fig. 4-12). Since the fourth correlation has the highest correlation peak value, this indicates that 112.2mm is the best estimated axial localization of the object. The corresponding detected displacement is 200μm, which matches with the ground truth.

![Figure 4-12: Pattern matching result. Images on the second row are the reconstruction results of the hologram taken after a single step displacement. Plots on the bottom show the normalized correlation results between the template and each reconstruction on different axial planes.](image)

The hybrid compressive holographic reconstruction and template matching are repeated for all the other holograms taken during the experiment. The detected displacements for all the 50-step movements are shown in Fig 4-13. The red line is the linear fit to the expected displacement for each movement. The average detected movement is 200.6μm. The standard deviation which measures the axial localization uncertainty is ~0.156mm, demonstrating a 20 times improvement over the theoretical axial localization uncertainty. The histogram figure on the right of Fig 4-13 indicate
that the proposed approach correctly detected most of the movements and only 2 out of the 50 movements are off by more than 1 step compared to the true displacement.

![Compressive reconstruction](image)

\[
\begin{align*}
\mu_{\text{ex}} &= 0.2\text{mm} \\
\mu &= 0.2006\text{mm} \\
\sigma &= 0.156\text{mm}
\end{align*}
\]

**Figure 4-13:** The axial movement detection result by the proposed algorithm.

For comparison, the same experimental data is also processed using the traditional back-propagation approach. The same pattern matching procedure is then applied to the back-propagated images. The detected displacement for each step is shown in Fig. 4-14. The average detected displacement by conventional approach is 192\text{\mu m} and the error is 13 times higher than that using the newly proposed method. The localization uncertainty is 1.172\text{mm}, which shows almost no improvement in reducing the theoretical localization uncertainty. The histogram in Fig. 4-14 indicates that only 22\% of the total displacement are correct.

To conclude, by combining compressive holography and pattern matching, the theoretical limit of axial localization uncertainty can be reduced by at least 20 times.
Figure 4-14: The axial movement detection result by conventional back propagation approach.
Chapter 5

Vortex and wake-induced vibration characterization using compressive holoography

In the deep ocean, harbor seals swim around and search for fishes as prey. Surprisingly, in the deep sea waters, there is almost no light passing through. To make it even worse, the high turbidity of deep ocean dramatically reduces visual acuity [90]. From the perspective of optics, deep ocean conditions are not good for seals to hunt. Therefore, this leads to an interesting question: how harbor seals manage to maneuver in such a dark environment.
5.1 Introduction to whisker sensing in harbor seals

Biologists at University of Rostock's Marine Science Center[91] have done extensive investigation on how harbor seals chasing fishes in dark and murky waters. They discovered that the harbor seal sense what is going on in the surrounding area, based on monitoring how its own whiskers vibrate in the water. This discovery inspired us to design a hydrodynamic whisker sensor and install it on the autonomous underwater vehicle (AUV). The biomimetic whisker sensor would greatly improve an AUV's ability to respond to its surroundings as well as collect additional useful data from deep sea, which is unattainable with current technology. Before designing such a biomimetic sensor, it is critical to uncover how harbor seals could use vibration as sensory information. Our approach is to study the fluid mechanics of the vibration induced by whiskers' interactions with surrounding fluidic environment. This work is done in collaboration with MIT towing tank group.

To investigate the sensing mechanism of seal whisker, the first question needs to be answered is how it is possible for seals to distinguish whether whiskers' vibration is caused by a flow pattern generated by nearby fishes or are vibrating on their own. For a cylindrical object moving through steady flow, it can develop significant vibrations. This vibration is a fundamental fluid mechanics phenomenon, known as vortex-induced vibration (VIV) [92, 93, 94]. Fluid traveling over round surfaces does not stay attached, but rather shed into swirling vortices. This vortex shedding happens in an alternating pattern, creating a periodic force on the moving body, thus inducing sustained vibrations on bluff bodies [93]. As seals are swimming forward,
their whiskers would naturally experience VIV. From the signal processing point of view, VIV actually becomes background noise when seals try to sense the part of vibration that is caused by other fishes. Ideally, to have an accurate detection of prey, the whiskers are not supposed to vibrate much on their own, meaning that they should experience low “background noise” during the detection process. On the contrary, once the whiskers do encounter unsteady flows, large vibrations are expected to develop. In other words, if the whiskers encounter vortex wakes of a fish that the seal wishes to pursue, the whiskers will most likely begin vibrating. Therefore, in order to perform well in fish detection, the unsteady flow induced vibrations must be much larger than the “background noise”, the whisker’s self-induced vibrations.

In Rostock, biologists mounted a camera onto a harbor seal’s head to get a closer observation of whisker’s VIV [91]. They noticed that the whiskers did not vibrate much when the seal is swimming forward normally. The small VIV in open flow indicates that seals may experience heightened sensitivity in detecting flow disturbances, which helps seals more easily detect unsteady flow features.

Work in [95, 96] reports that the reason why seal whiskers experience low VIV is due to their unique undulatory and elliptical geometry. An SEM image of a shed whisker from the New England Aquarium is shown in Fig. 5-1. Flow visualizations show that this undulatory shape causes desynchronization in vortex shedding, which helps reduce fluid forcing on the whisker.

Work in [96] has theoretically proved that the specialized undulatory shape could reduce VIV. To explore the problem more deeply, we desire to experimentally quantify the different VIV experienced by bluff objects with different edge and cross-sectional
5.2 Whisker's vortex–induced vibration characterization using compressive holography

The diameter of harbor seals' whiskers are generally on the order of millimeter. Traditional VIV measurement techniques require accurate force and position readings, thus demanding larger size models. Our collaborators at MIT towing tank group have experimentally explored this phenomenon using a 30× scale plastic whisker model [97]. However, increasing the size of the model might change its own VIV response. Therefore, it is desired to perform VIV experiments with models whose size is on the same order as the real whiskers. Our approach is to optically monitor the small-scale whisker's vibration using compressive holography technique, as detailed
in chap. 2-4.

Figure 5-2 shows the imaging geometry of the in-line digital holography setup used for imaging vortex-induced vibration from different whisker models. In front of the camera, it is a tank filled with distilled water. An upside down “L”-shaped post and a piece of silicon gel are used to hold the whisker vertically with the majority of its body immersed in water. This allows the whisker to freely vibrate. In order to accurately study VIV properties, we specially select a gel with appropriate stiffness which makes the natural frequency of the system close to Strouhal frequency (St) of 0.2 (St= fd/U, where f is the natural frequency of the system, d is the diameter of the cylinder object and U is the flow speed). The other end of the “L”-shaped post is fixed onto a motion stage. The motion stage can be moved along the direction parallel to the length of the tank and parallel to the optical axis. The motion stage used here has a maximum speed of 420mm/s and maximum travel range of 1000mm. Figure 5-3 shows the actual experimental setup in the lab. The magnified figure in Fig. 5-3 shows a real harbor seal whisker hanging in the water. In order to guarantee there are enough temporal samples during each VIV cycle, we use a high speed camera (up to 500 frames per second) to capture the hologram movies.

During the experiments, the motion stage is moved along the optical axis towards the camera, mimicking a seal is swimming along the same direction. As described above, vortices would be shed behind the whisker model while being towed and exert periodic forces back onto the whisker model. This transverse force will induce the whisker to vibrate along the transverse direction, which is perpendicular to the optical axis direction. Our goal is to use compressive holographic imaging approach to
validate that whiskers experience low VIV in open flow due to their special geometry.

In order to compare VIV responses, we carried out the VIV characterization experiments on three different models, shown in Fig. 5-4. To attain decent VIV observation while keeping the model size on the same order as real seal whiskers, the fabricated models are twice larger than the average size of a normal whisker. All the three models were 3D-printed with the same plastic materials. The first one, shown in Fig. 5-4(a) is a whisker model with undulatory edges. The period of the whisker’s edge is 10.8mm and its cross-section is an oval shape with a major axis of 3.2mm and a minor axis of 1.6mm. Figure 5-4(b) shows an elliptical cylinder model with the same oval cross-section as the whisker model, and the difference is that the edge of the elliptical model is smooth rather than undulatory. The third model, displayed in Fig. 5-4(c) is a smooth cylinder with a circular cross-section. The diameter of the circle is 1.6mm, matching the size of the minor axes of the whisker and elliptical cylinder models. All the three models have the same length of 114mm and for each
test, the submerged span is 45mm.

To compare the different VIV responses on the three models, five groups of experiments were carried out. The first two groups of tests were performed on the whisker model. One group was done with the major axis of whisker model facing the towing direction (90 degree orientation), while the other group was done with its minor axis facing the towing direction (0 degree orientation). The third and forth groups of experiments were done with the ellipse model, with 0 and 90 degree orientations respectively as well. In the fifth group, the experiments were performed with the circular cylinder. Since the circular cylinder has a uniform cross-section and there is no orientation difference. In each group, the model was towed with a uniform speed ranging from 100mm/s to 450mm/s.

Holographic movies were captured as each model was towed in the tank. A sample
Figure 5-1: Images of the three plastic, 3D-printed models used in the tests. Each model was held by a piece of silica gel at the base and hanged vertically in the water. (a) 3D undulatory whisker model, with an elliptical cross-section; (b) 3D elliptical model with smooth edges; (c) 3D cylindrical model with smooth edges.

hologram of the whisker model is shown in Fig. 5-5(a). In order to eliminate the effects from camera noise and twin image problem, compressive reconstruction was used (described in Chapter 3) to each captured hologram. The relationship between the edge function of each model $a_{\text{edge}}$ and the information recorded on each hologram $I$ can be expressed in the form of

$$S \cdot \mathcal{F} \cdot I = H \cdot \mathcal{F} \cdot a_{\text{edge}} + n,$$

(5.1)

where $S$ denotes the spiral phase mask function, $\mathcal{F}$ represents the 2D Fourier transform operator, $H$ is the Fourier transform of Fresnel kernel and $n$ stands for the sum of system noise and twin image. The edge signal $a_{\text{edge}}$ of each model can be obtained
by minimizing $\ell_1$-norm $\|a_{\text{edge}}\|_1$ with (5.1) as a constraint. We implemented the $\ell_1$ minimization by adapting the Two-step Iterative Shrinkage/Thresholding (TwIST) algorithm [65].

5.3 Vortex–induced vibration characterization results

The compressive reconstruction of Fig. 5-5(a) is displayed in Fig. 5-5(b). It clearly shows the undulatory shape edge of the whisker model. With the reconstructed edge signals at each frame, cross-correlation of the edges between successive frames allows for the object’s position over time to be tracked. Figure 5-6 displays an example of the cross-flow VIV pattern of the whisker model. The whisker was towed with its major axis facing the towing direction and was moving at a speed of 200mm/s. The plot in Fig. 5-6 shows that for this specific towing speed, the detected VIV has an average vibration amplitude of 3.25mm and vibration frequency of 12.5Hz.

5.3.1 VIV response comparison between whisker, ellipse and cylinder models

The VIV response results of the five groups of tests are plotted in Fig. 5-7. In order to exclude the influences from different natural frequencies $f_n$ and characteristic lengths $d$ in each model, the towing speed and detected VIV amplitude are normalized to non-dimensional quantities, reduced towing speed ($\frac{U}{d/f_n}$) and reduced vibration amplitude
Figure 5-5: (a) A sample hologram of the bottom part of the plastic whisker model (b) The compressive edge reconstruction of the sample whisker hologram shown in Fig. 5-5(a).

\[(A/d)\), where \(U\) is the actual towing speed and \(A\) is the detected VIV amplitude. Natural frequency \(f_n\) for each model can be obtained by tracking the model’s vibration after being naturally plucked. Table 5.1 shows the characteristic length and natural frequency for each group of tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>Whisker model</th>
<th>Ellipse model</th>
<th>Circular cylinder model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90 degree</td>
<td>0 degree</td>
<td>90 degree</td>
</tr>
<tr>
<td>Natural frequency (Hz)</td>
<td>orientation</td>
<td>orientation</td>
<td>orientation</td>
</tr>
<tr>
<td>Natural frequency (Hz)</td>
<td>11.21</td>
<td>10</td>
<td>12.5</td>
</tr>
<tr>
<td>Characteristic length (mm)</td>
<td>3.2</td>
<td>1.6</td>
<td>3.2</td>
</tr>
</tbody>
</table>

By comparing those two groups of results with both whisker model and ellipse model facing along 90 degree orientation (corresponding to the pink and green lines in Fig. 5-7), we can observe that whisker model experienced lower reduced VIV amplitude than the ellipse model. Moreover, if we compare the other three groups of
Figure 5-6: A sample plot of detected VIV response of the whisker model. The whisker was placed along 90 degree orientation and towed at a speed of 200mm/s. Tests where each model has the same orientation and the same characteristic length of 1.6mm (shown as blue, red and black lines in Fig. 5-7), whisker model still presents the lowest VIV amplitude among the three models. From the five sets of experiments, we can conclude that whisker model does experience low VIV in open flow due to its special geometry.

Figure 5-7: The VIV responses comparison among all the three models, with 0 and 90 degree orientations.
5.3.2 VIV response comparison between $3\times$ and $30\times$-scale whisker models

The VIV comparison between different scale models has been performed as well. Figure 5-8(a) shows the detected VIV responses in the groups of tests where both whisker and ellipse models face the 0 degree orientation. The plots have been shown in Fig. 5-7 and here they are plotted in a separated figure to show a more clear comparison with the larger size models. Fig. 5-8(b) shows the measured VIV responses of a $30\times$-scale whisker, ellipse and cylinder models. Each plastic model (average crossflow diameter $d = 1.59$cm, length $l = 27.5$cm) is connected to a four-armed flexing plate with strain gauges (Omega, Model KFH-03-120). Fig. 5-9 depicts the isometric view of the $30\times$-scale whisker sensor. As the whisker and flow field interact, the four arms bend accordingly, allowing free vibration in crossflow directions. A voltage signal is read from the strain gauges, and this is then calibrated to the deflection at the model tip. More detailed explanation about the flow sensor design can be found in [97]. The results obtained with the $30\times$-scale models have very similar trends to the results performed on the $3\times$-scale models. The peaks appear on the black lines in both plots are caused by the "lock-in" phenomena. This is because the fluid and the structure operate in a feedback loop. When the frequency of the transverse force exerted by vortex shedding synchronizes with the motion frequency of the model, the body force increases the strength of the vortices, which in turn increases lift and drag on the model and then results in high oscillation amplitudes.

Regarding the difference between the actual VIV amplitudes in the two plots,
there can be a few reasons for that. One is that the two different experiments are
done with different Reynolds numbers. Another reason is that the target areas on
the models are different in the two tests. The tests performed with $3 \times$-scale model
was targeted to the bottom tip of each model while the VIV on the $30 \times$-scale models
were measured at the top tip of each model.

Figure 5-8: (a) The measured VIV on $3 \times$ whisker model, using compressive holography approach. (b) The detected VIV response of the $30 \times$ whisker model, measured by strain gauges. Each model was positioned in its streamlined orientation(0 degree).
Figure 5-9: The 30× scaled-up whisker model used for VIV detection via strain measurement. Experiments were conducted by Dr. Heather Beem. The model whisker was mounted on a set of flexing plates that allow free vibration in in-line and crossflow directions. Bushings are used to reduce coupling between the axes.

5.4 Whisker in vortex wake: wake-induced vibration characterization

In Section 5.2, the experiments have demonstrated that due to the specialized undulatory geometry of the whiskers, seals experience low self-reduced noise from VIV while they are moving forward in the ocean. However, another question remaining unsolved is how the whiskers detect unsteady flows resulted from a passing-by body in the surrounding area.

To answer that question, we turned to the problem of characterizing wake-induced vibration (WIV) [98, 99]. WIV occurs when a flexibly mounted object is placed in the wake of an upstream object, subject to the unsteady forcing of its vortical structures. Studies carried out in [98, 99, 100, 101] have demonstrated that the vibration amplitude and frequency of the downstream body would be greatly affected by the upstream body. This indicates that whenever there is a fish passing by, the
vortical wake of the fish will have a significant impact on the vibration of seal’s whiskers. The changes in whiskers’ vibration would inform seal of the change in surrounding environment.

In this section, we studied the WIV response of a flexibly mounted plastic whisker model placed within the vortical wake of a larger upstream circular cylinder. Experiments in which the whisker model was first replaced by an elliptical model and then a circular model were also performed, in order to compare WIV responses between different models. The three models used in this section are the same as the ones used in Section 5.2.

5.4.1 Wake–induce vibration characterization experiments

The WIV responses of the models were characterized by the compressive holography approach as well. An in-line digital holography system was built in the lab to conduct the experiments. Fig. 5-10 shows a top view of the relative positions between the whisker model and the upstream cylinder model. The whisker was placed in its 0 degree orientation with a characteristic length $d_w$ of 1.6mm. The diameter of the cylinder $d_{cyl}$ is 6.4mm, which is 4× whisker diameter. The longitudinal separation $d_z$ between the whisker and cylinder models is 28mm, which is about 17.5× of whisker’s characteristic length. In order to prevent the axial overlapping between the two bodies, the whisker was placed 8.8mm away from the upstream cylinder along the transverse direction. This transverse separation is about 5.5× of whisker’s characteristic length. [98, 99, 100, 101] have shown that the upstream cylinder still has
significant effects on the WIV of downstream body even when they are placed 25 diameters apart. Fig. 5-11 shows the real laboratory test imaging system. The whisker model is hanging by a blue holder and a piece of silicon gel, allowing it to freely vibrate while being towed. A cylinder was placed to the front of the whisker model with the separation dimension described above. A digital camera (Basler A504k) was used to holographically capture the WIV responses of each model.

![Diagram of towing direction and relative positions](image)

**Figure 5-10**: Illustration of the relative positions between the upstream cylinder and the downstream whisker model in the WIV characterization experiments.

### 5.4.2 Detected Wake–induced vibration among different models

Three groups of experiments have been carried out on the three models respectively. All the tests were done with the models oriented along its streamlined direction (0 degree). The models' reduced WIV amplitude responses are plotted in Fig. 5-12 with solid lines. And the dash–line plots correspond to the VIV response for each
model with the same orientation as the WIV tests. By comparison, it is obvious that all the three models encountered significantly larger vibrations in the cylinder's wake compared to the open flow case. The whisker model would vibrate with approximately 10× higher amplitude when it encounters the wake than it does in the open water. Furthermore, among all the three models, the whisker model has the highest increased vibration ratio between WIV and VIV.

Fig. 5-13(a) and (b) display the detected WIV responses for whisker models made with two different scales. Fig. 5-13(a) plotted the WIV for 3× plastic whisker model (same as the plots shown in Fig. 5-4) measured by compressive holographic approach, while Fig. 5-13(b) shows a similarly conducted WIV experiment for 30×–scale up whisker model (shown in Fig. 5-9), measured by the strain gauge attached to the tip of the model [97]. In both figures, the red lines denote whisker models' in–wake
Figure 5-12: Comparison of VIV and WIV responses among the three models. The results shown here are all with the models positioned along their streamlined orientation (0 degree).

vibrations and the black lines represent the VIV responses in open flow. The two plots indicate that whisker models vibrate with more than an order-of-magnitude larger amplitude when in the wake of the upstream body than in open water. Although the two sets of experiments are performed with different scale models, the trends of how the detected WIV and VIV change along towing speeds in the two models are extremely similar. We attribute the absolute value differences of the reduced vibration amplitude to the different test parameters ($d_{cyt}$, $d_z$, $d_w$, Reynolds number, etc.) in the two sets of experiments.

Also of note is the frequency responses for the three models when they are in the wake of upstream cylinder. The detected WIV frequency is shown in Fig. 5-14 where the nondimensional frequency of vibration, $f/f_n$, is plotted versus the nondimensional towing speeds. Each frequency value was obtained by selecting the peak
Figure 5-13: Comparison of VIV and WIV responses measured on different scaled size models. For both models, the whisker vibrate with far higher amplitude in the wake of upstream body than it does on its own. (a) Detected vibration responses for the 3x–scale whisker model, measured by compressive holography approach. The whisker is within the wake of an upstream cylinder \( (d_{\text{cyl}} = 4d_w) \) at a distance \( d_z = 17.5d_w \), for WIV tests. (b) Characterized vibration responses of the 30x–scale model. The whisker is within the wake of an upstream cylinder \( (d_{\text{cyl}} = 2.5d_w) \) at a distance \( d_z = 8d_w \), for WIV tests.

value of the Fourier transform of corresponding WIV response. Each plotted WIV frequency response mainly contains three trends, which have been marked with different colors in Fig. 5-14. In the orange region where towing speeds are low, the wakes of upstream cylinder are still relatively weak and models responded with relatively constant frequency oscillation. As this stage, the vibration of the models are mainly caused by its own VIV. In the region with medium towing speeds (appearing as pink in the plot), the wake of upstream object becomes stronger and models' vibration starts to synchronize with the oncoming vortical wake. And at this stage the vibration response is a mixture of VIV and WIV. When the strength of wakes from the upstream cylinder continues to increase (shown in the blue region), the whisker will synchronize with the oncoming wake dominant frequency and it oscillates at the
vortex shedding frequency of the upstream body.

Figure 5-14: The detected reduced WIV frequency for $3 \times$ whisker model. When the wake of the upstream cylinder becomes strong, the peak frequency of the whisker’s vibration would synchronize with the frequency of the oncoming wake.

In conclusion, the experiments above prove two important points that can be used by whisker sensing. One is the whisker models normally vibrate with a relatively low amplitude in the open water, while the vibration amplitude can increase by an order-of-magnitude when encountering an oncoming vortical wake from an upstream object. The other is that the frequency of whisker’s vibration will synchronize with the dominant frequency of the oncoming wake from the upstream body.
Chapter 6

Weak signal detection via
null-space estimation and sparse
reconstruction

6.1 Introduction to weak signal detection

In this chapter, I will consider a class of problem that aims to detect weak signals from low signal-to-noise ratio (SNR) measurements along with low signal-to-background ratio (SBR). This problem is encountered in many practical situations. For example, in defect detection, the signal of interest locates at the scattered wave from the defect region, whose power decreases rapidly as the defect size becomes smaller. Detecting the defect signal is often difficult because the received signal is also accompanied with both random noise and other structured (non-random) signals from "background". If the total power of the signal of interest is small as compared to the total power of
the noise, we need to deal with the low SNR problem. If in addition, the total power
from the background signal is strong, we also need to consider the low SBR problem.

The good news is that in many cases, including defect inspection, one does not
need to accomplish the “signal reconstruction” task, which looks for a complete rep-
resentation of the signal (e.g. position and shape of the defect). Rather, it is only
necessary to perform the “signal detection” task, which tries to determine whether a
signal of interest (e.g. defect) exists. In the latter case, a rough estimate of relevant
coefficients representing the signals is sufficient. As a result, it is possible to achieve
signal detection with data having much lower quality (i.e. low SNR and low SBR).

To estimate relevant coefficients that represent the signals, sparse reconstruction
based method is of interest as it can accurately reconstruct a sparse representation of a
signal from a few linear measurements. Moreover, sparse reconstruction is robust from
low SNR measurements [102, 103]. Intuitively, this is because although the total power
of the noise is strong, it will be evenly spread over the reconstruction (sparsifying)
basis as the noise will remain random. On the other hand, most power of the signal
of interest will be concentrated to the few non-zero coefficients of the reconstruction
basis, the SNR is thus effectively enhanced using this sparse reconstruction process.

The low SBR problem, on the other hand, cannot be directly solved by sparse
reconstruction method. The reason is that the representation of the background
signal under the signal-sparsifying basis may also be sparse and occupy similar sets
of coefficients as the signal of interest. As a result, it is important to design a robust
pre-filtering process that can effectively separate the background from the signal of
interest. In this chapter, I will demonstrate a background–signal separation method
based on a null-space projection algorithm [104].

6.2 Null-space based background–signal separation

The null-space of an $m \times n$ matrix $A$ is the linear subspace that is constituted by all solutions to the equation

$$Ax = 0. \quad (6.1)$$

In set notation, that is

$$\text{Null}(A) = \{x : x \in \mathbb{R}^n \text{ and } Ax = 0\}. \quad (6.2)$$

For any vector $b$ belonging to $A$'s null-space, it satisfies Eq. (6.1). In other words, if the vector $b$ is projected onto $A$, i.e. $A \cdot b$, the result will be 0 vector. This implies that the information of vector $b$ will be totally removed after this projection operation. Next, we consider a mixture signal that contains a linear sum of $b$ and another vector $o$. Let us further assume that the vector $o$ does not belong to $A$'s null-space, i.e. it cannot be represented as the linear sum of any combinations of vectors in $A$'s null-space. After projecting this signal onto $A$, information about $o$ can be preserved from the mixture signal, as $b$ will be totally removed.

Typically, the null-space of a matrix can be estimated using the method of singular value decomposition (SVD) [105, 106]. The SVD of an $m \times n$ matrix $A$ takes a
factorization form as

\[ A = U\Sigma V^* \]

\[ = (u_1 \ldots u_r \ u_{r+1} \ldots u_m) \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \sigma_{r+1} \\ & & & \ddots \\ & & & & \sigma_n \end{pmatrix} \begin{pmatrix} v_1^* \\ \vdots \\ v_r^* \\ v_{r+1}^* \\ \vdots \\ v_n^* \end{pmatrix} \]  

(6.3)

where \( U \) is an \( m \times m \) unitary matrix, \( \Sigma \) is an \( m \times n \) rectangular diagonal matrix with non-negative real numbers on the diagonal, and \( V^* \) (the conjugate transpose of \( V \) ) is an \( n \times n \) unitary matrix. All the column vectors \( u_i \) in matrix \( U \) are the eigenvectors of matrix \( AA^* \) and all the column vectors \( v_i \) in \( V \) are the eigenvectors of matrix \( A^*A \). The diagonal entries \( \sigma_i \) in \( \Sigma \) are known as singular values of \( A \) and are typically arranged in descending order. The non-zero singular values of \( A \) are the square roots of the non-zero eigenvalues of both \( AA^* \) and \( A^*A \). As a result, all the three matrices \( U, \Sigma \) and \( V \) can be obtained through eigenvalue decomposition.

Mathematically, the null-space of \( A \) is a set of columns in \( V \) that correspond to those zero value singular values. Since the singular values are arranged in descending order, without loss of generality, we can assume that in Eq. (6.3), all the singular values corresponding to \( \sigma_{r+1} \) to \( \sigma_n \) are zero. Based on this assumption, we can make the conclusion that the null-space of \( A \) is a subspace that is constituted by the column

104
vectors \((v_{r+1} \ldots v_n)\) in \(V\). To show this, we can calculate the product of matrix \(A\) and any vector that is a combination of the column vectors \((v_{r+1} \ldots v_n)\). For a simple example, let us calculate \(A \cdot v_{r+1}\).

\[
A \cdot v_{r+1} = (u_1 \ldots u_r u_{r+1} \ldots u_m) \cdot \begin{pmatrix}
\sigma_1 \\
\vdots \\
\sigma_r \\
0 \\
\vdots \\
0
\end{pmatrix} \cdot \begin{pmatrix}
v_1^* \\
\vdots \\
v_r^* \\
v^*_{r+1} \\
\vdots \\
v_n^*
\end{pmatrix} = (u_1 \ldots u_r u_{r+1} \ldots u_m) \cdot \begin{pmatrix}
v_1^* \cdot v_{r+1} \\
\vdots \\
v_r^* \cdot v_{r+1} \\
v^*_{r+1} \cdot v_{r+1} \\
\vdots \\
v^*_{n} \cdot v_{r+1}
\end{pmatrix}
\]

(6.4)

Because of the unitary property of matrix \(V\), the inner product between different
columns in $V$ is all zero. As a result, Eq. (6.4) can be rewritten as

$$A \cdot v_{r+1} = (u_1 \ldots u_r u_{r+1} \ldots u_m) \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_r \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = 0. \quad (6.5)$$

From the result in Eq. (6.5), we see that the vector $v_{r+1}$ belongs to the null-space of $A$. For the same reason, we can also conclude that any vector that is a linear combination of $v_{r+1}, \ldots, v_n$ is a member of $A$'s null-space. In other words, the null-space of $A$ is the set of column vectors in $V$ that correspond to those having zero singular values.

The above review about matrix null-space suggests that an effective way to separate the background from the signal of interest can be constructed if one can find a projection matrix $A$ whose null-space contains the background signal vector $b$. To do so, we require the matrix $A$ satisfying the equation

$$A \cdot b = 0. \quad (6.6)$$

In practice, only the information about the background $b$ is contained in the measurements, one has to find $A$’s null-space based on the measured data. This "inverse
problem” can be considered by invoking the property of matrix transpose, and reformulate Eq. (6.6) as

\[ b^T \cdot A^T = 0^T, \]  \hspace{1cm} (6.7)

where \( T \) denotes the matrix transpose operator. Equation (6.7) indicates that the transpose of the projection matrix \( A \) can be treated as the null–space of a background matrix \( B \). In order to recover the entire null–space of \( B \), which is typically of dimension larger than one, multiple “repeated” measurements about the background is required, with the number of measurements larger than the dimension of the null-space. The background matrix \( B \) is then constructed with each column from a unique set of measurement \( b_i \). In practice, these repeated measurements are different due to random noise present, however, they contain the same underlying background structure, which defines the invariant null–space. By doing so, the projection matrix \( A \) can be estimated by calculating \( B \)'s null–space via SVD.

The projection matrix \( A \) defined from above process will guarantee that, given a mixture signal containing both signal and background, the background can be maximally removed, leaving only the signal of interest and noise after the projection operation. The background removed signal is then fed to the sparse reconstruction algorithm to remove the noise, leaving only an estimate about the significant coefficient about the signal of interest. Next, I will demonstrate how the proposed method is applied to a real industrial application.
6.3 Acoustic crack detection for underwater platforms

Underwater platforms are widely used in the deepsea oil drilling industry. The metallic platforms usually remain in water for years after being installed. Cracks are the prime safety concern after long deployment. The possible cause includes fatigue, occurring when a material is subjected to repeated loading and unloading, and corrosion by sea water due to its high acidity and alkalinity. Usually a crack starts to form in microscopic size and then grows. Eventually when the crack reaches a critical size, the crack will propagate and results in fracture in the platform structure. Therefore, detection of cracks in the platform frame is of critical importance; yet present technology does not have the ability to automatically detect fine cracks. Moreover, after being underwater for a long time, there would be a thick layer of marine growth on the platform, which creates a complicate background and makes it even harder to detect the crack. Currently, offshore operators rely on remotely controlled underwater vehicles (ROVs) with scrubbing tools, or divers, to remove marine growth on pipelines, or subsea structures, in order to perform crack detection analysis. This task is costly and dangerous for human divers, and is currently limited in depth. While using ROVs is safer, it is still limited by tether management and uncertainty in the sub-sea environment. In this section, I propose a new way to detect cracks that can potentially lead to technologies for untethered and automatic detection of cracks in deeper regions.

A typical crack is shown in Fig. 6-1. A lightening-shaped structure is used to
Figure 6-1: A lightning-shaped crack appears on the surface of a platform, whose surface is covered with a thick layer of marine growth

illustrate a crack on the platform and is covered by marine growth. The background interference created by the marine growth is too strong for the crack to be directly detected. Optical imaging is not effective in this application due to severe absorption and scattering of light in sea water [107, 108]. The proposed approach here is to use acoustic imaging method for crack detection. Figure 6-2 shows a schematics of the proposed setup. Acoustic wave is sent onto the platform surface and a sensor receives the returned acoustic signals reflected from the platform. Typically, the same ultrasound sensor can be used as both the sound emitter and receiver.

Figure 6-2: The underwater crack detection setup using acoustic wave
The platform is probed by wave produced with the acoustic sensor. If the platform does not contain crack, the only back-scattered signal to the detector is what we refer to as “background.” In our case, the background signal is mainly from the thick layer of marine growth attached to the platform surface. Let us denote this background signal as \( b(x) \), where \( x \) is the spatial (e.g. Cartesian) coordinates. If a crack exists, an additional signal \( c(x) \) will be back-scattered. The details about the signal \( c(t) \) are of interest, as they may reveal the location and shape of the crack.

As we have mentioned earlier, the determination of the scattered signal in its full representation is referred to as “reconstruction” problem. However, in many cases of practical interest and especially when the measurement conditions are challenging, the simpler “detection” problem is sufficient to decide if further action should be taken (e.g. detailed inspection of the platform.) In the detection problem, instead, only a binary classification, e.g. “1”; if the signal \( c(x) \) exists at all in the measured data and “0” otherwise, is sufficient. For simplicity, the following general notation will be used to denote the received signal \( r(x) \)

\[
    r(x) = S \cdot [b(x) + p \cdot c(x)].
\] (6.8)

where \( p \) takes the value 1 if there is a crack and 0 if there is none, and \( S \) denotes scattered wave propagation between platform and acoustic sensor. The received signal is also contaminated by noise in the measurement. In the case of additive noise, which
is most common in practice, the total received signal is then

$$r(x) = S \cdot [b(x) + p \cdot c(x)] + n(x).$$  \hspace{1cm} (6.9)

where $n(x)$ denotes noise that follows a random process with a given statistics, e.g. Gaussian. The detection problem can be cast as determining the value of $p$.

Even though we have reduced the problem to one of great simplicity, it remains challenging because $b(x)$ and $n(x)$ can have powers that are orders of magnitude higher than the crack signal $c(x)$. That is, the unknown signal $c(x)$, if it exists at all, is buried under the background and noise (i.e. low SNR and low SBR).

To extract $c(x)$ from such low SNR and SBR measurements, I propose the following three-step process. The general flow of the algorithm is shown in Fig. 6-3.

The first step of the algorithm is to remove the strong background signal $b(x)$ via null-space estimation and projection. The mathematical theory related to null-space based signal separation has been explained in section 2. To remove $S \cdot b(x)$, it is necessary to find the projection matrix $A$ whose null-space contains the background signal $S \cdot b(x)$. Therefore, the first step is to collect many repeated measurements of the scattered background signals (with no crack exists) and reformat them into a background training matrix $B$. The second step is to perform SVD on the background training matrix $B$ for estimating its null-space. From section 2, it is shown that a proper choice of the projection matrix $A$ can be the conjugate transpose of $B$'s null-space.

In the second step, a new measurement $r(x)$ (with or without crack) is processed by
projecting the data onto the pre-calibrated matrix $A$. The results can be interpreted as

$$ A \cdot r(x) = A \cdot \{S \cdot [b(x) + p \cdot c(x)] + n(x)\} $$

$$ = A \cdot S \cdot b(x) + p \cdot A \cdot S \cdot c(x) + A \cdot n(x) $$

$$ = p \cdot A \cdot S \cdot c(x) + A \cdot n(x). \quad (6.10) $$

Note that the strong background interference $S \cdot b(x)$ is removed since it is in the null-space of $A$.

The third step is to estimate $p$ from the remaining signal obtained in Eq. (6.10). Although background interference has been eliminated, there is still heavy noise existing in the remaining signals. In this step, sparsity based reconstruction algorithm will be developed to extract crack signal $p \cdot c(x)$ from the low SNR measurement. The linear model used can be formatted as

$$ y = A \cdot r(x) = (A \cdot S) \cdot [p \cdot c(x)] + A \cdot n(x), \quad (6.11) $$

where $A \cdot S$ represents the linear transformation, $p \cdot c(x)$ is the signal to detect and $A \cdot n(x)$ represents the projected noise.

The reconstruction algorithm used next relies on the assumption that the underlying signal of interest is sparse at some pre-determined “sparsifying” basis. There are many ways to choose the sparsifying basis. Standard bases that have been applied to natural signals include total variation [109], wavelets [32], discrete cosine
transform [36]. For dealing with a specific signal class (e.g. scattered signals from crack), it might be more fruitful to use basis obtained from a class of basis learning algorithm known as dictionary learning [110, 111]. The sparsifying basis, i.e. the dictionary, is "learnt" from a bank of pre-collected crack signals using algorithms such as K-SVD [112].

Let us denote the sparsifying basis for crack signal as $\psi(x)$ and the crack signal can be represented as

$$c(x) = c_s(x)\psi(x),$$

(6.12)

where $c_s(x)$ denotes crack's coefficients under the $\psi$-basis. The linear model in Eq. (6.11) can be reformatted as

$$y = (p \cdot c_s(x))(A \cdot S \cdot \psi) + A \cdot n(x).$$

(6.13)

Since the projection matrix $A$, the scattering model $S$ and the sparsifying basis $\psi$ are all known, $p \cdot c_s(x)$ can then be estimated using standard sparse reconstruction algorithms. It should be noted that this method assumes that the noise is not sparse in $\psi$, which is true in most applications. Finally, a criteria is then applied to the estimated coefficients $p \cdot c_s(x)$ for classification (i.e. if crack exists or not).
6.4 Fine crack detection simulation

6.4.1 Low SNR and SBR crack signal detection

To validate the proposed algorithm, a series of tests are performed with simulated crack signal and realistic platform background\footnote{downloaded from http://indianwildlifeblog.com/2014/06/how-to-prevent-marine-growth-on-marine-vessels/}.

Based on the description of the method, sample signals of platform background need to be collected before any crack detection. Figure 6-5 shows a sample picture of an underwater platform that is fully covered with a thick layer of marine growth (an optical image is used to approximate the 2D background structure encountered in acoustic imaging). To construct the background training matrix, the platform picture is first divided into 170 sub-images (shown in Fig. 6-5). Each sub-image is considered as a training region of the background. Next, the scattered signal from each region is simulated and used to construct $B$. Here, the near-field diffraction model is adopted to simulate acoustic scattering from the background structure using Fresnel propagation. After constructing the training matrix, the null-space of $B$ is calculated to find the projection matrix $A$ (shown in Fig 6-6).

Next, we consider a new measurement and determine if a crack exists. We first simulate a mixture signal containing both a background signal $b(x)$ and signals from three line-shaped cracks $c(x)$ (shown in Fig. 6-7). Since the simulated crack signal is sparse in its original pixel representation, so the sparsifying basis is taken as the naive pixel basis. The background $b(x)$ is similar to the platform sub-images shown
in Fig.6-5. The total scattering signal $r(x)$ is taken as the Fresnel propagated field from $b(x) + c(x)$ with added white Gaussian noise. The simulated SBR is -17dB and SNR is -2dB. Figure 6-8(a) shows the simulated scattering signal. It is clearly seen that the background resulting in a much stronger scattering signal than the crack signal. After projecting the total scattering data $r(x)$ onto the pre-calibrated background matrix $A$, the background is effectively removed and the remaining signal is shown in Fig.6-8(b). By comparing Fig. 6-8(a) and (b), we can see that most of the background has been removed. The remaining project signal can be used to estimate the sparse coefficients of the signal. Figure 6-8(c) shows the reconstruction, demonstrating successful reconstruction of the original line-shaped cracks.

To study how the proposed algorithm will perform when there is no crack exists, a background-only signal is simulated. Figure 6-9(a) shows the estimated coefficients using the proposed algorithm. The result contains many random spikes which originates from the random noise $n(x)$. As expected, the estimated coefficients are not sparse because the noise is not sparse in crack’s sparsifying basis. As compared to the previous case, the estimated coefficients is much more sparser when a crack is present.

The simulated data were also processed by the traditional back propagation method, shown in Fig. 6-9(b). Without removing the background, the crack signal is entirely swamped by the strong background interference.
6.4.2 Detection rate v.s. SNR

SNR can greatly affect the effectiveness of the proposed method for crack detection. To study how robust the proposed method performs under various SNR levels, simulations with 50 different SNRs between 40dB and -15dB are repeated. Furthermore, in order to quantify both the true positive detection rate (correctly output "1" when cracks are present) and true negative detection rate (correctly output "0" when no cracks are present), 500 randomly generated samples with cracks and 500 samples without cracks are simulated at each SNR level. Figure 6-10(a) shows how true positive detection rate changes as the SNR decreases. As long as SNR is above -5dB, the true positive detection rate remains higher than 80%. However, as SNR reaches -5dB and lower, the true positive detection rate drops dramatically and the algorithm will completely fail after SNR reaches -7dB. Figure 6-10(b) shows the true negative detection rate as SNR increases. It is seen that the true negative detection rate remains higher than 90% in all the SNR levels studied.

It is concluded that the proposed hybrid algorithm based on null-space estimation and projection, and sparse signal reconstruction, for weak crack signal detection is effective for SNR higher than -5dB.
Obtain underwater platform scattering signal training set B

SVD

Calculate the null space of B via SVD

\[ A = \text{Null}(B)^* \]

Get the projection matrix A

\[ A \cdot m \]

Remove Background

\[ y = A \cdot m = A \cdot S \cdot \Psi \cdot p \cdot c(x) + A \cdot n(x) \]

Reconstruct signal

\[ p \cdot c(x) \] via CS

Determine crack’s existence based on the sparsity of the reconstructed signal

A new test data

\[ m = S \cdot [p \cdot c(x)+b(x)]+n(x) \]

\[ c(x) = \Psi \cdot c_s(x) \]

Figure 6-3: The autonomous crack detection algorithm flow
Figure 6-4: An example of typical cracks on the platform

Figure 6-5: An example of platform buried under marine growth. The whole image is divided into 170 sub-images and each of the sub-image can represent an example of platform background
Figure 6-6: The estimated projection matrix $A$

Figure 6-7: The crack signal used for simulation

Figure 6-8: (a) The scattering measurement of the mixture signal. This image shows the scattering simulated at the acoustic sensor plane; (b) Plot of the remaining information after the projection. After the projection, most of background interference has been removed; (c) The compressively reconstructed crack signal.
Figure 6-9: (a) The compressively reconstructed signal when there is no crack in the simulation; (b) The reconstruction result from scattering measurement of the mixture signal using traditional back-propagation method.
Figure 6-10: (a) Plot of true positive detection rate v.s. SNR. Each detection rate is obtained from 500 simulations with randomly generated cracks; (b) Plot of true negative detection rate v.s. SNR. Each detection rate is obtained from 500 simulations without cracks.
Chapter 7

Conclusion

With the fast development in semiconductor fabrication field, CCD and CMOS cameras with high sampling rate, high frame rate and large dynamic range become available. Furthermore, advances in computing apparatus makes the implementation of complex algorithm more efficient. By combining these technologies with classical optics, computational imaging becomes the trend of optics for the next generation.

When imaging with digital sensor, to localize the object with sub-pixel accuracy or to detect the object signal from low SNR images can always be an issue. In this thesis, digital holography, the 3D imaging technique, has been explored to tackle those problems. Digital holography is widely used to reconstruct 3D profiles of objects from a single intensity frame. Unlike the traditional point-to-point imaging technique, during the free space propagation of the object field in holography, the information of the object is spread out onto the entire digital sensor and object’s diffraction pattern is recorded on the sensor. Thus by taking advantage of information spreading in holography, it can greatly increase the chance to detect the weak signals.
In the case of detecting sub-pixel movement or locating object with sub-pixel accuracy, the information is under-sampled. To tackle this problem, compressive sensing has been implemented here together with digital holography. Compressive sensing says that if a signal is sparse, it can be accurately reconstructed even from a small amount of linear measurements. The number of measurements has a lower bound, which is inversely proportional to the sparsity of the signal as well as the "incoherence" parameter that measures how incoherent the sensing matrix and the sparsifying basis are relative to each other. The solution of compressive sensing is also robust to noise.

Given that most of the general objects are not sparse in their natural spatial coordinates, here we chose different sparsifying basis to convert general signals to sparse signals. For a 1D object signal, the derivative operator is a good choice. However, since the 2D equivalent would constitute a Hilbert transform that does not exist, so a different sparsifying basis, namely the spiral phase mask, was implemented to sparsify a general 2D object. By implementing the sparsifying operations, compressive reconstruction can successfully recover the edges of the object directly from the hologram captured in the experiment, free from artifacts due to twin image and other sources of noise.

The experimental results also show that 1/45 sub-pixel movement can be successfully detected in the 1D case through compressive holography and 12nm accuracy can be achieved. In the 2D case, it has been experimentally demonstrated that 1/30 sub-pixel motion was successfully detected. It is worthy to note that this number is not the theoretical limit and here we only tested up to 1/30 sub-pixel motion due to
the computer memory limitation in zero-padding of the 2D Fourier transform. This can be easily overcome in the future versions. The theoretical limit for axial localization is subject to the system’s depth of field. By applying compressive holography and template matching, we are able to detect 1/16 depth of field axial displacement.

The developed motion detection algorithm has been implemented onto vortex-induced vibration (VIV) and wave-induced vibration (WIV) characterization. The VIV and WIV of seal whisker models were successfully detected using the compressive holography model. The detected vibration was also shown to match with collaborator’s experimental results which were obtained from large-scale models.

Compressive holography has also been applied to detect extremely low signal-to-noise ratio (SNR) and signal-to-background (SBR) signals. In this thesis, we demonstrated the principle by studying its application to underwater platform crack detection from acoustic signals. The signal reflected from underwater platform is a mixture signal of platform background, possible small cracks and system noise. The desired signal is crack signal, but in the measurement, both SNR and SBR are extremely low. By applying null-space projection to remove strong background and sparsity-based signal recovery, the proposed method is shown to be effective to detect the weak crack signals.
Bibliography


