RICE UNIVERSITY

Exploiting compressive matrices for dynamic infrared object tracking

by

Jianbo Chen

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Science

APPROVED, THESIS COMMITTEE

Dr. Kevin F. Kelly, Chair
Associate Professor, Electrical and Computer Engineering

Dr. Ashok Veeraraghavan
Assistant Professor, Electrical and Computer Engineering

Dr. Paul E. Hand
Assistant Professor, Computational and Applied Mathematics

HOUSTON, TEXAS
April 2016
Copyright

Jianbo Chen

2016
ABSTRACT

Exploiting compressive matrices for dynamic infrared object tracking

by

Jianbo Chen

Recent development on compressive sensing (CS) presents a great potential for this technique to be used in broader applications from hyper-spectroscopy microscopy to homeland security. And the new mathematics of CS has drastically benefited this field especially in imaging and video applications. Based on novel theoretical principles and experiments, it has been demonstrated that an image can be reconstruct with only $K \ll N$ measurements from an $N$-dimensional basis, which is much less than the sampling rate required by the Shannon-Nyquist sampling theorem.

The compressive single pixel camera is one embodiment of such an imaging system and has proven capable of capturing both static images and dynamic scenes using fewer measurements than the current schemes. In this thesis we will explore compressive dynamic scene acquisition with prior information or models, incorporating with different sensing matrixes. We demonstrate through simulations and experiments the effectiveness of knowledge-enhanced patterns over unbiased compressive measurements in a variety of applications including motion tracking and object recognition.
We also present using a SPC like system for high-speed anomaly detection. Despite its importance in a wide variety of machine vision applications, extending anomaly detection and tracking beyond the visible spectrum in a cost-effective manner presents a significant technological challenge. As a step in this direction, we present a compressive imaging system, specially designed patterns, and a set of metrics to identify the existence of short duration anomalies against a complex background. Our novel measurement design is chosen to be most sensitive to singular anomalies based on the Walsh-Hadamard transform. We illustrate the utility of our approach via a series of simulations and experiments on the compressive single-pixel camera system.
Acknowledgments

First and foremost, I would like to give my greatest appreciation to my advisor Dr. Kevin Kelly for his guidance and mentoring of my graduate study and research, for his patience, motivation, enthusiasm, and immense knowledge. I have learned many things since I became Dr. Kelly’s student. Thanks my committee members: Prof Ashok Veeraraghavan, Prof Paul Hand for their suggestions on my research thesis and future projects.

My sincere gratitude also goes to my fellow group members: Dr. Yun Li, who helped developed the fundamental of HdTV² and anomaly detection. Liyang Lu, Yibo Xu and Jesse Choe for their valuable support on my graduate research.

I also want to express my thanks towards our collaborators Dr. Matthew Herman, Dr. Lenore Mcmackin, Mr. Tyler Weston and Dr. Thomas Goldstein for their insightful comments and discussion on all the projects.

My special thanks also go to my parents, Liangyu Chen, Jianrong Hou, for giving birth to me at the first place and supporting me spiritually throughout my life, and my wife Yang Zhang, who unconditionally supported every aspects of my life, you are the true hero.

To all my friends, thank you for making the life colorful and always be there when I needed help, it is my greatest fortune to have you all.
Contents

Acknowledgments ........................................................................................................ iv

Contents .......................................................................................................................... v

List of Figures ................................................................................................................. vii

List of Tables ................................................................................................................... xii

Nomenclature ................................................................................................................ xiii

1 Introduction .................................................................................................................. 1
  1.1. Video recording .......................................................................................................... 1
  1.2. The dilemma of dynamic scene acquisition ................................................................. 2
  1.3. Why compressive video matters .................................................................................. 2

2 Compressive sensing and compressive imaging ............................................................. 4
  2.1. Introduction ................................................................................................................ 4
  2.2. Sparse representation ................................................................................................ 8
  2.3. Single pixel camera ..................................................................................................... 11
  2.4. Compressive video ...................................................................................................... 13
  2.5. Experimental setup ................................................................................................... 15

3 Extending knowledge-enhanced compressive video acquisition ................................. 17
  3.1. Introduction ................................................................................................................ 17
  3.2. Hadamard Total Variation Video (HdTV^2) ................................................................. 20
  3.3. The STOne transform ................................................................................................ 23
  3.4. HdTV^2 with STOne transforms laboratory testing ....................................................... 25

4 High speed compressive anomaly detection ................................................................. 32
  4.1. Simulation scenarios ................................................................................................... 33
  4.2. Statistical inference .................................................................................................... 35
  4.3. Simulation with different detection matrices ............................................................... 37
    4.3.1. Anomaly detection with Walsh –Hadamard and Sum-To-One transform ...... 38
    4.3.2. Anomaly detection with Partial Complete transforms ........................................... 39
  4.4. Quantizing detection performance ............................................................................. 47
  4.5. Anomaly detection experiment with SPC ................................................................. 48
4.5.1. Laser triggered anomaly control system ............................................................ 51
4.5.2. Improve ROC and PR curve through realistic signal broadening ................... 53
4.5.3. Experiment results and discussion .................................................................... 57
4.6. Anomaly detection with FPA .................................................................................. 61
  4.6.1. Virtual channel ................................................................................................. 63
  4.6.2. Point spread function correction to the imaging system ................................. 65
  4.6.3. Discussion of anomaly detection results with FPA ......................................... 67
References ..................................................................................................................... 72
List of Figures

Figure 2.1 A side-by-side comparison of RAW image without compression and JPEG format image of the same image, the difference is subtle can be hard to detect. .......................................................... 5

Figure 2.2 (a) cameraman image from MATLAB image library, (b) sparse representation of image with discrete cosine transform. ................................................. 6

Figure 2.3 (a) Blue line is the sorted DCT coefficients of ‘cameraman’, red line is the compressed DCT coefficients that used to reconstruct the image, (b) original image, (c) “compressed” image with using only the largest half of the DCT coefficients ................................................................. 7

Figure 2.4 Algebraic schematic of the compressive sensing method. .................. 9

Figure 2.5 An illustration of single pixel camera. .............................................. 12

Figure 2.6 An inside look of InView CS camera. ............................................. 13

Figure 2.7 Experimental setup for compressive imaging and video applications. The system constits with DMD, a single detector and a 64 × 64 FPA. ......................................................................................... 16

Figure 3.1 Background subtraction experimental results using the SPC. Reconstruction of background images (top row) and test images (second row) from compressive measurements. Third row: conventional subtraction using the above images. Fourth row: reconstruction of difference image directly from compressive measurements. The columns correspond to measurement rates M/N of 50%, 5%, 2%, 1% and 0.5%, from left to right. Background subtraction from compressive measurements is feasible at lower measurement rates than standard background subtraction. ....................... 18

Figure 3.2 Example of (a) background and (b) background. (c) Hadamard pattern spectral comparison with (pattern number on the x-axis and frequency of occurrence on the y-axis) between background and foreground with (d) a part of spectrums zoomed in inset. .................................................. 21

Figure 3.3 Comparison of recovered frames: (a) ground truth of foreground from LHM dataset, (b) recoveries via 3DTV from randomly selected WHT coefficients, (c) recoveries via 3DTV from the proposed measurements......... 22
Figure 3.4 (a)-(c) The heatmap of combining first 20, 50, and 100 Walsh-Hadamard patterns associate with the coefficients have the biggest response to the moving vehicle. (d) The overlap of the heat map and the vehicle with the background, the high energy in the heatmap matches the location of the car.

Figure 3.5 A simple illustration of sum-to-one property of STOne transforms patterns.

Figure 3.6 A laboratory dynamic test scene with a toy car (white one in the middle) being pulled across the checkerboard background, there are also some static toy cars (dark colors) in the background.

Figure 3.7 The reconstructed image using HdTV$^2$ versus STOne transform, from left to right are direct STOne transform, in the middle is the HdTV$^2$ result reconstructed with 3DTV, and the fast low direct STOne resolution preview via is $\ell^2$ shown on the right.

Figure 3.8 An illustration of STOne pattern sequence of multiple resolution, the left side is $8 \times 8$ patterns, middle ones are $4 \times 4$ and the right column is $2 \times 2$, the lower resolution patterns are obtained through the property of STOne transform.

Figure 3.9 The reconstructed image using HdTV$^2$ with modified STOne transform, from left to right are direct STOne transform, in the middle is the result reconstructed with 3DTV, and the fast low direct STOne resolution preview via $\ell^2$ is shown on the right.

Figure 3.10 (a) Ground truth image of the car with the background. (b) The foreground object subtracted from the background. (c) Background reconstruction with 5% selected coefficients of foreground using WHT patterns and (d) STOne patterns. (e) Background reconstruction with 5% selected coefficients of foreground using WHT patterns and (f) STOne patterns. The STOne transformation has the capability to retrieve a better background image.

Figure 4.1 From left to right are three frames of scene images ($256 \times 256$) showing two toy cars moving towards each other. In (b), a bright pixel-block of size $4 \times 4$ in the background represents an anomaly point; (d) shows the zoom in on the anomaly.
Figure 4.2 An example of z-score, MMD score and change plots of detecting anomaly in the raw data, the change detection is the product of z-score and MMD score.

Figure 4.3 Anomaly detection results using the change plot for both PWHT and STOne, the spikes in the plots indicate the appearance and disappearance of the anomaly.

Figure 4.4 The partial-complete signature kernels.

Figure 4.5 The full PC Hadamard spectrum of the toy car image with $F = 64$ Blocks, some of the blocks are indicated.

Figure 4.6 Examples of patterns from Block 36 (left) and Block 54 (right), with the finer structures labeled in red.

Figure 4.7 Examples of spectrum when varying the location of the anomaly in the simulated scene with moving cars and swinging Newton’s cradle. The x-axis is the block number and y-axis is the coefficients. The anomaly locations are also shown in the top of each graph.

Figure 4.8 Examples of shifting nodal patterns convolved with the anomaly shifting in the scene at different locations. Locations that one signature Block fails to cover actually are fully covered by one of the rest Blocks.

Figure 4.9 The change detection using PWHT and PC block sets with different noise add to the simulation frames, images with different noise level are shown on the right.

Figure 4.10 Spectrum response of Thorlab PDA36A single detector.

Figure 4.11 Ground truth of the testing scene with (b) and without (a) the anomaly. (c) shows intensity profile across the line in (b) that pass through the anomaly.

Figure 4.12 2-channel Relay set used to precisely control the appearance of a laser anomaly in anomaly detection experiments.

Figure 4.13 A comparison between the laser trigger ground truth (red) and the measurement data (blue), the DMD is running random Walsh-Hadamard pattern at 1 kHz.
Figure 4.14 Measurements of two laser trigger signals with two experiments that have the same parameters.................................................................................................................................53

Figure 4.15 ROC and PR curve before the improvement. ...........................................54

Figure 4.16 The comparison between the real data and the ground truth. The blue line is the change detection result, and red line is the anomaly indication. ........................................................................................................................................55

Figure 4.17 The broadened ground truth that matches the detection data......56

Figure 4.18 ROC and PR curve after the improvement, the precision rate raises to above 50%......................................................................................................................................................56

Figure 4.19 An example image of the scene, the anomaly is zoomed in and the toy car is moving across the scene during the experiment...............................57

Figure 4.20 Raw measurement data of anomaly location #5, the anomaly appeared three times (red), which can also be seen from the intensity increase in the measurement (blue)........................................................................................................................................58

Figure 4.21 The change detection result of two different anomaly locations (location 5 and 8). Each detection has the anomaly turned on and off three times (red and shaded area). The change detection is combination of z-score and MMD. ..............................................................................................................................................59

Figure 4.22 ROC curve and associated PR curve with laser trigger control....60

Figure 4.23 ROC and PR curve plots for PC and PWHT patterns under different SNR circumstances, the SNR is labeled at the top of each graph. ................61

Figure 4.24 Image captured by the Hamamatsu 64 × 64 indium gallium arsenide (InGaAs) FPA. There exist defect pixels at location (60, 26) and (60, 27) of the FPA, which always give high intensity readouts........................................62

Figure 4.25 The detection signal from using different methods, SPC versus FPA. The FPA virtual channel has a significant improvement on SNR as each detector collects signal from a smaller ROI and the anomaly will have a bigger impact on the measurement. ........................................................................................................64

Figure 4.26 ROC and PR curve for SPC and FPA anomaly detection results. The FPA method elevates the performance under these metrics, especially for the PR curve........................................................................................................................................65
Figure 4.27 PSFs of multiple DMD pixels imaged on the FPA, the measured PSF (right) shows the distortion and the skew of the imperfect imaging system, the ground truth of pixel displayed in the DMD is shown on the left. 66

Figure 4.28 FPA captured image with and without the correction of the PSF. The left image is reconstructed small image patches stitched together. The right image is recovered with the consideration of the PSF. 67

Figure 4.29 Maximum $F_\beta$-score calculated from the PR curve results, each corresponds to an anomaly intensity (y axis) and duration (x axis). Three different $\beta$ numbers are chosen. 68

Figure 4.30 Detection threshold that determined by the maximum $F_\beta$ score associates with the corresponding condition, anomaly intensity in y axis and duration in x axis. 69

Figure 4.31 Arbitrary test of the anomaly detection using the pre-determined threshold, each row represents the anomaly intensity, which decreases from top to bottom. Each column represents the anomaly duration, which shortens from left to right. 70
List of Tables

Table 2.1 – The acquisition and reconstruction of image with different resolution, the time is tested using the InView single pixel camera....................... 14
### Nomenclature

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-N</td>
<td>Shannon-Nyquist</td>
</tr>
<tr>
<td>CS</td>
<td>Compressive Sensing</td>
</tr>
<tr>
<td>SPC</td>
<td>Single Pixel Camera</td>
</tr>
<tr>
<td>FPA</td>
<td>Focal Plane Array</td>
</tr>
<tr>
<td>DMD</td>
<td>Digital Micro-mirror Device</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge Coupled Device</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal–Oxide–Semiconductor</td>
</tr>
<tr>
<td>RIP</td>
<td>Restricted Isometry Prosperity</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete Cosine Transformation</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transformation</td>
</tr>
<tr>
<td>MMD</td>
<td>Maximum Mean Discrepancy</td>
</tr>
<tr>
<td>TV</td>
<td>Total Variation</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spread Function</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>PR</td>
<td>Precision Recall</td>
</tr>
<tr>
<td>FOV</td>
<td>Field Of View</td>
</tr>
<tr>
<td>ROI</td>
<td>Region Of Interest</td>
</tr>
</tbody>
</table>
1 Introduction

1.1. Video recording

Many have seen video clips of a helicopter circling the sky without the blade moving, or spotted a car on TV moving forward with the wheels rotating backwards. These magical scenes, also known as the wagon-wheel effect[1], reveal a profound issue in video capture and processing realm known as aliasing. In these cases, the camera sensor discretely samples continuously and rapidly varying scenes at some frame rate, and temporal aliasing can occur from a limited frame rate. Aliasing will change the apparent frequency of motion to a slower or even negative frequency.

The Shannon-Nyquist (S-N) theorem[2] addresses the signal aliasing problem, requiring the sampling rate to be twice of the highest signal frequency of interest. Imaging hardware has been built to operate faster and faster to record
events with shorter duration[3], the fastest camera is even able to visualize the speed of light[4].

1.2. The dilemma of dynamic scene acquisition

Meanwhile, the development in high-speed cameras becomes prohibitively expensive for video camera that tries to capture information outside the visible spectrum. In addition, the increase on frame number and image size simultaneously requires more storage, which brings huge demand for new storage infrastructure and processing power. Last but not least, one may want to analyze and extract information from the video in real time, thus the battle between more bandwidth and better video quality associates with the video acquisition and compression all the time.

1.3. Why compressive video matters

The compressive sensing (CS) theory brings light on providing solutions to the previously mentioned problems. CS attempts to overcome the limitation of S-N theory with much less signal acquisition, and reconstruct the original signal from a sampling rate much smaller than the S-N rate. Besides the advantage of reconstructing the image with less data, the CS also mitigates the requirement of high-resolution focal plane array (FPA), which makes the imaging beyond the visible spectrum feasible at low cost. With a spatial light modulator and few sensors or even a single one in extreme case, a high quality image can be robustly
reconstructed using the compressive imaging techniques. Compressive video extends the CS theory and has the potential to outperform the conventional video cameras since the DMD and a low resolution FPA are capable of running at a speed as high as tens of kilohertz. Another benefit of compressive video is that fewer measurements are required to reconstruct the frame, which means smaller data size and less transmission bandwidth is required.

However, despite its importance, only recently has significant headway been made in compressive video acquisition and recovery. Compressive video is complicate by the ephemeral nature of dynamic events, which makes direct extensions of standard CS imaging architectures and signal models difficult. Some possible solutions attempt to solve the compressive video problem including CS-MUVI[5] and 3DTV[6]. In this thesis I am proposing two applications that use compressive video algorithms and the single pixel camera system for dynamic scene acquisition and detection. The rest of my thesis is organized as follows: Chapter 2 gives a general introduction to compressive imaging and video, Chapter 3 discusses extending knowledge enhanced compressive video with background subtraction method, and high-speed anomaly detection directly from compressive measurements will be presented in Chapter 4.
Chapter 2

2 Compressive sensing and compressive imaging

2.1. Introduction

Compressive sensing (CS) is a signal processing technique for acquiring and reconstructing signals at less than Shannon-Nyquist (S-N) sampling rate, by solving an ill-posed linear equation using regularization and optimization methods[7]. Since the first introduction of CS a decade ago by Candes et al.[8]–[10], it has drawn great attention and development. When it comes to its imaging applications, one normally needs a large number of photon collecting sensors and storage to reconstruct a high-resolution image. Take a commercial digital camera for example, a traditional camera contains a focal plane array (CCD or CMOS) of millions photo-detectors, and an uncompressed raw 42 megapixel image is 64 megabytes[11]. However, to ease the burden of storage and display, most often the raw image is compressed and its
final size is no more than few megabytes. The uncompressed and compressed images have tiny differences that are hard to distinguish by eye (Figure 2.1).

Figure 2.1 A side-by-side comparison of RAW image without compression and JPEG format image of the same image, the difference is subtle can be hard to detect.

This is due to the fact that most nature scene images have sparse representations under certain transform such as discrete wavelet transform (DWT) [12] or discrete cosine transforms (DCT) [13]. Figure 2.2 shows an image and its DCT transform coefficient, the sparsity is obvious since most of the energy coefficients are close to zero.
In Figure 2.3 (a), blue line shows the sorted DCT coefficients of the example image, we find most of the coefficient energies concentrate in a small portion of the whole spectrum and the rest of the energies are small. Using a compression scheme similar to that of the JPEG format[14], we reserve the largest portion of the coefficient and set the small coefficient to zero (red line). Using the “compressed” data and inverse transform back to the image domain, the differences between the compressed and original images are subtle and difficult to notice (Figure 2.3(b), (c)).
Figure 2.3 (a) Blue line is the sorted DCT coefficients of ‘cameraman’, red line is the compressed DCT coefficients that used to reconstruct the image, (b) original image, (c) “compressed” image with using only the largest half of the DCT coefficients

As the conventional imaging method requires large amount of detectors to get a high-resolution image, it will be impractical for applications in non-visible wavelengths such as infrared and beyond. For example, the silicon detectors have little or none response to the infrared light and indium gallium arsenide (InGaAs) is the most common materials that used to manufacturing the photon counting devices in such spectrum region. However InGaAs materials are rare and poses an inherent lattice mismatch with conventional silicon chips[15], therefore the manufacturing cost for large density detectors arrays is too high for such wavelength regions. The data storage cost may also be expensive depending on size of image or video.
Instead of capturing the redundant data and compressing it, the idea of compressive sensing is to only measure a portion of coefficients under the transformed basis, by assuming most coefficients are near zero in this basis we are able to solve an ill-posed linear equation using regularization to reconstruct the original image. The compressive approach lessens the burden of acquisition, storage and data transmission. The biggest advantage of CS imaging for non-visible wavelengths or higher dimensional data is that the acquisition of data can be accomplished using limited amount of detectors, a single pixel in the extreme case, thus allowing the possibility of taking higher resolution images with exotic wavelengths at minimum cost.

2.2. Sparse representation

Consider a vectorized image denoted as \( x \in \mathbb{R}^N \), it is modulated by some random basis \( \Phi \in \mathbb{R}^{M \times N} \), producing an output measurement \( b = \Phi x \) as an \( M \times 1 \) vector. In addition the image \( x \) can be represented in a sparse form \( S \) under particular transform basis \( \Psi \) (Figure 2.4), therefore the equation can be re-written as \( b = AS \), where \( A \) is the product of the random basis \( \Phi \) and the sparsity basis \( \Psi \). \( S \) is a sparse vector that has only \( K \) numbers of non-zero values, and \( K < M << N \).
Figure 2.4 Algebraic schematic of the compressive sensing method.

Two conditions have to be satisfied for the sparse signal $S$ to be robustly reconstructed: the incoherence of the random and sparse bases and restricted isometry property (RIP)\cite{16} of the sensing matrix. If $\Phi$ is incoherent with $\Psi$, the signal $b$ can be recovered from $M = O(K \log N)$\cite{7} measurements. A measure of mutual incoherence $\mu$ of the two bases is given by

$$\mu(\Phi, \Psi) = \max |\langle \Phi_m, \Psi_n \rangle|$$

Eq (2.1)

The greater the mutual incoherence, the smaller the number of measurements needed. In particular, this incoherence holds with high probability between an arbitrarily fixed basis (wavelet, curvelet, Fourier, etc.) and a randomly generated one, such as independent identically distributed Gaussian ($\Psi_{ij} \in N (0,1)$) or Bernoulli ($\Psi_{ij} \in \pm 1$ with equal probability). Meanwhile, the incoherence also ensures $A = \Phi \Psi$ is a random matrix. The RIP is also well preserved for these random matrices.
To find the sparsest solution, it is a natural to solve the problem with an $\ell_0$ minimization method

\[
s^* = \min_x \{\|s\|_0: As = b\} \tag{2.2}
\]

Where the $\ell_0$ norm $\|x\|_0$ counts the number of non-zero elements in $x$. However this method is proven to be NP-hard and computationally impractical to solve. An alternative method is using $\ell_1$ norm minimization or Lasso that aims to look for sparse solution and find non-zero elements in $S$. Consider the presence of noise, the equation can be further generalized as

\[
\min_x \{\|s\|_1: \|As - b\|_2 \leq \sigma\} \tag{2.3}
\]

Beside $\ell_1$ norm, another popular method that has been used to solve this inverse problem is total variation (TV) regulation[17]. TV can be regarded as a generalized $\ell_1$ problem in CS. Instead of assuming the signal $x$ is sparse, TV model proposes the gradient of the underlying signal is sparse and the TV minimization problem is trying to seek the unique solution with the sparsest gradient, which is expressed as

\[
\min_x \sum_i \|D_i x\| \text{ s.t } \Phi x = y \tag{2.4}
\]

where $\sum_i \|D_i x\|$ is the sum of the discrete gradient magnitude at each point. The TV method is proven to be a robust method for sparse signal recovery[18] and in our project the TVAL3 solver based on TV is used that was developed by Chengbo and Ying[19].
2.3. Single pixel camera

Based on the previous discussion on sparse signal recovery, it is not difficult to see that we can reconstruct a two-dimension image without using a focal plane array (FPA), instead a limited amount of detectors and an optical modulator could be used to collect the signal and reconstruct the image from the measured data. The single pixel camera (SPC) was introduced in 2006 and is a perfect demonstration that the concept of CS is feasible for many real-world imaging applications[20].

The setup of the SPC is illustrated in Figure 2.5, in which a digital micro-mirror device (DMD) is used to modulate the light. Each DMD consists of hundreds of thousands of micro-mirrors whose orientation can be precisely controlled to tilt for example either +12 degree or -12 degree about its diagonal corresponding to the ON and OFF states. And the ON and OFF states directly correspond to the elements of a binary matrix, as shown in the bottom left of Figure 2.5. The binary pattern of the DMD along with the image projected on it acts as the inner product of a column of random binary matrix A and the vectorized image x. The total number of patterns consists as matrix A from ΦΨ. The modulated light is then collected after the DMD and focused on a single pixel, filling one element in vector b. The DMD can flip at speed in the tens of kHz, which allows fast signal acquisition.
Based on this idea, the SPC has also been commercialized by InView technology, a company that manufactures the InView camera. The InView camera has a more compact design that consists of a total internal reflection (TIR) prism and a $1024 \times 768$ resolution DMD as light modulator, the integrated analog to digital circuit and FPGA allowing small volume factor and programmability (Figure 2.6). The InView camera also comes with different versions depending on what detector is used. The visible version features a silicon detector while the near infrared (NIR) one has an InGaAs detector. Furthermore the Inview camera is equipped with a PCI-E interface that can be used to communicate with a computer, so any arbitrary patterns can be generated on the fly to perform the measurements, without being prestored in the memory.
2.4. Compressive video

A big challenge for CS application is video acquisition, because it increases measurement time to decrease sensor resolution. Table 2.1 shows a relationship between the image resolution and reconstruction time with the InView camera, from which we can see that it takes as many as 50 seconds to acquire and reconstruct a high resolution image. This will be problematic for motion image capturing.
Table 2.1 – The acquisition and reconstruction of image with different resolution, the time is tested using the InView single pixel camera

<table>
<thead>
<tr>
<th>Resolution</th>
<th>1024×768</th>
<th>512×384</th>
<th>256×192</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>57 s</td>
<td>14 s</td>
<td>5 s</td>
</tr>
<tr>
<td>20%</td>
<td>25 s</td>
<td>8 s</td>
<td>3 s</td>
</tr>
<tr>
<td>10%</td>
<td>18 s</td>
<td>6 s</td>
<td>1.5 s</td>
</tr>
</tbody>
</table>

One recent approach is to solve the compressive video problem by using dual-resolution patterns obtained through the Kronecker production of a low-resolution Hadamard pattern with a high-resolution random Hadamard pattern. This hybridization enables reconstruction of a low resolution motion image with only a small amount of measurements. The optical flow technique is also used to predict the pixel varying directions through the video and use that prior information to reconstruct the high definition image frames[5]. Another solution is to incorporate the temporal sparsity into the TV minimization algorithm (3DTV) with the assumption that a big portion of pixels in video frames merely change between the adjacent frames[6]. The 3DTV has the merits of fast reconstruction speed and flexibility to work with multiple type of sensing matrices.
Beyond dual resolution, multi-scale resolution preview can be achieved by use of the Sum-To-One (STOne) patterns[6]. Even with these accomplishments it is not clear full-scene compressive video recovery will ever be accomplished in real-time. However, it is also not clear it will ever need to be for many machine-vision applications.

2.5. Experimental setup

The majority the compressive imaging and video experiments were conducted in the lab with a system shown in Figure 2.7. The scene with moving objects are projected to the DMD through an imaging lens, and the modulated measurements are collected by different types of sensors regarding to particular applications. A TIR prism and a rotating mirror are used to channel the light signal to specific photon counting devices. This system allows a convenient switch between multiple types of detectors and has the flexibility of extending sensing and detection applications.
Figure 2.7 Experimental setup for compressive imaging and video applications. The system consists with DMD, a single detector and a 64 × 64 FPA.
3 Extending knowledge-enhanced compressive video acquisition

3.1. Introduction

In this chapter we are considering a scenario to perform compressive video application with foreground object tracking. For some applications such as surveillance and object detection, people only care about a moving object rather than the entire scene, and the background is neglected. Image background subtraction is utilized in traditional object tracking, which is accomplished by subtracting successive image frames with background, and identifying any differences that may be labeled as targets of interest[21]. High-resolution video acquisition requires high-density detector arrays, which can be prohibitively expensive and of limited performance for applications beyond the visible spectrum. Compressive sensing for background subtraction (CSBS) based on the single pixel
camera (SPC) system was proposed to extract the foreground object from pseudo-random compressive measurements[22]. CSBS shows that higher compression ratio can be achieved to robustly reconstruct the object of interest since the foreground only image is spatially sparser (Figure 3.1).

Figure 3.1 Background subtraction experimental results using the SPC. Reconstruction of background images (top row) and test images (second row) from compressive measurements. Third row: conventional subtraction using the above images. Fourth row: reconstruction of difference image directly from compressive measurements. The columns correspond to measurement rates M/N of 50%, 5%, 2%, 1% and 0.5%, from left to right. Background subtraction from compressive measurements is feasible at lower measurement rates than standard background subtraction.
Hadamard Total Variation Video (HdTV²) was proposed to extend the CSBS concept in the CS-MUVI framework to achieve exclusive foreground object video acquisition. It utilizes a training process to acquire the prior information of the background scene and the object of interest, and then use the three-dimensional total variation (3DTV) to reconstruct the video frames. A detailed process of performing HdTV² is given in the following section and we proved it has a good performance in foreground isolation with a small amount of measurements.

The 3DTV algorithm was developed to reconstruct the CS video by considering the time variant in the $\ell_1$ optimization process, as the sparsity is observed not only in image space but also in temporal basis. Denote $\mu_{i,j}^f$ as the $(i,j)$th pixel $\mu$ in frame $f$, gives a generalized mathematic form of the 3DTV operator, where the first two terms computes the gradient of pixels within each frame and the third term computes the individual pixel difference between adjacent frames.

$$|\nabla_3 \mu| = \Sigma_{i,j,f} \sqrt{(\mu_{i+1,j}^f - \mu_{i,j}^f)^2 + (\mu_{i,j+1}^f - \mu_{i,j}^f)^2 + |\mu_{i,j}^{f+1} - \mu_{i,j}^f|}$$  \text{Eq (3.1)}$$

The computational cost is much smaller since it doesn’t require reconstruction of the optical flow map to provide the pixel movement information, which is a computationally expensive process[5].
3.2. Hadamard Total Variation Video (HdTV²)

In conventional SPC compressive imaging, the scene is modulated with permuted Walsh-Hadamard patterns. The Hadamard matrices are the perfect choices for modulation pattern when the ON and OFF status of DMD mirror can be considered as +1 and -1 entries in Hadamard matrices. Random permutation is used because they mathematically represent a sparse-signal sensing matrix fulfilling RIP conditions and have the capability of transforming a canonical scene to a sparse representation. While using randomly permuted Walsh-Hadamard modulation vectors can reconstruct scenes with high accuracy by ‘democratically’ sampling over all frequencies and using no prior information, for instances where there is some prior knowledge it lacks the capability to automatically discern foreground objects in the scene from background.

Taking a closer look at the Walsh-Hadamard spectrum of a typical road scene, where a car moving across the highway and the vehicle is the moving target that we isolated from the static background as shown on the left side of Figure 3.2. The graph on the right side of Figure 3.2 maps the integrated optical energy in each of the modulated scene measurements in the complete set, we see that for specific WHT patterns, foreground and background image components have significantly different energies at some locations[23].
Figure 3.2 Example of (a) background and (b) background. (c) Hadamard pattern spectral comparison with (pattern number on the x-axis and frequency of occurrence on the y-axis) between background and foreground with (d) a part of spectrums zoomed in inset.

A higher system response to the target of interest, compared to using randomly ordered patterns as the sensing matrix, is realized by training the system with prior information about the scene to select only the patterns with a large foreground versus background energy difference. To be specific, we first acquire some training data through the CS measurement, and then compare the coefficients between images that contain background with foreground and images with background only. The training process uses the background subtraction method to pick up the coefficients that are larger for the foreground while trying to avoid the coefficients that are big for the background, as shown in Figure 3.2. The selected coefficients usually comprise 5-20% of the total measurement set. After training, we use the
patterns that correspond only to the selected training coefficients to perform the CS measurement and reconstruct the foreground and background frames.

Figure 3.3 shows the target isolation simulation result using HdTV² versus conventional randomly ordered Walsh-Hadamard transform (WHT) matrix. The compression ratio is 10%. Compared with the random measurements, HdTV² displays better performance with 3DTV reconstruction.

![Figure 3.3 Comparison of recovered frames: (a) ground truth of foreground from LHM dataset, (b) recoveries via 3DTV from randomly selected WHT coefficients, (c) recoveries via 3DTV from the proposed measurements.](image)

Another benefit of the HdTV² strategy is the patterns associate with the biggest response to the foreground provides location information of the target. The selected patterns have higher probability on measuring where the target of interest locates. As shown in Figure 3.4, with a few number of measurements, the combination of the patterns reveals the position of the foreground object. The HdTV² gives the possibility of object location detection in real-time without
reconstructing the image, and through learning the prior information of the object location and size, the HdTV² can have a superior advantage in extracting foreground information.

![Figure 3.4](image)

**Figure 3.4** (a)-(c) The heatmap of combining first 20, 50, and 100 Walsh-Hadamard patterns associate with the coefficients have the biggest response to the moving vehicle. (d) The overlap of the heat map and the vehicle with the background, the high energy in the heatmap matches the location of the car.

### 3.3. The STOne transform

Besides Walsh-Hadamard matrix, another Hadamard-based transform known as Sum-To-One (STOne) transform[6] has also proposed as the sensing matrix in the SPC. The STOne pattern is a multi-resolution Hadamard pattern, designed to reconstruct any intermediate resolution image between a single value and the highest resolution image within one measurement via $\ell_2$ reconstruction method.
Consider a matrix $S_4$ as shown in Eq (3.2), where the row vectors are orthogonal to others and elements in the vector sum to 1.

\[
S_4 = \frac{1}{2} \begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\quad \text{Eq (3.2)}
\]

Eq (3.3) shows using $S_4$ as a stencil to construct a new set of transform matrices, each column of $S_4^k$ represents a square sensing matrix of resolution $2^k \times 2^k$, higher resolution patterns can be obtained through the iterative Kronecker product. The sum-to-one property is satisfied for $S_4^k$, $k \in (1, \cdots n)$.

\[
S_{4^{k+1}} = S_4 \otimes S_4^k = \frac{1}{2} \begin{pmatrix}
-S_4^k & S_4^k & S_4^k & S_4^k \\
S_4^k & -S_4^k & S_4^k & S_4^k \\
S_4^k & S_4^k & -S_4^k & S_4^k \\
S_4^k & S_4^k & S_4^k & -S_4^k
\end{pmatrix}
\quad \text{Eq (3.3)}
\]

Figure 3.5 gives an intuitive illustration of the multi-scale property of STOne, where each $2 \times 2$ subset of the high-resolution STOne pattern adds to either +1 or -1, resulting in a lower resolution of STOne pattern. The same approach applies for the next power of two lower resolutions.
Figure 3.5 A simple illustration of sum-to-one property of STOne transforms patterns.

STOne allows measurements be reconstructed to image instantly at Nyquist rates at any power-of-two resolution, so that low resolution preview can be generated with inverse sensing matrix multiplication. The same data can also be upscaled to higher resolutions using compressive signal reconstruction methods that leverage sparsity to overcome the Nyquist limit.

3.4. **HdTV² with STOne transforms laboratory testing**

The fast preview generation and multi-scale properties make STOne an attractive alternative as compressive video sensing matrix. To perform a preliminary simulation test on a dynamic scene captured from the SPC system as shown in Figure 3.6, we simply applied the HdTV² techniques with STOne transform by selecting the patterns corresponding to the coefficients which stands out for the foreground.
Figure 3.6 A laboratory dynamic test scene with a toy car (white one in the middle) being pulled across the checkerboard background, there are also some static toy cars (dark colors) in the background.

The results did not compare favorably with the ones from HdTV² method. As shown in Figure 3.7, the noise level for the reconstruction of moving car frames was large.
Figure 3.7 The reconstructed image using HdTV$^2$ versus STOne transform, from left to right are direct STOne transform, in the middle is the HdTV$^2$ result reconstructed with 3DTV, and the fast low direct STOne resolution preview via is $\ell_2$ shown on the right.

The possible reason that leads to the poor reconstruction was the pattern matrix selection based on target energy broke the inherent multi-scale structure of STOne transforms across the multiple scales of the STOne transform, which caused redundant sampling in the lower resolution pattern selections resulting in lower quality reconstruction.

To better understand the multi-resolution STOne schema, Figure 3.8 illustrates the three different resolution levels of STOne pattern, where a set of 20 STOne patterns with the resolution of $8 \times 8$, denoted as H, is shown at the left. Within each $8 \times 8$ STOne pattern, sub-blocks of $2 \times 2$ pixel groups can be summed up to either $+1$ or $-1$, resulting in a medium resolution STOne pattern of $4 \times 4$ (M) in the middle column. The binning process can continue to get lowest resolution of $2 \times 2$ (L) at right. Thus, as shown by the arrows in Figure 3.8, using the STOne patterns at
the lowest resolution, we obtain a complete data set to reconstruct a $2 \times 2$ image in only 4 measurements, and in 16 measurements for a $4 \times 4$ image.
Figure 3.8 An illustration of STOne pattern sequence of multiple resolution, the left side is $8 \times 8$ patterns, middle ones are $4 \times 4$ and the right column is $2 \times 2$, the lower resolution patterns are obtained through the property of STOne transform.

Note that this is the conventional way to generate the STOne pattern, while there is no necessity for it to be sequenced. For example, choosing pattern set L1 L2 L3 L4 and L1 L6 L7 L16 has the same effect on reconstruct a $2 \times 2$ image.

The previous idea of applying the HdTV\(^2\) directly with STOne transform was to try to look for the patterns in $[Hn]$ that have a big response to foreground without considering the lower level sets of $[Mn]$ or $[Ln]$. This caused redundancy in the lower level STOne patterns. For example choosing M1, M3, M7 and M11 reconstructs a bad $2 \times 2$ preview because L3, L7 and L11 have the same signature that fails to provide complete low-resolution information. Thus to properly apply HdTV\(^2\) with STOne pattern, we have to take its unique nested property into consideration when choosing the training patterns. The new simulation results HdTV\(^2\) with modified STOne strategy is shown in Figure 3.9, which yielded great improvements over the naïve implementation.
Figure 3.9 The reconstructed image using HdTV² with modified STOne transform, from left to right are direct STOne transform, in the middle is the result reconstructed with 3DTV, and the fast low direct STOne resolution preview via $\ell_2$ is shown on the right.

Another advantage of this new model-based STOne object tracking approach is that the randomness and finer structure of the patterns in the measurement matrices should still allow for fairly reasonable reconstruction of the background as well as the foreground, which does not occur in the HdTV² method. As shown in Figure 3.10, the background reconstruction with STOne is better compare to WHT strategy under different compressive ratio. A SNR of 27dB vs. 8 dB are obtained for STOne and WHT respectively at 5% compression ratio, and 30 dB vs. 17 dB at 10% compression ratio. Given that the target motion is likely occurring on a moving background, e.g. a panning camera, the STOne based tracking should prove more useful than HdTV² which focuses on the target only and performs best with a static background.
Figure 3.10 (a) Ground truth image of the car with the background. (b) The foreground object subtracted from the background. (c) Background reconstruction with 5% selected coefficients of foreground using WHT patterns and (d) STOne patterns. (e) Background reconstruction with 5% selected coefficients of foreground using WHT patterns and (f) STOne patterns. The STOne transformation has the capability to retrieve a better background image.
Chapter 4

4 High speed compressive anomaly detection

Anomaly detection finds importance in a wide variety of applications. It is the identification of items, events, or observations that do not conform to an expected pattern or to other items in a dataset. Typically the anomalous signal will translate to some kind of problem such as bank fraud, a structural defect, medical problems, or errors in a text[24]–[27].

The scenario we consider falls into the category of imaging, specifically, a situation with one or multiple anomalies in a scene with moving objects, for example detecting a bright flash in the scene or finding fluorescence molecular (fluorophores et.al) in microscope image. A conventional method of doing such anomaly detection is to capture the scene using a focal plane array (FPA) as a sequence of video frames; the frames are then processed on computer and the anomaly can be detected by subtracting each one with static background frames
that do not contain an anomaly[28]. However this method is resource demanding, which means it will be extremely impractical to achieve anomaly detection in wavelengths other than visible light for the reason we discussed in Chapter 2. Another problem is that transferring the video frame data requires a large bandwidth while most of the information will be redundant for the purpose of anomaly detection[24].

Our approach to detect anomalies will address the shortcomings of the conventional methods along with some additional benefits. It is based on our single pixel camera (SPC) system using a single photon detector, yet the detection portion can be expanded with different FPAs for specific detection wavelength regions. We tested the anomaly detection system with both single detectors and a short wave infrared (SWIR) FPA consisting of 64×64 detectors. The single detector has the merits of fast operation speed and low costs, and it can be conveniently switched with detectors working under other wavelength regions such as mid-wave infrared where an FPA will be prohibitively expensive for most applications. On the other hand, the FPA has the advantage of having better detecting SNR, as each detector ‘sees’ a smaller region of interest (ROI) and thus more sensitive to changes in the scene[29], [30].

4.1. Simulation scenarios

We begin with a series of simulation of anomaly detection on a video set. The video frames that consist two toy cars moving towards each other, and there are
color checkerboard and metal Newton's cradle in the background. The simulation also tried to match the real-world SPC setup: the scene is modulated with detection matrix (binary DMD patterns in the experiment), and then summed to a single value as the simulated system is SPC, or the scene is divided into patches and each patch is summed up and 'detected' by a single detector if the simulated detection hardware is a FPA. Examples of testing scene are shown in Figure 4.1.

![Figure 4.1](image)

**Figure 4.1** From left to right are three frames of scene images (256 × 256) showing two toy cars moving towards each other. In (b), a bright pixel-block of size 4 × 4 in the background represents an anomaly point; (d) shows the zoom in on the anomaly.

The simulation parameters used are also matched to the experiment, assuming that the DMD is operating at 15-kilohertz speed and there are 100 frames in total. Assuming the video length is 10 seconds, the frame rate is 30, so for each frame the scene is modulated 500 times before the next frame. Thus for a 128×128 resolution scene, the compression ratio will be 3%.
4.2. Statistical inference

In the simulation, we repeatedly modulated the scene with patterns designed for anomaly detection and form the data metric by summing up the post-modulation image. When an anomaly is present, the measured modulations show different statistical information compared with the case without the anomaly, which we relied on to make the prediction. Given the acquired time series \( \{b_j\} \) from the selection of patterns, the metrics is calculated through the following steps:

1. We calculate the change, \( x_j = \sum_{k-1}^{M} |b_{j+M+k} - b_{j+k}| \), where \( M \) is the number of patterns in this selection. This purely records the difference of measurements between the current set of patterns and the previous set of patterns.

2. Then we use a window of size \( W \) to slide through the change series with one step at a time. Based on the hypothesis that the data collected from the sensor has a normal distribution with a mean and standard deviation, and the anomaly data has a huge deviation from the data mean value. So for each window, we calculate the standard score \( (z\text{-score})[31] \) from the two nearest non-overlapping window. The ratio of mean value of the difference between those two windows and the smaller variance of the changes within two windows forms the \( z \)-value (Eq 4.1).

\[
z = \frac{x - \mu}{\sigma}
\]  
Eq (4.1)
3. Based on the changes, we calculate the maximum mean discrepancy (MMD)[32], [33], which is used to identify whether data are from the same or different probability distributions. There are two advantages of using MMD to identify the anomaly in this particular problem: first, MMD can be easily estimated based on samples, and hence leads to low complexity tests. Second, MMD-based approaches do not require estimation of probability density functions as an intermediate step, but directly estimate the distance of distributions to build tests.

\[
MMD_b[\mathcal{F}, X, Y] = \\
\left[ \frac{1}{m^2} \sum_{i,j=1}^m k(x_i, x_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^n k(y_i, y_j) \right]^{\frac{1}{2}}
\]

Eq (4.2)

4. The product of Z-score and MMD score is the indicator of the anomaly’s presence.

Figure 4.2 shows detection examples of using z-score, MMD score and change plot as anomaly indication metrics, the change detection is the product of z-score and MMD score. An observation from the calculated change detection results is that there is always a secondary negative spike that is associated with the signal spike. The reason that causes this phenomenon is due to the sliding window calculation process. As we are comparing a section of data (with length W) compare to the same length of data that k steps before it. The negative spike has a distance W apart from the positive true signal spike. This can be used as an extra constraint for us to determine whether the signal is a true positive or not by looking for the secondary
spikes. In the rest part of the chapter, the absolute value of the change plot will be used as the only criteria for our anomaly detection performance evaluation.

![Raw Data, Change, Z-Score, MMD Score](image)

**Figure 4.2** An example of z-score, MMD score and change plots of detecting anomaly in the raw data, the change detection is the product of z-score and MMD score.

### 4.3. Simulation with different detection matrices

There are multiples types of patterns can be put onto the DMD acting as the anomaly detection matrix, in our case, three types of detection matrixes have been studied and compared in detection capabilities. Permut Walsh-Hadamard transform (PWHT)[20], Sum-To-One transform (STOne) and Partial-Complete transform (PC)[34] are the main detection transform we are considering for the purpose. Each transform has its own properties that fit particular applications.
4.3.1. Anomaly detection with Walsh–Hadamard and Sum-To-One transform

The PWHT is a randomly permuted column of a complete Hadamard matrix, the sequence of the row is also randomly chosen so that there will be minimum coherence between modulation patterns. Meanwhile the STOne has a unique property of getting low resolution preview instantly with the measurement which is becomes a benefit in compressive image and video reconstruction. However in general both permuted PWHT and STOne patterns are random Hadamard matrices that ‘democratically’ sensing the whole scene and having no preference on the prior information of the anomaly.

Figure 4.3 shows the simulated anomaly detection results using both PWHT and STOne patterns on the DMD with the statistic plots. The spikes in the change plot indicate the detected event of anomaly on or off.
Figure 4.3 Anomaly detection results using the change plot for both PWHT and STOne, the spikes in the plots indicate the appearance and disappearance of the anomaly

The magnitude of the spikes can be also be used to indicate the effectiveness of the detection respect to the anomaly intensity to background illumination ratio, as the intensity ratio increases, the spikes' intensities become larger. When the anomaly becomes dimmer, at some level of intensity it will not be able to be detected by the algorithm and the number of false positive spikes will be too many to distinct the true anomaly from the false alarm. The detection capability between PWHT and STOne are similar as the randomness and sometimes has limited performance when the signal to background illumination ratio (SBR) is low.

4.3.2. Anomaly detection with Partial Complete transforms

Partial Complete (PC) transform is a sensing matrix that is a variation of Hadamard transform and was proposed by Matthew Herman[34]. The PC strategy takes advantage of the fact that the high resolution Walsh Hadamard matrix is the Kronecker product of two smaller Walsh matrixes, as described in the following equations:

\[ H_{m+n} = H_m \otimes H_n \]

When \( m \gg n \), \( H_n \) would perform as a seed pattern that determines the finer structure of the grand pattern. In another words, fixing the seed pattern and changing \( H_m \), the patterns would have identical smallest elements. Take a DMD pattern of \( 256 \times 256 \).
for example, we specifically set \( n = 64 \) and \( m = 1024 \). Thus we will have 64 signatures and each signature can generate a Block of 1024 patterns. Figure 4.4 shows the partial-complete signature kernels used in partial complete strategy to generate high-resolution patterns. Those kernels are 2-dimensional visualizations of all signatures.

Figure 4.4 The partial-complete signature kernels.
Figure 4.5 The full PC Hadamard spectrum of the toy car image with $F = 64$ Blocks, some of the blocks are indicated.

Figure 4.5 shows the Hadamard domain partitioned into Walsh-ordered signature blocks. Some of the blocks with greater energy have been identified. The effect of the coarse-scale random modulator in PC algorithm is primarily to spread out the information in Block $\mathcal{B}0$. Moreover, since it complements the signature tile sizes, the 2-D sequences within each block are preserved. For instance, the patterns from Block $\mathcal{B}0$ captures the coarse $8 \times 8$ information, while the patterns from Block $\mathcal{B}4$ captures the $8 \times 4$ details, etc.

In Figure 4.6, we list three examples of patterns for both Block 36 and Block 54. They are examples to shows that the patterns from the same block have identical finer features determined the corresponding kernels.
The newly arranged order in the PC strategy, redistributes images’ energy concentration according to the finer features revealed in the field of view. Accordingly, the energy concentrates in a few blocks. The PC patterns have a good performance in anomaly detection applications, as the anomaly’s specific size and shape will show very specific feature in the PC spectrum. Then by choosing the kernels that have big coefficients associated with the anomalies, we can form a selection of patterns for such anomaly detection.
We look further into the Partial Complete algorithm and analyze groups of blocks in the Partial-Complete Hadamard spectrum. For a scene modulated by DMD with \(N\) pixels, we split the Hadamard domain into \(F = 4^K\) non-overlapping blocks, with \(B = N/F\) measurements per block. An important dual feature is that the pixel domain is partitioned at the same time into \(B\) non-overlapping tiles, each of size \(2^K\times2^K\) pixels. In this case the test scene has pixels of \(N = 256\times256 = 216\) pixels and \(F = 64\) blocks (labeled as \(B0, \ldots, B63\)), which results in \(B = 2^{10} = 1024\) measurements per block in the Hadamard domain; in the pixel domain this results in \(B = 2^{10}\) tiles partitioning the field of view (FOV), where each tile contains \(8\times8\) pixels.

We have shown that the Hadamard spectrum of an anomaly appearing in an arbitrary location within an \(8\times8\) tile of pixels was essentially the same regardless of which tile in the FOV. That is, there is a strong periodicity induced in the pixel domain by the structure of the Hadamard waveforms as a result of their Kronecker product structure. Thus it suffices to examine any \(8\times8\) tile.

However, certain groups of four blocks \(Bk\) in the Hadamard domain are complementary in the sense that they cover up nulls in the measurements corresponding to specific locations within an \(8\times8\) tile of pixels. Nulls in the measurements are undesirable since an anomaly corresponding to that location could result in it not being detected, i.e., a false negative.

The energy disturbance with and without the anomaly can also been clearly seen from the measured coefficients plots which is arranged with blocks. For
example in Figure 4.7, 64 blocks of coefficients are plotted and it can be clearly seen that some blocks show bigger response when the anomaly is in different locations. If one chooses blocks that have little responses to anomaly appearances, the detection will miss the anomaly.

**Figure 4.7** Examples of spectrum when varying the location of the anomaly in the simulated scene with moving cars and swinging Newton’s cradle. The x-axis is the block number and y-axis is the coefficients. The anomaly locations are also shown in the top of each graph.

This representation proves the necessity of choosing complementary block sets of patterns that maximize the anomaly response, which will also have better detection results in statistical analysis. The reason for those different features is caused by the finer structure defined by the corresponding signature. The signature of B36 shows a $2 \times 2$ checkerboard pattern with the basic element having the same size as the anomaly ($4 \times 4$). The signatures for other blocks have at least one group of pixels with the same size of anomaly. Another observation we have is the locations that B36 fails to cover actually are fully covered by either block of B38, B52 or B54, as shown in Figure 4.8.
Figure 4.8 Examples of shifting nodal patterns convolved with the anomaly shifting in the scene at different locations. Locations that one signature Block fails to cover actually are fully covered by one of the rest Blocks.

In the following simulation, we fix our anomaly size to $4 \times 4$ in a $256 \times 256$ scene, we then use partial complete algorithm generated patterns that contains the block sets of $B36$, $B38$, $B52$ and $B54$ to do the anomaly detection and compare the results with other block sets and algorithms. The simulation results are shown in Figure 4.9
Figure 4.9 The change detection using PWHT and PC block sets with different noise add to the simulation frames, images with different noise level are shown on the right.

The results show the chosen block sets has a significant improvement in detecting anomalies at different locations. We also decrease the intensity of the anomaly and test the detection performance, which also demonstrate that the proposed approach has a better chance to catch the anomaly with low intensity.
4.4. Quantizing detection performance

The evaluation of the detection performance by merely looking at the change graph is insufficient and lacks quantitative justification, thus we need some metric that can be used to better evaluate performance between different methods. In this project we choose receiver operating characteristic (ROC)[35], [36] and precision recall (PR)[37] as our comparison criteria.

In statistics, ROC curve is created by plotting the true positive rate (TPR) on y-axis against the false positive rate (FPR) on the x-axis at various threshold settings that illustrates the performance of a binary classifier system where discrimination threshold is varied. The true-positive rate is also known as sensitivity or recall in machine learning, it indicates the rate of correct detection. The false-positive rate reflects a given condition has been fulfilled, when it actually has not been fulfilled, in other word, a false alarm. The ROC curve is thus the sensitivity as a function of false alarm.

\[
TPR (Recall) = \frac{true \, positive}{true \, positive + false \, negative}
\]

\[
FPR = \frac{false \, positive}{false \, positive + true \, negative}
\]

The PR curve is a further method of quantitation and plots the precision as a function of recall (TPR). In information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant. Both precision and recall are therefore based on an understanding and
measure of relevance. Thus high precision means that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results.

\[
\text{Precision} = \frac{true \text{ positive}}{true \text{ positive} + false \text{ positive}}
\]

A measure that combines precision and recall is the weighted-harmonic mean of precision and recall, the traditional F-measure or balanced F-score\cite{38}:

\[
F_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}
\]

Where \(\beta\) is a weighting factor can be used as a selection to either put more emphasis either on precision or recall. In general the \(\beta\) is chosen to be 1 so the precision and recall are evenly weighted and this is noted as F1-score. In this project, F1-score is used as our primary parameter to evaluate the performance of the anomaly detection results. Other method such as G-score, which is the geometric mean of recall and precision, can be used for further validation of the detection performance.

### 4.5. Anomaly detection experiment with SPC

With the simulation results showing an advantage of using PC block sets as detection patterns, the next step is implementation and evaluation. A big merit of the anomaly detection system is that it can be implemented in the original SPC setup with only computational modification. This also allows the anomaly detection
function to be realized on the InView SPC camera devices that are compact and portable with all the optics and analog to digital conversion (ADC) electronics integrated as a whole system.

A Texas Instrument DMD chip (DLP7000) which has a resolution of 1024 × 768 is used to modulate the scene. The DMD has a maximum operation speed of 30 kilohertz and has a wide spectrum range of reflection, from 400 nm to near Infrared (NIR) of 2000 nm. The micro-mirrors have size of 13.6 μm in diagonal and tilt at two angles. The DMD is put at the location where the image plane of the scene is focused, and a total reflection (TIR) prism is used to channel one direction of the reflected light into a path where other optics elements are set to project the scene into light collection devices, in the initial case, a single detector. The scene we are using employs a toy car being pulled across a printed background at a constant speed; the background is a gray scale image of a parking lot and buildings. The photon collection device is a Thorlab (PDA36A) silicon detector has a maximum readout speed of 5 kilohertz at 70 dB noise reduction option, the effective detecting spectrum range from 400 nm to 1000 nm (Figure 4.10), which covers the whole visible wavelength.
Figure 4.10 Spectrum response of Thorlab PDA36A single detector

We used a red laser pointer as the anomaly, which has a wavelength of 700 nm. And has a maximum intensity of 5 mw/mm². The scene with and without the anomaly is shown in Figure 4.11. An intensity indication curve is plot along the line drawn across the scene to give an intuitive view of the anomaly intensity respect to the background illumination intensity.
Figure 4.11 Ground truth of the testing scene with (b) and without (a) the anomaly. (c) shows intensity profile across the line in (b) that pass through the anomaly

4.5.1. Laser triggered anomaly control system

In order to plot the ROC and PR curve, a ground truth showing the anomaly appearance is necessary. Thus we also developed a laser triggering system that verifies the synchronization between the laser trigger ground truth and the anomaly measurement data.

Figure 4.12 2-channel Relay set used to precisely control the appearance of a laser anomaly in anomaly detection experiments.

The trigger system is realized using a two-channel relay set (SunFounder 2 Channel 5V Relay) like the one shown in Figure 4.12. A relay is an electrically operated switch and is used where it is necessary to control a circuit by a low-power signal (with complete electrical isolation between control and controlled circuits such as TTL or ADC voltage output), or where several circuits must be controlled by
one signal. In our experiment, the relay set has two channels: one controls the laser as the anomaly and the other one control the motor that pulls the car moving across the scene. Figure 4.13 shows the results of the measurement data overlay with the laser trigger ground truth that obtained both from the National Instrument analog and digital convert device. We can see that the trigger edges of the laser match well with the light intensity increase in the measurement data.

![Figure 4.13 A comparison between the laser trigger ground truth (red) and the measurement data (blue), the DMD is running random Walsh-Hadamard pattern at 1 kHz.](image)

The relay system allows us to control the desire event precisely, providing confidence that consistency with measurements can be achieved even with small signal to noise ratio (SNR). Figure 4.14 displays a repetition of two experiments
with the exactly same parameters, and the measured triggers shows good consistency.

![Figure 4.14 Measurements of two laser trigger signals with two experiments that have the same parameters.](image)

With the functional laser trigger system, we are able to run the anomaly detection experiment with known trigger edge and repeatable experimental conditions, which will allow us to generate the ROC curve.

### 4.5.2. Improve ROC and PR curve through realistic signal broadening

The initial plot of precision recall (PR) curve has a low precision rate across the recall region, which gives us the impression that the detection algorithm is not very effective (Figure 4.15). However the real data shows the otherwise, thus a
closer look on the detection results and the comparison with the ground truth was performed.

![Average ROC Curve and Average Precision-Recall Curve](image)

**Figure 4.15 ROC and PR curve before the improvement.**

We found out that the detection results, which is the change plot in our case, always contains an ascending and descending slope when the anomaly is detected; on the other hand, the ground truth of the anomaly is more like a delta function, as shown in Figure 4.16. The difference between the real data and the ground truth leads to the low precision rate in the PR curve.
Figure 4.16 The comparison between the real data and the ground truth. The blue line is the change detection result, and red line is the anomaly indication.

Our approach to improve the PR curve plot is trying to match the real data with the ground truth; we used a curve broadening method to make the indication of the anomaly in the ground truth rise slower and comparable to the real data (Figure 4.17). The broadening of the anomaly ground truth improves the PR results significantly and that allows us to calculate more accurate F1-score for locating the optimized threshold (Figure 4.18).
Figure 4.17 The broadened ground truth that matches the detection data

Figure 4.18 ROC and PR curve after the improvement, the precision rate raises to above 50%.
4.5.3. Experiment results and discussion

As discussed previously, we have found that using the Hadamard patterns that contains particular signatures will provide higher SNR in detecting anomalies, in this experiment we use the signature combination of B36, B38, B52 and B54, with 100% patterns in each signature, so the total number of patterns are 4096, the DMD is running at 2000 frames per second and the patterns are repeated 10 times, the toy car is also moving across the scene, as shown in Figure 4.19, where the anomaly size is peaked about 4 × 4 in a 256 × 256 scene.

Figure 4.19 An example image of the scene, the anomaly is zoomed in and the toy car is moving across the scene during the experiment.

The laser pointer used as an anomaly is turned on and off three times in each measurement, a total of 10 measurements are performed to ensure that the
anomaly will be covered by all four signatures, each one represents the anomaly in a random location of the scene. Figure 4.20 shows the data collected when the anomaly appeared at 5th out of 10 locations, from which we can find the time stamp when the laser was turned on.

![Figure 4.20](image)

**Figure 4.20** Raw measurement data of anomaly location #5, the anomaly appeared three times (red), which can also be seen from the intensity increase in the measurement (blue).

Figure 4.21 shows the change detection results of anomaly location #5 and #8. As usual, the change detection algorithm is a combination of z-score and MMD. The spikes in the results of Figure 4.21 indicate detection of an anomaly, where the spikes aligned with the trigger ground truth (red lines) are the true detection (or true positive) while other spikes are false detections (or false positives).
The change detection result of two different anomaly locations (location 5 and 8). Each detection has the anomaly turned on and off three times (red and shaded area). The change detection is combination of z-score and MMD.

The ROC curve is generated using the 10 change detection measurements, each of which is compared with their own trigger ground truth using the ‘perfcurve’ function in the MATLAB. This will thus give us the true and false positive number in each measurement, along with the thresholds for each result. By averaging the number of true and false positives with respect to the threshold, we have the ROC and PR curve plotted and shown in Figure 4.22.
Figure 4.22 ROC curve and associated PR curve with laser trigger control

With the assistance of the laser trigger system, we are now able to compare the anomaly detection performance between the PC method and PWHT method through ROC and PR curve, Figure 4.23 shows the detection result of the two methods under different noise level. As in simulation, the PC method outperforms the PWHT approach under each circumstance.
Figure 4.23 ROC and PR curve plots for PC and PWHT patterns under different SNR circumstances, the SNR is labeled at the top of each graph.

4.6. Anomaly detection with FPA

The former experiments have proved that the SPC system is feasible for high-speed anomaly detection. However it has some limitations too. The single detector collects all the light reflected from the DMD and it becomes difficult to detect intensity increase contributed by anomaly that usually small compare to total light in the scene. Even with our proposed PC patterns that enhance the appearance of the anomaly, the SPC system’s performance suffers when the anomaly become dimmer or the background illumination increases. Another disadvantage of the SPC
anomaly detection system is its incapability of detecting the location of the anomaly in real time. The measured data can be used to reconstruct the scene for the purpose of locating the anomaly, yet this process is computationally demanding and the reconstruction may take seconds for a high-resolution scene. In order to improve the anomaly detection system and overcome the limitations of SPC imager, we replaced the single detector with a Hamamatsu focal plane array (FPA) that consists of $64 \times 64$ indium gallium arsenide (InGaAs) detectors. The FPA has a spectrum respond region of 700 nm to 2000 nm, with a maximum data readout speed of 1 kilohertz. Figure 4.24 shows some sample images taken by the FPA.

![Hamamatsu FPA Image](image)

**Figure 4.24** Image captured by the Hamamatsu $64 \times 64$ indium gallium arsenide (InGaAs) FPA. There exist defect pixels at location (60, 26) and (60, 27) of the FPA, which always give high intensity readouts.

The FPA is placed at secondary image plane after the DMD, and the optical system maps a $512 \times 512$ DMD region to the $64 \times 64$ pixel array. Ideally we would
use the data collected by the total 4096 detectors, and each detector on the FPA will collect light from a block of 12 × 12 DMD mirrors, the smaller region of interest (ROI) will significantly increase the SNR of the anomaly compare to the SPC system where the single detector ‘sees’ the whole DMD area.

4.6.1. Virtual channel

In previous analysis we have decided that some particular PC block set will be the optimum patterns to detect the anomaly. However the 12 × 12 mirror block on the DMD is insufficient for us to generate the block sets for the PC strategies. Also, smaller ROI means fewer measurements are necessary. To be consistent with our simulation process we need at least a DMD ROI resolution of 64 × 64 for each detector in the 64 × 64 Hamamatsu FPA, so a minimum of 4096 × 4096 resolution DMD is required to reproduce the results in the simulation, which is unavailable. Therefore we introduce the idea of virtual channel, which down samples the 64 × 64 FPA into an 8×8 FPA by grouping sub-block of 8 × 8 detectors in the FPA into one virtual detector. The virtual channel strategy has another benefit of fast processing speed. As currently we have limited computation power and data transmission bandwidth, it will take tremendous time to process the data from 4096 detectors without the help of parallel computing. In this project we will use an 8 × 8 virtual channel and each virtual channel will have a collection of light from 64 × 64 DMD mirrors. The whole image resolution is still 512 × 512. A comparison of data collected using SPC and FPA virtual channel strategy are illustrated in Figure 4.25,
and the ROC and PR are also plotted respectively so show the improvement of the FPA system, especially for the PR curve (Figure 4.26).

Figure 4.25 The detection signal from using different methods, SPC versus FPA. The FPA virtual channel has a significant improvement on SNR as each detector collects signal from a smaller ROI and the anomaly will have a bigger impact on the measurement.
4.6.2. Point spread function correction to the imaging system

Besides the use of virtual channel, we also compute the point-spread function (PSF) to correct the distortion of the image by the optical system. The PSF describes the response of an imaging system to a point source or point object. Suppose that we modulate the scene with an $m \times n$ DMD, because of optical skew and other alignment issues the DMD does not always line up perfectly with the FPA, so each virtual channel does not necessarily precisely consist of $f_m/v_m \times f_n/v_n$ detector elements, where $f_m \times f_n$ are the detector elements on the FPA and $v_m \times v_n$ are the mirror...
element in each virtual channel. Figure 4.27 shows the measured PSF of one DMD pixel on the FPA (right), and a comparison with the ideal PSF (left).

![Figure 4.27 PSFs of multiple DMD pixels imaged on the FPA, the measured PSF (right) shows the distortion and the skew of the imperfect imaging system, the ground truth of pixel displayed in the DMD is shown on the left.](image)

In addition, optical crosstalk between adjacent detector elements form the individual mirrors on the DMD is another factor that will affect the imaging and detecting performance for the FPA system. A point-spread function of the optical system is used to represent the mapping and response for the detector elements from the DMD mirrors. The PSF can be written as a $f \times N$ matrix $Q_{FPA}$ where $f = f_m \times f_n$ is the total number of detector elements and $N$ is total number of mirrors of the DMD, the light modulation matrix then can be written as $Q = Q_{virt} \times Q_{PSF} \times M$, the $Q_{virt}$ is a $v \times f$ matrix that models how the FPA’s $f$ detectors contribute to each of the $v$ virtual
channels, and $M$ is the dot product of the modulation matrix and the scene that has a dimension of $N \times m$, $m$ is how many measurements are performed and $m/N$ gives the compression ratio.

In practice we obtain $Q_{PSF}$ by raster scanning each pixel of the DMD and record the response from the FPA as an $f \times 1$ column vector in the $Q_{PSF}$, going through all the N mirrors on the DMD will eventually fill out the $f \times N$ matrix $Q_{PSF}$. The image before and after the PSF correction is shown in Figure 4.28

![Figure 4.28 FPA captured image with and without the correction of the PSF. The left image is reconstructed small image patches stitched together. The right image is recovered with the consideration of the PSF.](image)

4.6.3. Discussion of anomaly detection results with FPA

Possessing all the building blocks both in hardware and algorithm, we then performed a series of experiments to test our system, and monitored the detection threshold under each condition. We considered many testing situations that consist
of different anomaly intensity and duration, to be specific we considered the anomaly intensity of 980 μw/mm^2, 489 μw/mm^2 and 290μw/mm^2, the durations of 150 ms, 60 ms and 25 ms. For each pair of conditions we performed the anomaly detection process in ten different locations within the ROI of one particular virtual channel, so the ROC and PR curve can be generated for each.

The maximum F1-score are calculated from the generated ROC and PR curve from each condition and plotted (Figure 4.29), along with the optimum threshold for the detection purpose (Figure 4.30).

Figure 4.29 Maximum F_β-score calculated from the PR curve results, each corresponds to an anomaly intensity (y axis) and duration (x axis). Three different β numbers are chosen.
Figure 4.30 Detection threshold that determined by the maximum $F_\beta$ score associates with the corresponding condition, anomaly intensity in y axis and duration in x axis.

The maximum $F_\beta$-score is plotted with the correspondent threshold in Figure 4.30, as we can see from the plot that the max $F_\beta$-score decreases when the anomaly intensity is getting smaller. The bigger $\beta$ we choose the lower threshold we will have, which means it weights recall higher than precision and the algorithm becomes more conservative in detecting the anomaly.

As previously discussed we are choosing F1-score as our criteria to determine the threshold. As the threshold is differed as the anomaly has different intensities, in the arbitrary test we also characterize the SBR for these arbitrary tests so that we are able to decide which threshold to use. Figure 4.31 shows some of the testing results of detecting anomaly with arbitrary parameters, SBR is calculated according to the measured raw data and the corresponding threshold is chosen.
Figure 4.31 Arbitrary test of the anomaly detection using the pre-determined threshold, each row represents the anomaly intensity, which decreases from top to bottom. Each column represents the anomaly duration, which shortens from left to right.

We can see that the detection catches the anomaly with different durations and SBR above 1.02, also the threshold is optimal as that most of the false positive in the detection result is excluded. As the SBR is getting lower and the detection change contains more noise, no matter what threshold we choose the FPR is big and the detection will fail. However this is not due to the incapability of the detection algorithm or threshold determination method, rather a signal to noise problem that
defines the limit of our system. Yet there lies the potential to improve the detection performance by using smaller virtual channel or eventually the original resolution of the FPA, so that the SNR will greatly increase with lower anomaly intensity.
References


