Three-dimensional Imaging, Visualization and Recognition in Low Light Environments

Adam Markman
University of Connecticut - Storrs, amarkman89@gmail.com

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/1953
Three-dimensional Imaging, Visualization and Recognition in Low Light Environments

Adam Markman, Doctor of Philosophy

University of Connecticut, 2018

Abstract

There has been significant research in developing novel image sensors to capture a scene in low illumination environments. Using conventional passive imaging sensors that operate in the visible spectrum, camera read-noise becomes higher than the signal information from the scene resulting in read-noise dominant images. These images suffer from poor signal-to-noise ratio making scene visualization unobtainable. Thus, imaging in photon-limited environments has been viewed as a nuisance. This dissertation will consist of three separate parts that discuss utilizing photon-counting principals for secure information storage or imaging in photon-limited environments. First, an optical security system for information storage is introduced that uses secure three-dimensional optical phase codes whose unique optical signature can be authenticated using the random-forest classifier. Next, a system for securing three-dimensional integral imaging displays using a quick-response encoded elemental image array will be discussed that combines photon-counting with traditional optical encryption methods for secure information storage. The
second part of the dissertation will discuss integrated circuit authentication in photon-limited environments. More specifically, a photon-counting model will be applied to an image of an integrated circuit (IC) captured using an x-ray to demonstrate that IC authentication can be achieved when few photons are available in the scene. In the third part of this dissertation, three-dimensional (3D) imaging and object recognition in low light environments using real experimental data will be discussed. Afterwards, an object recognition framework for 3D reconstructed images obtained from elemental images taken in low light conditions using Convolutional Neural Networks will be introduced.
Three-dimensional Imaging, Visualization and Recognition in Low Light Environments

B.S., University of Connecticut, Storrs, CT, 2011

M.Sc., University of Connecticut, Storrs, CT, 2014

M.S., University of Connecticut, Connecticut, USA, 2016

A Thesis
Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Connecticut 2018
Doctor of Philosophy Dissertation

Three-dimensional Imaging, Visualization and Recognition in Low Light Environments

Presented by

Adam Markman, B.S., M.Sc., M.S.

Major Advisor ______________________________________________
Bahram Javidi

Associate Advisor ______________________________________________
Rajeev Bansal

Associate Advisor ______________________________________________
John Chandy
University of Connecticut

2018
ACKNOWLEDGMENTS

My graduate studies have been an adventure filled with challenging and rewarding experiences. The intellectual growth and friends I have met on this journey is invaluable. Without the support of my colleagues and mentorship of my advisors, this dissertation would not be possible.

I would like to first and foremost thank my major advisor, Prof. Bahram Javidi, for his guidance and mentorship. He awarded me the opportunity to pursue a Ph.D. under his guidance and supported me throughout my graduate studies by ensuring the projects we work on are novel and impactful with along with taking the time to have fruitful discussions on technical topics related to our research. He has taught me to question common assumptions and how to approach technical problems that have never be addressed before. Furthermore, he has provided incredible resources to help with my research by collaborating with world renowned researchers and major corporations.

I would also like to thank individuals that have helped me along the way whether it is through technical collaborations or mentorship. First, I would like to extend my sincere appreciation to my committee members, in no particular order: Prof. Rajeev Bansal, Prof. John Chandy, Prof. Helena Silva, and Prof. Faquir Jain. They have aided me and provided valuable feedback on my research. Furthermore, I would like to thank Prof. Artur Carnicer, Prof. Pedro Latorre Carmona and Prof. Arun Anand for their guidance and technical collaborations.
I wish to thank my lab mates, Xin Shen, Satoru Komatsu, Siddharth Rawat, Hee-Seung Kim, Timothy O' Connor, and Hyun-Woo Kim for their technical guidance, collaboration on projects, and providing advice and unique insights on research projects I have worked on. Their drive for success and learning is contagious and without their help I would not have been able to complete this dissertation.

Lastly, I am thankful to the support of my parents, Jayne and Wes Markman, along with my sister Rachel Markman. They have supported me throughout Ph.D. program and have offered their unconditional encouragement, love and support.
Table of Contents

List of Figures ........................................................................................................ viii
List of Tables ........................................................................................................... xiii

Part I: Secure Object Information using Photon-Counting Imaging ...................... 1
Chapter 1 Overview of imaging in low illumination conditions and its applications 2
  1.1 Introduction ........................................................................................................... 2
  1.2 Principle of the photon-counting model .............................................................. 4
      1.2.1 Optical Security .......................................................................................... 6
      1.2.2 Three-Dimensional Imaging ........................................................................ 15
  1.3 Imaging in Low illumination Environments ......................................................... 23
      1.3.1 Synthetic aperture integral imaging (SAII) experimental data .................... 23
  1.4 Organization of Thesis ....................................................................................... 24

Chapter 2 3D integral imaging displays using a QR encoded elemental image array ................................................................. 26
  2.1 Introduction ........................................................................................................... 27
  2.2 Overview of Integral Imaging .............................................................................. 28
  2.3 Double random phase encryption with Run-length encoding a Huffman coding 29
  2.4 Experimental results ......................................................................................... 37
  2.5 Compression with Photon-Counting ................................................................... 41
  2.6 Conclusion ......................................................................................................... 43

Chapter 3 Security authentication with a Three-Dimensional Optical Phase Code Using Random Forest Classifier ................................................................. 45
  3.1 Introduction ....................................................................................................... 45
  3.2 3D Optical Phase Code Design and Feature Selection ...................................... 47
  3.3 3D Optical Phase Code Design and Feature Selection ...................................... 53
  3.4 3D Optical Phase Code Authentication ............................................................... 56
  3.5 Conclusion ....................................................................................................... 60

Part II: Integrated Circuit Authentication under Photon-Limited Conditions ...... 61
Chapter 4 Integrated Circuit Authentication using Photon-limited X-ray Microscopy ................................................................. 62
  4.1 Introduction ....................................................................................................... 62
  4.2 Photon-limited X-ray images ............................................................................ 64
4.3 Photon-Counting authentication

4.4 Classification Results

4.5 Conclusion

Part III: Three-dimensional Imaging in Low Light Environments

Chapter 5 Three-dimensional object visualization and detection in low light illumination using integral imaging

5.1 Introduction

5.2 Statistical Analysis of Low Illumination Scene

5.3 Experimental Results

5.4 Face Detection

5.5 Conclusion

Chapter 6 Learning in the dark: 3D Integral Imaging object recognition in very low illumination conditions using Convolutional Neural Networks

6.1 Introduction

6.2 Three-Dimensional Integral Imaging in Low Illumination Conditions

6.3 Experimental Results

6.4 Object Classification Using Convolutional Neural Networks

6.5 Conclusion

BIBLIOGRAPHY
List of Figures

Fig. 1.2.1. (a) the original 256 x 256 grayscale image and the photon-limited image using approximately (b) 300,000 or 4.577 photons/pixel, (c) 30,000 photons in the scene or 0.457 photons/pixel and (d) 3,000 photons in the scene or 0.0457 photons/pixel.

Fig. 1.2.2. Schematic of (a) double-random-phase encryption (DRPE) encryption process and (b) the DRPE decryption process.

Fig. 1.2.3. (a) 256 x 256 gray scale image. The (b) amplitude and (c) the phase of the encrypted image using the double-random-phase encryption. (d) Corresponding decrypted image.

Fig. 1.2.4. (a) 256 x 256 gray scale image. The (b) phase and (c) photon-limited amplitude of the DRPE using 3e8 photons/pixel and (d) the decrypted image.

Fig. 1.2.5. (a) Authenticated decrypted image using the decrypted image shown in Fig. 1.2.4. A (b) false class reference image and the (c) corresponding authenticated image the false class image as the reference image.

Fig. 1.2.6. (a) 256 x 256 gray scale image. The (b) amplitude and the (c) phase of the full-phase encrypted image using the double-random-phase encryption. The (d) corresponding decrypted image.

Fig. 1.2.7. Integral imaging setup. The (a) pick up stage and (b) 3D reconstruction.

Fig. 1.2.8. 3D axially distributed sensing (a) pick up process and (b) reconstruction process.

Fig. 2.2.1. (a) Pick up and (b) display stages of an integral imaging system.

Fig. 2.3.1. Building the binary tree used in Huffman coding. (a) Generating the binary tree with root nodes and (b) assigning labels to each branch of the binary tree.

Fig. 2.3.2. (a) 19 x 19 pixels RGB image; (b) and (c) display the QR codes containing the real and imaginary information of for the red color channel from the compressed and ciphered image shown in (a), respectively; (d) and (e) depict a scanned QR code using the iPhone SCAN application revealing the real and imaginary information from the compressed and ciphered image, respectively, for the red color channel of the image shown in (a); (f) shows the decrypted and decompressed image.

Fig. 2.3.3. Storing multiple RGB encrypted elemental images inside of multiple QR codes.
Fig. 2.4.1. 19 x 19 pixels RGB (a) elemental images which were compressed using run-length encoding and Huffman coding followed by encryption using the double-random-phase encryption; computational 3D integral imaging reconstructions with 4 EIs at a distance (range) of (b) at 65 mm focused on “3” and (c) at 115 mm focused on “D”.

Fig. 2.4.2. Optical 3D integral imaging display setup using a Smartphone and a lenslet array.

Fig. 2.4.3. 3D displays results using Smartphone. (a) 3D Optical reconstruction with integral imaging using the primary elemental images. (b) Optical 3D reconstruction using the decrypted and decompressed elemental images obtained from the compressed double-random-phase encryption elemental images in QR codes.

Fig. 2.5.1. 3D photon counting integral imaging compression and security experiments. A 54 x 96 elemental image (EI) array consisting of 19 x 19 pixels RGB EIs was used. (a) shows an EI while (b) depicts the corresponding photon-limited elemental image using about 8 photons per pixel on each color channel of the EI. 3D optical reconstruction after decryption and decompression is shown using (c) the original EIs and (d) when photon counting on the EIs was used.

Fig. 3.2.1. Workflow for developing the 3D optical phase code for (a) 3D optical phase code A, (b) 3D optical phase code B, and (c) 3D optical phase code C.

Fig. 3.2.2. (a) Experimental 4 mm x 4 mm QR code and a picture of (b) 3D optical phase code A, which consists of a QR code with an optical phase mask consisting of a glass slide and diffuser paper.

Fig. 3.2.3. Optical experimental setup. A 455 nm blue laser diode is transmitted through a polarizer and lens. The laser is then transmitted through the QR code which has a 3D optical phase mask placed on it. A CCD sensor, a distance d away from the QR code, records the speckle pattern.

Fig. 3.2.4. (a,b,c) The speckle patterns obtained using 3D optical phase codes A, B, and C, respectively. The 3D optical phase codes are a distance of 110 mm from the CCD sensor. The corresponding histograms and a fitted Gamma distribution are also shown.

Fig. 3.4.1. Overview of the 3D optical phase code authentication system.

Fig. 3.4.2. Out-of-bag classification error using 100 trees which converges to about 0.0067.

Fig. 3.4.3. A false class speckle pattern obtained by placing 3D optical phase code C 100 mm from the CCD sensor.
Fig. 3.4.4. Effect of displacement of 3D optical phase code A on the class confidence score output from the random forest classifier for distances 70 mm, 110 mm, and 150 mm from the CCD sensor.

Fig. 4.2.1. (a) X-ray microscope used in the experiments (interior view). (b) An integrated circuit (IC) x-ray image obtained using 80 kV and 7 W with x-ray microscope. (c) The IC x-ray image obtained using 40 kV and 1 W.

Fig. 4.2.2. Example of a real authentic 256 (W) x 512 (V) pixels (a) integrated circuit. A photon-limited IC is generated based on the (b) photon-counting model with no weights and (c) photon-counting model using the weighted photon-counting method with 1.5e6 photons in the image, or 11.4 photons/pixel.

Fig. 4.3.1. Constructing an integral image.

Fig. 4.3.2. Example of a 256 (W) x 512 (V) pixel counterfeit images where (a) the wires have been tampered along and (c) the die has been changed. (b) and (d) depict their corresponding photon-limited IC image, respectively, using 1.5e6 photons, or 11.4 photons/pixel based on the weighted photon-counting method.

Fig. 4.4.1. Graph of percent accuracy vs. total number of photons for SURF feature matching using 16 test images. The images were counterfeit and authentic 256 (W) x 512 (V) pixel photon-limited IC x-ray images using the photon-counting model based on the weighted photon-counting method (PC-W) and photon-counting model without weights (PC-NW). There were 16 images used for training which were photon-limited.

Fig. 5.2.1. Concept of integral imaging (II). (a) optical pickup, and (b) 3D computational reconstruction. z is the reconstruction distance, g is the focal length, p is the pitch between imaging sensors, and c is the sensor size.

Fig. 5.2.2. Read-noise dominant elemental image and its corresponding histogram.

Fig. 5.3.1. The (a) synthetic aperture integral imaging (SAII) used in the outdoor, low light experiment and the (b) USB board camera used in the experiment. z is the reconstruction distance, g is the focal length, p is the pitch between imaging sensors, and c is the sensor size.

Fig. 5.3.2. The (a) first elemental image with SNR of 2.04 and (b) the 3D reconstructed image of a scene reconstructed at z = 5 m with an SNR of 2.66.

Fig. 5.3.3. Experimental results under low light conditions for Scene 1: (a) read-noise limited EI with an SNR of 0.646, and (b) corresponding 3D reconstructed image at z = 3.9 m having an SNR of 0.942. The red box denotes a detected face using
the Viola-Jones algorithm. The face was detected using 3D imaging while it was unable to be detected using 2D imaging.

Fig. 5.3.4. Experimental results under low light conditions for Scene 2. (a) read-noise limited EI under low light conditions with an SNR of 0.486 along with (b) the corresponding reconstructed 3D image at \( z = 4.3 \) m having an SNR of 0.893. Moreover, (c) is the face after total-variation regularization denoising that was able to be detected. Experiments with a third scene (scene 3): (d) read-noise limited EI under low light conditions with an SNR of 0.272 and (e) the corresponding reconstructed 3D image at \( z = 5.1 \) m having an SNR of 0.796. Moreover, (f) is the face after total-variation regularization denoising that was able to be detected. The red boxes denote detected faces using the Viola-Jones algorithm.

Fig. 5.3.5. Histogram of (a) object (i.e. the person’s face) from the elemental image [see Fig. 5.3.4(a)] with an entropy of 3.97 bits/pixel, and (b) the histogram of the object in the 3D reconstructed image shown in Fig. 5.3.4(b) with a reduced entropy of 2.17 bits/pixel.

Fig. 6.2.1. Integral Imaging. (a) pickup, and (b) 3D reconstruction stages. \( c \) = sensor size, \( p \) = pitch, \( f \) = focal length, \( z \) = distance. (c) Parameters details in (a). \( R_i \) = chief ray, \( \ell_i \) = the i-th lens; \( \theta \) = azimuth angle; \( \phi \) = zenith angle.

Fig. 6.2.2. Synthetic aperture integral imaging (SAII) pick-up and reconstruction stage

Fig. 6.3.1. Sample elemental images with corresponding SNR and photon/pixel. The SNR of the images shown in Fig. 6.3.1(d-f) cannot be computed as the average power of the object regions is less than that of the background. N/A = not applicable.

Fig. 6.3.2. Three-dimensional (3D) reconstructed images at \( z = 4.5 \) m corresponding to the elemental images shown in Fig. 6.3.1. The SNR and photons/pixel can be computed for low light levels.

Fig. 6.3.3. The (a) elemental image and (b) 3D reconstructed image at \( z = 4500 \) mm. The \( SNR_{contrasts} \) are 31.5 dB and 33.76 dB, respectively.

Fig. 6.3.4. 2D captured images having an exposure time of (a) 0.010 s, (b) 0.015 s and (c) 0.203 s under low light conditions. The \( SNR_{contrasts} \) of the images shown in (a) and (b) cannot be computed as the object region intensity is less than that of the background. The \( SNR_{contrasts} \) of (c) is 8.213 dB.

Fig. 6.3.5. (a) Average of 72 elemental 2D images, and (b) the 3D InIm reconstructed image at \( z = 4.5 \) m using an exposure time of 0.015 s for each elemental image. The \( SNR_{contrasts} \) is 6.38 dB in (a) and 16.702 dB in (b), respectively. (c) Average of 72 elemental 2D images and (d) the corresponding 3D InIm reconstructed image
at $z = 4.5$ m using an exposure time of 0.010 s for each elemental image. The SNR$_{\text{contrasts}}$ is 2.152 dB in (c) and 15.94 dB in (d), respectively.

Fig. 6.4.1 Correlation between 3D reconstructed image at $z = 4.5$ m after TV-denoising using EIs (a) obtained under SNR of 10.41 dB as a reference. This 3D image was correlated with 3D reconstructed images after TV-denoising whose EIs were obtained under an SNR of -12.75 dB for (b) true class object, and (c) false class object; with correlation values of 0.58 and 0.48, respectively. Classification is difficult.

Fig. 6.4.2. Example denoised 3D reconstructed training images acquired at various SNRs, reconstruction depths, additive noise and rotations
List of Tables

Table 2.4.1. Specifications of 3D display setup.

Table 3.2.1. Example of mean, variance, skewness, kurtosis, and entropy calculated for recorded speckle patterns.

Table 3.2.1. Example of mean, variance, skewness, kurtosis, and entropy calculated for recorded speckle patterns.

Table 3.4.1. A 12 class system was developed using 3D optical phase codes (OPC) A, B, and C and placing the 3D codes at different distances from the CCD sensor.


Table. 5.3.1. Scene metrics. \( z \) = reconstruction distance; EI = elemental image; 3D Rec. = 3D reconstructed image; SNR = signal-to-noise ratio; inc. = increase; dec. = decrease.

Table 6.4.1. Confusion matrix for face recognition using the CNN trained under low illumination conditions with 3D reconstructed images using 4 test scenes for each of the 6 subjects.
Part I: Secure Object Information using Photon-Counting Imaging
Chapter 1

Overview of imaging in low illumination conditions and its applications

1.1 Introduction

Imaging in environments under poor illumination conditions has garnered great interest in recent years such as for pedestrian detection [1], mobile devices [2], astronomy [3], and biological imaging [4][5]. However, imaging in these environments proves difficult using passive imaging sensors that operate in the visible spectrum. The recorded images suffer from read-noise [6], which occurs due to the imaging sensor’s on-chip signal amplification prior to analog-to-digital conversion, is higher than the signal recorded from the scene resulting in noise-like images. To overcome this issue, specialized hardware has been developed that can operate in low illumination conditions leading to improved scene visualization, such as the electron-multiplying charged coupled device (EM-CCD) [7], scientific complementary metal oxide semiconductor (sCMOS) [8], photon-counting detectors [9], to name a few. In addition, computational approaches have been developed to improve image quality [11][12][13].

Photon-counting devices may be employed to operate in low light conditions; however, they are bulky and expensive. It is possible to computationally generate photon-counting
images to simulate a photon-counting device [14][15][16]. In recent years, the photon-counting computational model has found applications in optical security [17][18][19][21][22] such as the well-known double-random-phase encryption [23] and full-phase double-random phase encryption [24]. The photon-counting model has been shown to improve the security of the optical security system by adding a non-linear transformation to the amplitude of the complex encrypted image making the encryption technique more robust to attacks such as plain-text, cypher-text, and brute-force techniques.

Photon-counting has also found applications in a three-dimensional (3D) imaging technique known as 3D integral imaging (InIm) [16][25]-[31]. 3D Integral imaging is a passive, multi-perspective imaging system that uses incoherent light to record multiple perspectives of a scene, known as elemental images, using a lenslet array, array of cameras, or a single camera on a moving platform. Integral imaging reconstruction may be performed optically or computationally [32]-[38]. 3D InIm has found applications in a wide range of fields such as 3D display [34]-[38], imaging in turbid media [39], gesture recognition [41], medical imaging [42], scattering media [39][43], and target recognition [44][45]. In particular, photon-counting has been applied to the recorded elemental images [16][25]-[30] used in 3D InIm. The scene information depicted in each elemental image is severely degraded; however, upon computational 3D reconstruction, scene visualization is improved. As a result, object detection [28][29] can be achieved. Recently, real data was acquired under low illumination conditions using a 16-bit cooled CCD device. Each elemental image provided little visual information as the images were read-noise dominate;
however, scene visualization was achieved upon 3D reconstruction and an advanced denoising algorithms known as total-variation regularization [46].

In this chapter, we will overview the photon-counting model when applied to the well known double-random-phase encryption and 3D imaging techniques. We will also discuss imaging in low illumination conditions using 3D imaging.

1.2 Principle of the photon-counting model

It is of great interest to model the rate of arrival of photons arriving on a sensor. A common approach is to assume that the photon-flux from the scene follows a Poisson distribution. However, other distributions may be used such as the binomial distribution, negative binomial distribution, multinomial distribution, or negative multinomial distribution depending on the coherent state of light [15]. The energy of a photon can be computed as:

\[ E = \frac{hc}{\lambda}, \]

where \( E \) is the energy, \( h \) is Planck’s constant, \( c \) is the speed of light in a vacuum, and \( \lambda \) is the wavelength of light. Thus, energy is inversely proportional to the wavelength of light.

Moreover, the fewer the number of photons, the sparser the scene becomes due to less photons arriving at a pixel. A photon-counting detector can be mathematically modeled and the number of photons arriving at pixel \( j \) can be modeled as [14][15]:
\[ P(l_j; \lambda_j) = \frac{[\lambda_j]^l_j e^{-\lambda_j}}{l_j!} \quad \text{for } \lambda_j > 0, \ l_j \in \{0,1,2,\ldots\} \]

where \( P(.) \) denotes probability, \( l_j \) is the number of photons detected at pixel \( j \) and \( \lambda_j \) is the Poisson parameter defined as \( N_p x_j \), where \( N_p \) is the number of photons in the scene and \( x_j \) is the normalized irradiance at pixel \( j \) such that \( \sum_{j=1}^{M} x_j = 1 \) with \( M \) being the total number of pixels.

Figure 1.2.1 depicts an example of a simulated photon-limited image. Fig. 1.2.1(a) depicts the original 256 x 256 grayscale image. Fig. 1.2.1(b) is the photon-limited image using approximately 300,000 or 4.577 photons/pixel. Fig. 1.2.1(c) and Fig. 1.2.1(d) are the photon-limited images using 30,000 photons in the scene (0.457 photons/pixel) and 3,000 photons in the scene (0.0457 photons/pixel), respectively. As the number of photons decreases, image visualization deteriorates as fewer photons are arriving at each pixel.
1.2.1 Optical Security

A. Double-Random-Phase encryption

The double-random-phase encryption is an optical encryption technique that can be implemented optically or computationally [23][24][47]-[51]. The double-random-phase encryption will be reviewed using one-dimensional notation. We define \( (x) \) and \( (v) \) as coordinates in the spatial and frequency domain. Moreover, \( f(x) \) is the input image while
\( n(x) \) and \( b(\nu) \) are two random noises that are uniformly distributed over the interval \([0,1]\).

The encrypted image for the amplitude-based double-random phase encryption is:

\[
\xi(x) = \{ f(x) \exp[j2\pi n(x)] \} * h(x),
\]

where \(*\) denotes convolution, \( \exp[j2\pi n(x)] \) is a phase mask in the spatial domain, and \( h(x) \) is a phase mask whose Fourier transform is \( \exp[j2\pi b(\nu)] \).

Using a random phase mask in both the spatial and frequency domain is necessary. For example, using a phase mask in the spatial domain results in a non-stationary signal:

\[
\langle g(x_1) g^*(x_2) \rangle = \langle f(x_1) e^{j\phi(x_1)} f^*(x_2) e^{-j\phi(x_2)} \rangle,
\]

\[
= f(x_1) f^*(x_2) \langle e^{j\phi(x_1)} e^{-j\phi(x_2)} \rangle,
\]

\[
= f(x_1) f^*(x_2) \delta(x_1 - x_2),
\]

where \( \delta(.) \) denotes a Dirichlet delta function, \( \phi(x) \) is a random function uniformly distributed on the interval \([0,1]\), and \( f(x) \) is the input signal, \( \langle \cdot \rangle \) denotes mean ensemble, \( g(x) = f(x) e^{j\phi(x)} \), and \( e^{j\phi(x)} \) denotes the phase mask in the spatial domain.

It can be shown that for the double-diffuser system [see Eq. 1.2.3], the autocorrelation

\[
\langle g(x_1) g^*(x_2) \rangle = E_{u_0} \delta(\tau_1 - \tau_2),
\]

where \( E_{u_0} \) denotes the energy of the signal, and \( (\tau) \) and \( (x) \) are spatial coordinates. The resulting encrypted image is a complex signal that is white and wide sense stationary signal. Moreover, information scattered throughout the encrypted image allowing the encrypted image to be robust to additive noise or cropping attacks. Decryption is the reverse of the encryption process:
\[ f_{\text{decrypt}}(x) = \left| F \{ F \{ \xi(x) \} \exp(-j2\pi b(\nu)) \} \right|^2, \]
\[ = \left| f_{\text{decrypt}}(x) \right|^2, \]

where \( F(.) \) denotes the Fourier transform. Note that \( f_{\text{decrypt}}(x) \) is equivalent to \( \left| f_{\text{decrypt}}(x) \right|^2 \) as the input signal is assumed to be real and positive.

Fig. 1.2.2 shows an optical representation of the DRPE encryption and decryption process. More specifically, Fig. 1.2.2(a) shows the encryption process implement using a 4-f system where \( f \) is the focal length of the lens. Using this architecture, the size of the input image is identical to size of the output image. A plane wave illuminates the input image \( f(x,y) \), which is then multiplied by the random phase mask \( \psi_1(x,y) = \exp[j2\pi n(x,y)] \) by placing it sufficiently close to the input image. The complex signal then propagates through a lens, \( \ell_1 \), resulting in the Fourier transform followed by multiplication by a second phase mask located in the focal plane of the lens, denoted as \( \psi_2(\nu,w) = \exp[j2\pi b(\nu,w)] \). The resulting signal passes through a second lens, \( \ell_2 \), located a distance \( f \) away and recorded by a sensor located on the focal plane of the second lens. The complex encrypted image \( \xi(x,y) \) has amplitude \( |\xi(x,y)| \) and phase \( \angle \xi(x,y) \). Fig. 1.2.2(b) depicts the decryption process, which is simply the reverse of the encryption process. Note, the complex conjugate of the phase masks are used for decryption.

An example of a computational implementation of the DRPE is shown in Fig. 1.2.3. Fig. 1.2.3(a) depicts the 256 x 256 grayscale image. The image is then encrypted. Fig. 1.2.3 (b) and (c) shows the amplitude and phase information of the complex encrypted image,
respectively. Fig. 1.2.3(d) shows the corresponding decrypted image revealing the input image.

We note that the double-random-phase encryption is prone to numerous attacks, such as plain-text and cypher-text attacks [52]. If the phase mask in the frequency is compromised i.e. an attacker gains access to this information, the DRPE system is compromised. Although many attacks can be prevented by updating the phase masks used in the encryption system.

![Diagram of double-random-phase encryption (DRPE) encryption and decryption processes.](image)

**Fig. 1.2.2.** Schematic of (a) double-random-phase encryption (DRPE) encryption process and (b) the DRPE decryption process. \( f \) = focal length, \( \ell_1 \) and \( \ell_2 \) denotes lenses.
Fig. 1.2.3. (a) 256 x 256 gray scale image. The (b) amplitude and (c) the phase of the encrypted image using the double-random-phase encryption. (d) Corresponding decrypted image.

B. Double-random-phase encryption with photon-counting

The photon-counting model [see Eq. 1.2.2] can be used in the double-random-phase encryption system [19][20]. As the DRPE is a linear encryption system, it is vulnerable to a multitude of attacks such as cyphertext and plaintext attacks. The photon-counting model is a non-linear operation and may be applied to the amplitude of the encrypted image.
improving the security of the DRPE. Fig. 1.2.4(a) depicts a 256 x 256 input image. Fig. 1.2.4(b) depicts the phase of the encrypted image. Photon-counting was then applied to the amplitude of the encrypted image resulting in image shown in Fig. 1.2.4(c), where the photon-counting model used approximately 3.8 photons/pixel.

Using the double-random-phase encryption with photon-counting, the decrypted image is still noise-like [Fig. 1.2.4(d)] and cannot be verified visually. One possible way to authenticate the decrypted image is by using nonlinear correlation filter [53][54]. The \( k^{th} \) order nonlinear filter [53] is an example of a nonlinear correlation filter and was used for authentication due to its simplicity and ease of implementation. The filter is presented, using \((v,w)\) to represent the coordinates in frequency domain, as:

\[
c(x, y) = \text{IFT} \left\{ |F_{ph}(v,w) F(v,w)|^k \exp\left[ j \left( \phi_{ph}(v,w) - \phi(v,w) \right) \right] \right\},
\]

where \( k \) is the strength of the applied nonlinear that suppresses the amplitude thus determining the performance of the filter, \( \text{IFT} \) is the inverse Fourier transform, \( F(v,w) \) is the Fourier transform of the input image, \( F_{ph}(v,w) \) is the Fourier transform of the photon-limited decrypted image, \( \phi_{ph}(v,w) \) and \( \phi(v,w) \) is the Fourier phase obtained for the decrypted image and input image, respectively, and \(||\) is the modulus operator.
Fig. 1.2. 256 x 256 gray scale image. The (b) phase and (c) photon-limited amplitude of the DRPE using 3.8 photon/pixel and (d) the decrypted image.

Fig. 1.2.5 below depicts an example of an authenticate image containing a significant peak. Fig. 1.2.5(b) shows a false class image. Moreover, Fig. 1.2.5(c) shows the corresponding authenticated image using the false class image as the reference image resulting in a correlation output without a significant peak.
C. Full Phase Double-Random-Phase Encryption with Photon-Counting

It has been shown that the full-phase DRPE [24] may improve the security of the encryption system. The full-phase processor is defined as:

$$
\psi(x) = \{\exp[j\pi f(x)] \times \exp[j2\pi n(x)]\} * h(x),
$$

1.2.7
where * denotes convolution, \( \exp[j2\pi n(x)] \) is a phase mask in the spatial domain, and \( h(x) \) is a phase mask whose Fourier transform is \( \exp[j2\pi b(\nu)] \).

The full phase PC-DRPE encrypted image can be decrypted in a similar fashion. Since the input image is real and positive, the photon-limited decrypted image, \( f_{ph\text{full}}(x) \), is:

\[
\left| f_{ph\text{full}}(x) \right| = \left| \text{Arg} \left\{ A \exp \left[ j\pi f_{ph\text{full}}(x) \right] \right\} / \pi \right|,  \tag{1.2.8}
\]

where \( A \) is the amplitude of the decrypted image, \( || \) is the modulus operation, and \( \text{Arg}(\cdot) \) is the argument function that restricts the phase angle from \(-\pi\) to \( \pi \).

Fig. 1.2.6(a) depicts an example of the original image. Fig. 1.2.6(b) shows the amplitude of the encrypted image while Fig. 1.2.6(c) shows the phase of the encrypted image while Fig. 1.2.7(d) is the decrypted image, which was able to be fully recovered.

In [21], the photon-counting model was applied to the amplitude of the encrypted image. At a low number of photons, the full-phase DRPE with photon-counting was shown to require fewer photons for authentication when using a binary input image.
1.2.6 Three-Dimensional Imaging

A. Synthetic Aperture Integral Imaging with Photon-counting

Three-dimensional (3D) imaging provides lateral information of a scene in addition to depth information. Passive imaging techniques have been investigated to provide true 3D imaging. In particular, integral photography was first proposed by Lippmann in 1908 [55]. Prof. Lippman proposed capturing multiple perspectives of a scene using a lenslet array
or a camera array that are recorded on a photographic film. The captured information is then reconstructed with full parallax and quasi-continuous viewing angles by back projecting the captured light rays. Moreover, the 3D reconstructed image can be seen without glasses or any viewing devices. Integral imaging [25][26][56][57] is a passive 3D imaging technique based on Prof. Lippman’s work. However, digital image sensors are used to capture the light rays. By capturing different perspectives of the scene, known as elemental images, a 3D image can be reconstructed either optically or digitally providing lateral and depth information of a scene.

Scene visualization using passive imaging techniques in photon-limited environments is a difficult task. As the number of photons decrease, an image becomes read-noise dominant thus generating an image with few photons. There are passive imaging techniques to overcome this limitation. Night vision amplifies the few photons available. If not enough photons are present in a scene, an NIR source is used to illuminate the scene. Moreover, longwave infrared (LWIR) imaging records the black body radiation of a scene. However, this method suffers from poor resolution thus making it difficult to identify objects in a scene e.g. identify a person’s face. 3D integral imaging [26] is a technique that uses multiple cameras that operate in the visible domain to capture different perspectives of a scene. By using multiple cameras, the total number of photons captured from the scene increases. Thus, by combining the information obtained using integral imaging, a 3D scene may be reconstructed using passive imaging techniques with improved visualization compared with a single elemental image. An advantage of this method is that low-cost, off the shelf sensors may be used to perform scene visualization in low light environments. As
no active sources are used, numerous applications can utilize this imaging technique such as in the military scenarios where an adversary would not be notified of the image sources locations.

Integral imaging (InIm) is a three-dimensional (3D) imaging technique that captures multiple perspectives of a scene, known as elemental images, using an array of cameras or a lenslet array. Fig. 1.2.7 depicts an InIm setup using an array of cameras. The technique is able to capture both angular information and intensity information. By combining all of the EI perspectives, a 3D reconstructed image can be computationally or optically reconstructed. In synthetic aperture integral imaging (SAII), a single camera is used to capture the different perspectives by translating the camera horizontally or vertically. To generate a 3D reconstructed image, computational reconstruction can be used based on the pinhole model. The captured light rays from the EIs are back propagated to their pixels on each EI through their corresponding virtual pinholes and overlap in the 3D imaging space. At the reconstruction plane, \( z \), all objects that lie on that plane are in focus while information not located on that plane are blurred. The computational integral imaging reconstruction algorithm is defined as [26]:

\[
I(x, y; z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \left[ E_{k,b} \left( x-k \frac{L_x}{c_x \times M}, y-b \frac{L_y}{c_y \times M} \right) \right],
\]

where \((x, y)\) is the pixel index, \( z \) is the reconstruction distance, \( O(x, y) \) is the overlapping number on \((x, y)\), \( K \) and \( B \) are the total number of EIs obtained in each column and row, respectively; \( E_{k,b} \) is the ideal EI in the \( k \)-th column and \( b \)-th row, \( L_x \) and \( L_y \) are the total number of pixels in each column and row, respectively, for each \( E_{k,b} \), \( M \) is the magnification.
factor and equals \( z/g \), \( g \) is the focal length, \( p \) is the pitch between image sensors, \( c_x \) and \( c_y \) are the size of the image sensor.

Fig. 1.2.7. Integral imaging setup pick up and reconstruction stage.
In [15], photon-counting was applied to each elemental image followed by computational 3D reconstruction. The resulting 3D reconstructed image was shown to have a higher peak signal-to-noise ratio compared with a single 2D elemental image.

Denoising algorithms have been applied to the elemental images to improve the 3D reconstructed images by applying total-variation regularization expectation maximization (TV EM-MAP) [30]. It was shown that by using TV EM-MAP on photon-limited elemental images, the quality of the elemental images improved such that the image sensor positions used to acquire the elemental images may be estimated assuming two are known a priori [31].
B. Axially Distributed Sensing with Photon-Counting

Axially distributed sensing (ADS) [59] is a passive sensing 3D imaging technique that captures a scene by moving a camera at different depths along its optical axis which is perpendicular to a 3D scene. Each captured 2D image, known as an elemental image (EI), is then used to reconstruct the 3D scene since each EI obtains a unique perspective of the scene. Figure 1.2.8 depicts an example of the ADS pick up and reconstruction stages. A camera captures 2D images of a 3D scene at multiple distances along its optical axis with a step size of ΔZ, as shown in Fig. 1.2.8(a). For the reconstruction process, as shown in Fig. 1.2.8(b), a reference camera position is set at the first capture position (C₀). With the pinhole model, the distance between the pinhole and captured elemental image is g. The pixels on each captured EI can be mapped into multiple planes in the 3D space to reconstruct the 3D scene. Mathematically, 3D reconstruction is performed by:

\[ I_{z_0}(x, y) = \frac{1}{K} \sum_{n=0}^{K-1} I_n \left( \frac{x}{M_n}, \frac{y}{M_n} \right), \quad \text{where} \quad M_n = \frac{Z_n}{Z_o}, \quad 1.2.10 \]

where K is the total number of the elemental images obtained by the ADS pick up process, \( I_n \) is the \( n \)-th elemental image, \((x, y)\) is the pixel index, \( M_n \) is the relative magnification of the \( n \)-th image with respect to the closest image, \( Z_o \) is the initial distance of the camera from the scene and \( Z_n \) is the distance of the camera when it is at position \( n \).

In [60], photon-counting was applied to each elemental image generating a photon-limited image:
\[ I_{Zo}(x, y) = \frac{1}{K} \sum_{n=0}^{K-1} I_{n}^{ph}\left(\frac{x}{M_n}, \frac{y}{M_n}\right), \quad \text{where} \quad M_n = \frac{Z_n}{Z_o}, \]

where \(I_{n}^{ph}\) denotes the \(n\)th photon-limited image.

To improve the quality of the 3D reconstructed image, a total-variation regularization maximum a priori expectation maximization approach (TV) was used on each 2D elemental image. Upon 3D reconstruction, the peak signal-to-noise ratio is higher than when TV MAP-EM was not applied to the EIs.

Fig. 1.2.8. 3D axially distributed sensing (a) pick up process and (b) reconstruction process [61].
C. Polarimetric Imaging with Photon-Counting

A method for estimating the degree of polarization (DoP) in photon-counting 3D integral images has been developed [62]. As the information recorded in low light illumination condition is very sparse, calculation of the Stokes parameters and the DoP becomes a difficult task. In [62], polarimetric images were acquired for a set of elemental images for 3D integral imaging. 3D image reconstruction was then performed for each polarimetric image. At a fixed 3D reconstruction distance, the Stokes parameters were computed. This information was then used to find the degree of polarization (DoP) for a particular depth plane. These results were then compared to when photon-counting was performed on the elemental images followed by 3D reconstruction and computation of the DoP. To improve the reconstruction results, a total-variation regularization denoising algorithm was applied to the 3D reconstructed image. Pattern recognition was also shown to be achievable using the 3D reconstructed images after image denoising.

The Stokes parameters were computed by first capturing size polarimetric images denoted as $I^{0.0}, I^{90.0}, I^{45.0}, I^{135.0}, I^{135.\pi/2}, I^{135.\pi/2}$, where $I^{0.0}$ is the captured intensity for when the linear polarizer is at angle of $\alpha$ relative to the x-direction and $I^{\alpha.\pi/2}$ denotes the intensity image capture when a quarter wave-plate is used in addition to the polarizer. The Stokes parameters are then computed followed by the degree of polarization. The Stokes parameters are computed as:
The degree of polarization can then be computed as:

\[ \text{DoP} = \frac{1}{S_o} \sqrt{S_1^2 + S_2^2 + S_3^2} \], where DoP \leq 1, \tag{1.2.13} \]

The polarimetric distribution was computed for each elemental image i.e. \( i_{k,b}^{0^\circ,0^\circ} \), \( i_{k,b}^{90^\circ} \), \( i_{k,b}^{45^\circ,0^\circ} \), \( i_{k,b}^{135^\circ} \), \( i_{k,b}^{45^\circ,\pi/2} \), \( i_{k,b}^{135^\circ,\pi/2} \) where \( k \) and \( b \) indicate the elemental image index. The reconstructed image for each degree of polarization becomes:

\[ I^{\alpha,\beta}(x, y, z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} i^{\alpha,\beta}_{k,b} \left( x - k \frac{L_x \times p}{c_x \times M}, y - b \frac{L_y \times p}{c_y \times M} \right), \tag{1.2.14} \]

where \( \alpha, \beta \) denotes the degree of polarization.

The photon-counting model was then applied to each recorded polarimetric distribution followed by 3D reconstruction. After 3D InIm reconstruction, the stokes parameters are computed using the 3D InIm reconstructed images followed by the DoP. Total-variation denoising was applied to the 3D reconstructed images followed by computation of the DoP. The polarimetric information is retained even in low light environment.

### 1.3 Imaging in Low illumination Environments

#### 1.3.1 Synthetic aperture integral imaging (SAII) experimental data
In [46] a 16-bit cooled camera was used to record real-world data under low illumination conditions. A 3D InIm pickup stage was used to acquire elemental images. Due to the low photon-flux of the scene, the elemental images were read-noise dominant. A penalized maximum-likelihood expectation maximization algorithm (PM-TV) was used on the elemental images to improve signal-to-noise ratio (SNR) followed by 3D InIm reconstruction. Moreover, the photons per pixel and SNR were estimated. The mean squared error of the 3D reconstructed images were computed and found that by utilizing the PM-TV algorithm, the MSE was lower in the reconstructed images when PM-TV was used for denoising.

1.4 Organization of Thesis

This thesis will consist of three separate parts that discuss utilizing photon-counting principals for secure information storage or imaging in photon-limited environments. In chapter 2, an optical security system for information storage is introduced that uses secure three-dimensional optical phase codes whose unique optical signature can be authenticated using the random-forest classifier. In chapter 3, a system for securing three-dimensional integral imaging displays using a quick-response encoded elemental image array will be discussed that combines photon-counting with traditional optical encryption methods for secure information storage. The second part of the dissertation will discuss integrated circuit authentication in photon-limited environments. More specifically, in chapter 4 a photon-counting model will be applied to an image of an integrated circuit (IC) captured using an x-ray to demonstrate that IC authentication can be achieved when few photons are
available in the scene. In the third part of this dissertation, imaging in low-light environments using 3D imaging with passive imaging sensors that operate in the visible spectrum will be discussed. In chapter 5, a three-dimensional (3D) imaging and object recognition in low light environments using real experimental data will be discussed. In chapter 6, an object recognition framework for 3D reconstructed images obtained from elemental images taken in low light conditions using Convolutional Neural Networks will be introduced.
3D integral imaging displays using a QR encoded elemental image array

Mobile devices are a ubiquitous technology and many researchers are trying to implement 3D displays on mobile devices for a variety of applications. We investigate a method to store compressed and encrypted elemental images (EI) used for three-dimensional integral imaging displays in multiple QR codes. This approach allows user friendly access, read out and 3D display with mobile devices. We first compress the elemental images and then use the double-random-phase encryption to encrypt the compressed image and store this information in multiple QR codes. The QR codes are then scanned using a commercial Smartphone to reveal the encrypted information which can be decrypted and decompressed. We also introduce an alternative scheme by applying photon-counting to each color channel of the EIs prior to the aforementioned compression and encryption scheme to generate sparsity and nonlinearity for improved compression and security. Experimental results are presented to demonstrate both 3D computational reconstruction and optical 3D integral imaging display with a Smartphone using elemental images from the QR codes. This work utilizing compressed QR encoded elemental images for secure integral imaging displays using mobile devices may enable secure 3D displays with mobile devices [63].
2.1 Introduction

Integral imaging, a promising 3D imaging technique, has been extensively investigated in disciplines as diverse as entertainment, medical sciences, robotics, manufacturing, and defense [16][22][25][26][55][56][57][64][65][66]. Photon-counting has been incorporated with integral imaging [16][22] to reconstruct a 3D image from a photon-starved environment. Recently, the double-random-phase encryption with photon-counting has been incorporated with integral imaging [22]. It was shown that integral imaging can provide a better security performance along with the ability to secure 3D information. The double-random-phase encryption (DRPE) technique [23][47][52][67][68][69] [70][71][72] is able to encrypt and decrypt an image. There have been numerous improvements to the encryption system to enhance security [20][47][69] in the system. One encryption scheme involves applying a photon-counting technique to the amplitude of the encrypted image generating a photon-limited [16][20][22][51][71]. This technique limits the number of photons arriving at a pixel. As a result, the decrypted image is sparse, noise-like and difficult to visually authenticate. In [71], a binary image was encrypted and compressed using the full-phase DRPE with photon counting and an iterative Huffman coding technique, respectively, and stored in a quick-response (QR) code. The QR code was then scanned by a QR reader built in a commercial Smartphones revealing the data. The scanned data was then decompressed and decrypted revealing a noise-like decrypted image that can be authenticated using non-linear correlation filters [73][74][75]. In this letter, we propose a method to securely store a RGB elemental image in a QR code by combining run-length encoding (RLE) [76] with the Huffman coding.
compression [77] scheme along with the double-random-phase encryption [23]. By storing data in a QR code [78][79][80], it is possible to reduce the amount of information needed to transfer the elemental image. We can store multiple elemental images in QR codes that can be used in 3D integral imaging (II) reconstruction [26]. In addition, we present an alternative scheme by first applying photon-counting on the elemental image followed by RLE, Huffman coding, and encryption to generate sparsity and nonlinearity due to Poisson transformation for improved compression and security. Both computational and optical reconstruction of the elemental images is presented.

2.2 Overview of Integral Imaging

Integral imaging (II) [16][22][25][55][56][57][64][65] [66] can produce a three-dimensional (3D) image of a scene by recording multiple two-dimensional (2D) images of different perspectives of the scene, known as elemental images. This is achieved by using various pickup methods such as an array of image sensors or a camera with a lenslet array, as shown in Fig. 2.2.1(a). Reconstruction of the 3D scene from the recorded elemental images can be performed optically, as shown in Fig. 2.2.1(b), or computationally. Optical reconstruction is performed by projecting the elemental images through a lenslet array to the image plane forming a 3D scene. Computational reconstruction uses a virtual pinhole array to inversely project the elemental images to the image plane to obtain the reconstructed scene as follows [16][25][65]:

\[ R(x, y, z) = \frac{1}{O(x, y)} \times \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} E_{kl} \left( x - k \frac{N_x \times p}{c_x \times M}, y - l \frac{N_y \times p}{c_y \times M} \right), \]  

where \( R(x,y,z) \) is the intensity of the reconstructed 3D image at depth \( z \), \( x \) and \( y \) are the index pixels, \( E_{kl} \) is the intensity of the \( k \)th column and the \( l \)th row of the elemental images, \( K \) and \( L \) are the total number of elemental images in the column and row, respectively, \( N_x \) and \( N_y \) are the total number of pixels for each elemental image, \( M \) is the magnification factor \( z/g \), \( g \) is the focal length, \( p \) is the pitch between image sensors, \( c_x, c_y \) are the spatial size of image sensor, and \( O(x,y) \) is the overlapping number matrix.

Fig. 2.2.1. (a) Pick up and (b) display stages of an integral imaging system.

### 2.3 Double random phase encryption with Run-length encoding a Huffman coding

Double-random-phase encryption (DRPE) is used to encrypt the input image [23]. 1-D notation will be used in explaining the encryption method. To implement the encryption, and denote the spatial and frequency domains, respectively, \( f(x) \) is the primary input
image, $n(x)$ and $b(v)$ are random noise that are uniformly distributed over the interval $[0, 1]$. The encrypted image is:

$$\psi(x) = \{ f(x) \times \exp[i2\pi n(x)] \} \ast h(x),$$

where $\ast$ denotes convolution and $\times$ denotes multiplication, $\exp[i2\pi n(x)]$ is a phase mask, and $h(x)$ is a function whose Fourier transform is $\exp[i2\pi b(v)]$.

The decryption process is the reverse of the encryption process. The Fourier transform of the DRPE image, $\psi(x)$, is multiplied by the complex conjugate image of the phase mask, $\exp[-i2\pi b(v)]$. The Fourier transform is taken once more. The intensity of the image can then be recorded which produces the decrypted image, $f(x)$, if the primary input image is real and positive.

Run-length encoding (RLE) [76] is a lossless compression technique that represents a series of repeated data by stating the number of times the data is repeated followed by the repeated integer. For example, $[8 \ 8 \ 8 \ 8 \ 8]$ is represented by $[5 \ 8]$. Huffman coding [77] is a dictionary based compression technique that can be combined with run-length encoding to allow for sufficiently small images to be stored in a QR code. By combining the two compression methods, less information is stored in the QR code improving the ability of the smartphone to scan the QR code [78].

Huffman coding [77] is a lossless compression technique that is implemented by first recording the frequency of each character of the data to be compressed. Since each character can be written using 8 bits, Huffman coding seeks to reduce the number of bits
in the data. A probability of occurrence is assigned to each symbol, as shown in Table 2.3.1. Once this is performed, a binary tree is created by first connecting the nodes of the two symbols with the lowest probability of occurrence. These probabilities are then added together and the nodes are combined. The probabilities are then rearranged and the nodes of the two least probable symbols are connected and their probabilities are added together. This process is repeated until all of the symbols have been combined as shown in Fig. 2.3.1(a). From there, the leftward branches are assigned a 0 and the rightward branches assigned 1. Starting from the bottom node of 2.3.1(b), C5, each character is assigned a binary representation, or Huffman code, corresponding to the tree, which is shown in Table 2.3.1. Note that the maximum binary representation is 8 bits.

Table 2.3.1. Frequency of symbols along with their respective probability of occurrence and Huffman code.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Frequency</th>
<th>Probability</th>
<th>Huffman Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>20</td>
<td>0.3571</td>
<td>11</td>
</tr>
<tr>
<td>S2</td>
<td>15</td>
<td>0.2679</td>
<td>10</td>
</tr>
<tr>
<td>S3</td>
<td>11</td>
<td>0.1964</td>
<td>01</td>
</tr>
<tr>
<td>S4</td>
<td>6</td>
<td>0.1071</td>
<td>001</td>
</tr>
<tr>
<td>S5</td>
<td>3</td>
<td>0.0536</td>
<td>0000</td>
</tr>
<tr>
<td>S6</td>
<td>1</td>
<td>0.0179</td>
<td>0001</td>
</tr>
</tbody>
</table>
Fig. 2.3.1. Building the binary tree used in Huffman coding. (a) Generating the binary tree with root nodes and (b) assigning labels to each branch of the binary tree.

Using Huffman coding, we reduced the amount of information, shown in Table 2.3.1, from 448 bits to 126 bits. By providing a dictionary that allows for the reconstruction of
the Huffman tree, the compressed data can be decompressed and fully recovered by reading the binary tree beginning with the top of the tree.

We compress each channel of the RGB image using RLE followed by Huffman coding. The double-random-phase encryption is used to encrypt each channel of the compressed image. Due to the ciphered images being complex, both the real and imaginary parts must be retained. As a result, both are stored in separate QR codes which are generated using the ZXing project [79]. In addition, the encrypted images are rounded to one decimal place to minimize the amount of necessary information stored in the image. Each ciphered channel of the RGB image can then be stored in separate QR codes. Moreover, we can use the same encryption keys for each color channel. For a single RGB image, 6 QR codes are used in our experiments. The EI must be sufficiently small so that the compressed and encrypted data can be stored in a single QR code.

Once the QR code has been scanned, the data can be successfully decrypted if the decryption keys are known. In addition, the image can be decompressed if the dictionary associated with the Huffman code compression is known; no dictionary is required to decompress the RLE algorithm.

Figure 2.3.2(a) depicts a 19 x 19 pixels input RGB image. For brevity, the red channel of the RGB image is explained. The process can be repeated for the other channels. Figure 2.3.2(b) shows a QR code that stores the real part of the ciphered input image while Figure 2.3.1(c) shows the imaginary part of the ciphered input image. Figures 2.3.2(d) and Fig. 2.3.2(e) display the compressed and encrypted data stored in the QR codes using the iPhone SCAN application. After decompressing and decrypting the necessary data, the three color
channels of the decrypted image are combined to form the decrypted RGB image as shown in Fig. 2.3.2(f).

Fig. 2.3.2. (a) 19 x 19 pixels RGB image; (b) and (c) display the QR codes containing the real and imaginary information of for the red color channel from the compressed and ciphered image shown in (a), respectively; (d) and (e) depict a scanned QR code using the iPhone SCAN application revealing the real and imaginary information from the compressed and ciphered image, respectively, for the red color channel of the image shown in (a); (f) shows the decrypted and decompressed image.
To determine the degree of degradation (if any) due to the double-random-phase encryption combined with the RLE and Huffman coding compression schemes, the mean squared error (MSE) is calculated as:

$$MSE = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left[ f_{\text{decrypt}}(x_n, y_m) - f(x_n, y_m) \right]^2,$$

where $n$ and $m$ are pixels in the $x$ and $y$ directions, respectively, $N$ and $M$ represent the total number of pixels in the $x$ and $y$ direction, respectively, $f_{\text{decrypt}}(x_n, y_m)$ is the decrypted and decompressed RGB image and $f(x_n, y_m)$ is the original RGB image.

Since both Huffman coding and RLE are lossless, and no information is lost during the DRPE, the MSE calculated between the images shown in Figs. 2.3.2(a) and Fig. 2.3.2(f) is 0.

It is possible to store multiple encrypted and compressed images in multiple QR codes as shown in Fig. 2.3.3. Each QR code can be scanned allowing for every elemental image to be recovered. Thus, this approach can have applications in integral imaging reconstruction by storing the elemental images in QR codes.
Fig. 2.3.3. Storing multiple RGB encrypted elemental images inside of multiple QR codes.
2.4 Experimental results

A 3D integral imaging experiment was conducted for computational integral imaging reconstruction. Each elemental image used identical encryption and decryption keys. Moreover, identical encryption keys were used in the spatial and frequency domains. Four 19 x 19 pixels RGB elemental images, shown in Fig. 2.4.1(a) were generated using 3Ds Max. The 3D scene consists of a greed letter “3” and red letter “D” located at 65 mm and 115 mm away from a lenslet array. These images were compressed using RLE followed by Huffman Coding compression and then encrypted with the DRPE. These elemental images were decrypted and decompressed revealing the elemental images.

![Elemental Images](image1)

**Fig. 2.4.1.** 19 x 19 pixels RGB (a) elemental images which were compressed using run-length encoding and Huffman coding followed by encryption using the double-random-phase encryption; computational 3D integral imaging reconstructions with 4 EIs at a distance (range) of (b) at 65 mm focused on “3” and (c) at 115 mm focused on “D”.

37
Computational reconstruction results using Eq. 2.2.1 were obtained. Figure 2.4.1(b) depicts the computational reconstruction at 65 mm focused on “3” while Fig. 2.4.1(c) depicts computational results at 115 mm focused on “D”.

Storing a large quantity of elemental images in multiple QR codes may prove burdensome and inconvenient. Regardless, with improving camera resolution and hardware of Smartphones, it will be possible in the future to store more information in QR codes. Thus, it may be feasible to store a prodigious number of encrypted and compressed elemental images in QR codes. An optical experiment was carried out using the proposed compression and encryption scheme to secure elemental images for 3D display. Using 3Ds Max, 54 x 96 RGB elemental images, similar to the elemental images shown in Fig. 2.4.1 (a), were used which were each 19 x 19 pixels. As before, each elemental image used identical encryption and decryption keys. These images were then compressed using RLE followed by Huffman Coding compression and encrypted with the DRPE. Each elemental image had identical encryption keys for both the spatial and frequency domains. These elemental images were then decrypted and decompressed revealing the elemental images. An HTC One Smartphone was used to display the optical reconstruction. Fig. 2.4.2 depicts the experimental setup for optical reconstruction using an HTC One Smartphone and lenslet array. Table 2.4.1 below presents the experimental parameters.
Fig. 2.4.2. Optical 3D integral imaging display setup using a Smartphone and a lenslet array.

Table 2.4.1. Specifications of 3D display setup.

<table>
<thead>
<tr>
<th></th>
<th>Resolution</th>
<th>1080(V) x 1920(H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTC one smart phone display panel</td>
<td>Pixel size</td>
<td>54.1 um</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>0.985 mm</td>
</tr>
<tr>
<td>Lenslet array</td>
<td>Focal length</td>
<td>3.3 mm</td>
</tr>
<tr>
<td></td>
<td>Numbers of lenses</td>
<td>54 (V) x 96(H)</td>
</tr>
</tbody>
</table>
Fig. 2.4.3(a) depicts the 3D optical reconstruction using the primary elemental images while Fig. 2.4.3(b) depicts the 3D optical reconstruction using the decompressed and decrypted elemental images. The reconstruction results using our proposed method are as good as the conventional method. There is no visible loss of information due to the compression, encryption, and QR encoding.

Storing data in QR codes holds many practical advantages due to not requiring any specialized hardware to scan the QR code. Moreover, Smartphones are a ubiquitous technology. The proposed technique of storing QR codes can potentially be modified by creating a video barcode as described in [16] to allow a faster user experience when scanning multiple QR code.

Fig. 2.4.3 3D displays results using Smartphone. (a) 3D Optical reconstruction with integral imaging using the primary elemental images. (b) Optical 3D reconstruction using the decrypted and decompressed elemental images obtained from the compressed double-random-phase encryption elemental images in QR codes.
2.5 Compression with Photon-Counting

We now introduce an alternative compression and security approach. We first apply photon-counting \cite{16,20,22,71} to each channel of an RGB elemental image, $f(x,y)$. Photon-counting is described by a statistical method for a limited number of photons that arrive at a pixel. Thus, the lower the total number of photons in the image, the fewer the number of photons that arrive at a pixel. By reducing the number of photons, it is possible to have more sparsity improving the run-length encoding compression scheme used prior to Huffman coding. Also, the nonlinear transformation introduced by photon counting improves the security. Photon-counting can be modeled as a Poisson distribution:

$$P(l_j; \lambda_j) = \frac{\lambda_j^{l_j} e^{-\lambda_j}}{l_j!}, \text{for } \lambda_j > 0, l_j \in \{0,1,2,...\},$$

where $l_j$ is the number of photons detected at pixel $j$ and $\lambda_j$ is the Poisson parameter defined as $N_p x_j$ where $N_p$ is the number of photons in the scene and $x_j$ is the normalized irradiance at pixel $j$ such that $\sum_{j=1}^{J} x_j = 1$ with $J$ being the total number of pixels. Moreover, the normalized irradiance is defined as $|f(x_m,y_n)|/\sum_{m=1}^{M}\sum_{n=1}^{N}|f(x_m,y_n)|$, where $m$ and $n$ denote pixels in the $x$ and $y$ direction, respectively, $M$ and $N$ represent the total number of pixels in the $x$ and $y$ direction, respectively.

A 54 x 96 RGB elemental image array with each elemental image being 19 x 19 pixels, similar to the elemental image depicted in Fig. 2.5.1(a), was used for the 3D photon-counting imaging compression and security. The object chosen was a color gradient teapot generated by 3Ds Max.
Fig. 2.5.1. 3D photon counting integral imaging compression and security experiments. A 54 x 96 elemental image (EI) array consisting of 19 x 19 pixels RGB EIs was used. (a) shows an EI while (b) depicts the corresponding photon-limited elemental image using about 8 photons per pixel on each color channel of the EI. 3D optical reconstruction after decryption and decompression is shown using (c) the original EIs and (d) when photon counting on the EIs was used.

Photon-counting was applied to each channel of the RGB EI as shown in Fig. 2.5.1(b) using 3000 photons or about 8 photons per pixel on each channel of the 19 x 19 RGB image. Run-length encoding is then applied followed by Huffman Coding and the DRPE. The image was then decrypted and decompressed. Optical reconstruction using an HTC phone
[see Table 2.4.1] was used to reconstruct the image [See Eq. 2.2.1]. Fig. 2.5.1(c) depicts the optical reconstruction showing the original EI while Fig. 2.5.1(d) depicts the optical reconstruction using the photon-limited EI. The mean squared error [See Eq. 2.2.3] was calculated to determine the degree of degradation for each individual color channel. For comparison, both images were normalized to [0, 255]. The MSE for the red color channel was calculated as 62.94, the green channel was 1.07, and the blue channel was 19.73. Thus, there was some degree of degradation; however, visually the 3D optical reconstruction of the photon-counted image can still be discerned.

A comparison of the compression with and without photon-counting was conducted. It was found that for a sufficiently lower number of photons, the photon-counting method helped to improve the compression scheme. Using the EIs in Fig. 2.5.1(c) and Fig. 2.5.1(d), the total length of the Huffman code following run-length encoding was compared. It was found that the red channel was 21.7% shorter and the blue channel was 18.53% shorter; there were no green channel components in the test image. These results show that photon-counting may allow for larger EIs to be stored in QR codes after compression and encryption.

2.6 Conclusion

We have presented a method and experiments for secure integral imaging as a potential approach to store elemental images in QR codes for 3D displays. The compressed and encrypted elemental images using double-random-phase encryption can be stored in multiple QR codes to be scanned by a commercial Smartphone revealing the encrypted and
compressed elemental images. This data can then be decrypted and decompressed to obtain the elemental images for secure 3D integral imaging displays. We also present an alternative compression and security scheme by applying photon-counting to the input image followed by compression and encryption. By using photon-counting, the compression and security system improve at a cost of diminished image quality. Future work may include alternative algorithms for encrypting the elemental images, compressing the elemental images, and video scanning of the QR codes.
Chapter 3

Security authentication with a Three-Dimensional Optical Phase Code Using Random Forest Classifier

In this chapter, an object with a unique three-dimensional (3D) optical phase mask attached is analyzed for security and authentication. These 3D optical phase masks are more difficult to duplicate or to have a mathematical formulation compared with 2D masks, and thus have improved security capabilities. A quick response code was modulated using a random 3D optical phase mask generating a 3D optical phase code (OPC). Due to the scattering of light through the 3D OPC, a unique speckle pattern based on the materials and structure in the 3D optical phase mask is generated and recorded on a CCD device. Feature extraction is performed by calculating the mean, variance, skewness, kurtosis, and entropy for each recorded speckle pattern. The random forest classifier is used for authentication. Optical experiments demonstrate the feasibility of the authentication scheme [81].

3.1 Introduction

Optical information security has sought to ensure the secure transmission of an image to a recipient. This area of research includes image encryption [20][23][47][68][69][82][83][84][85][86][87][88][89][90], authentication [19][21][71][91][92][93][94][95][96],
and compression or secure storage [97][98][99]. Authenticating sensitive information is critical to discovering tampering caused by a miscreant. Methodologies for image authentication includes both optical [91][92][93][94] and simulated [19][21] authentication schemes.

Recently, authentication schemes have been investigated using optically tagged security codes [71][94][95]. In these authentication schemes, an object is optically tagged using a phase mask. In [71], these phase masks were as simple as Scotch tape. In [94], optical codes based on thin-film technology were produced for security applications. These structures generate distinctive polarimetric information that can be utilized to authenticate the message encoded. In [95], more complex phase masks consisted of embedding nanoparticle structures such as gold in an object. An optical set up was then used to authenticate objects containing the phase mask by illuminating the object with a laser diode. The polarimetric information from the object was recorded and used for authentication. In [91], an authentication scheme using a three-dimensional (3D) phase object was created by illuminating a 3D phase object with two different wavelengths and recording the resulting speckle pattern with a CCD device. These speckle patterns were then correlated with authentic speckle patterns from a database to verify the veracity of the 3D phase object.

In this work, we propose a 3D optical phase code (OPC) by encoding a quick response (QR) code with a 3D optical phase mask. An advantage of using a 3D optical phase mask compared with a 2D mask is its difficulty in being duplicated by simple examination of the optical phase mask or the resulting speckle pattern. The 3D optical phase mask may be generated in a variety of methods. In our experiments, it consists of a combination of glass
slides and diffuser material. A 445 nm wavelength blue laser diode is transmitted through the 3D OPC generating a unique speckle pattern that is recorded on a CCD. From the recorded speckle pattern, the mean, variance, skewness, kurtosis, and entropy is computed. The random forest classifier is then used to authenticate the phase masks.

3.2 3D Optical Phase Code Design and Feature Selection

Three-dimensional OPCs were created as shown in Fig. 3.2.1. A 4 mm × 4 mm QR code was first printed on transparency paper. A 3D optical phase mask was then placed on the QR code. In the experiment, three phase mask configurations were used. As shown in Fig. 3.2.1(a), a glass slide and diffuser paper were placed on a QR code; we denote this configuration as 3D OPC A. Figure 3.2.1(b) depicts a glass slide and diffuser paper along with an additional glass slide and diffuser paper placed on a QR code; we denote this configuration as 3D OPC B. Lastly, Fig. 3.2.1(c) depicts a glass slide and diffuser paper along with an additional glass slide, diffuser paper, and glass slide placed on a QR code, generating 3D OPC C. We note that phase codes were held together by Scotch tape; however we verified that the tape was placed sufficiently far from the QR code. Thus, when illuminated by a laser source, the laser would not be transmitted through the tape. Fig. 3.2.2(a) shows the experimental 4 mm × 4 mm QR code used while Fig. 3.2.2(b) depicts 3D OPC A. The 3D OPCs generate a highly nonlinear scattering of light due to being an inhomogeneous material [99][100][101][102]. In addition, the light transmitted through the 3D OPCs cannot be easily described with conventional wave propagation models [103]. This highly nonlinear light propagation, though difficult to model, can be used as a unique
phase mask. Having this complex phase mask is ideal to serve as an optical tag to create a unique signature for an object.

Fig. 3.2.1. Workflow for developing the 3D optical phase code for (a) 3D optical phase code A, (b) 3D optical phase code B, and (c) 3D optical phase code C.
Fig. 3.2.2. (a) Experimental 4 mm × 4 mm QR code and a picture of (b) 3D optical phase code A, which consists of a QR code with an optical phase mask consisting of a glass slide and diffuser paper.

Once the 3D optical phase codes were developed, an optical experiment was carried out as shown in Fig. 3.2.3. A 3D OPC was placed on a translation stage. A blue laser diode having a wavelength of 445 nm was transmitted through first a polarizer to lower the intensity followed by a lens to expand the light. The light was then transmitted through the 3D OPC. A Canon EOS 600D with a CCD sensor size of 14.9 mm (v) × 22.3 mm (w) was used to record the resulting speckle pattern which was 2784 (v) × 1856 (w) pixels. Twenty speckle patterns were recorded for when the 3D OPC was 30 mm, 70 mm, 110 mm, and 150 mm from the CCD sensor. Note that the statistical properties of a speckle, which is a
A nonstationary process, can be influenced by environmental effects including vibrations [103].

Fig. 3.2.3. Optical experimental setup. A 455 nm blue laser diode is transmitted through a polarizer and lens. The laser is then transmitted through the QR code which has a 3D optical phase mask placed on it. A CCD sensor, a distance $d$ away from the QR code, records the speckle pattern.

A CCD is an intensity recording device and the recorded speckle patterns can be approximated as a statistical distribution. It can be shown that the statistical pattern can be approximated as a Gamma distribution [104]:

$$\Gamma(I) = \left( \frac{I}{\bar{I}} \right) \frac{\Gamma(n_v)}{\Gamma(n_v)} \exp\left( -\frac{I}{\bar{I}} \right)$$

where $I$, $\bar{I}$, and $\sigma$ are the intensity data points, its average and the corresponding standard deviation, respectively.

An example of the speckle patterns captured is shown in Fig. 3.2.4 for a distance, $d$, of 110 mm from the CCD using 3D OPCs A, B and C, respectively, along with their corresponding histograms. We note that the color map was adjusted to improve the
visualization of the speckle. Using Eq. 3.2.1, a gamma distribution was also fitted to the histograms. The images were normalized to lie between the interval [0, 1].

Fig. 3.2.4. (a,b,c) The speckle patterns obtained using 3D optical phase codes A, B, and C, respectively. The 3D optical phase codes are a distance of 110 mm from the CCD sensor. The corresponding histograms and a fitted Gamma distribution are also shown.
In our proposed authentication scheme, we extract statistical features from each speckle pattern to be used for classification. The chosen features were: mean, variance, skewness, kurtosis, and entropy. The skewness and kurtosis can be used to examine the location and variability of a distribution, respectively [105]. The skewness measures the third moment of a distribution and measures the symmetry of a distribution. Since the speckle patterns are unimodal, negative skewness values mean the left tail is longer than the right tail of the distribution. Moreover, a positive skewness indicates the right tail is longer than the left tail. The kurtosis measures the fourth moment of a distribution and describes the curvature of the distribution. This metric measures how much the data is peaked or flat relative to a standard normal distribution. For feature extraction, the unbiased skewness and kurtosis was used [106]. Lastly, the entropy [107][108] measures the average uncertainty, or variability, of an image. The minimum uncertainty occurs at an entropy of 0. The mean, unbiased variance, skewness and kurtosis along with the entropy are defined as:

\[
\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad 3.2.2
\]

\[
\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2, \quad 3.2.3
\]

\[
\hat{s} = \sqrt{\frac{N(N-1)}{N-2}} \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^3 \left( \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^2 \right)^{-\frac{3}{2}}, \quad 3.2.4
\]
\[
\hat{k} = \frac{N - 1}{(N - 2)(N - 3)} \left[ \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^4 \right]^{\frac{1}{4}} \left[ \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^4 \right]^{-\frac{3}{4}} + 3,
\]

where \(\hat{\mu}\) is the sample mean, \(\hat{\sigma}^2\) is the unbiased sample variance, \(\hat{s}\) is the unbiased skewness, \(\hat{k}\) is the unbiased kurtosis, \(H\) is the entropy, \(p(x_i)\) denotes the probability mass function of \(x_i\) found by using the relative frequency distribution [109], and \(N\) is the total number of pixels.

By calculating the mean, variance, skewness, kurtosis, and entropy, pixel intensities no longer need to be stored; the only information needed are the five feature values and the classification model. An example of the features extracted at distances of 70 mm and 110 mm are shown in Table 3.2.1.

<table>
<thead>
<tr>
<th>Table 3.2.1. Example of mean, variance, skewness, kurtosis, and entropy calculated for recorded speckle patterns.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D Optical Phase Code</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>(d=70) mm</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(d=110) mm</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

### 3.3 3D Optical Phase Code Design and Feature Selection

We chose to use the random forest (RF) classifier [110][112] for the classification model. This supervised, non-parametric classifier has reduced variance and is robust to
overfitting. In essence, the random forest combines the outputs of many independent decision trees, which is a type of binary tree that contains nodes, branches, and leaves. A “vote” is made by averaging the final results of each decision and the majority vote indicates the predicted class of an input.

The splits are based on the Gini’s diversity index (GDI) [110]. This metric measures the node impurity. The lower the GDI, the better the split. For a data set $S$ at node $M$, the GDI is defined as:

$$\text{GDI}(S) = 1 - \sum_{k=1}^{K} \left[ p_k(x) \right]^2,$$

where $K$ is the number of predefined classes, $p_k(x)$ is the relative frequency [36] of class $k$ at node $M$ defined as:

$$p_k(x) = \frac{1}{N} \sum_{s_k \in S} I(x),$$

where $s_k$ is the number of data points in class $k$, $I$ is the indicator function, and $N$ is the total number of data points in $S$ at node $M$.

The number of features selected, at random, is also calculated for each split. The advantage of using a limited number of features is that it helps to decorrelate trees since strong predictors will not appear in every tree. The minimum node size is one and the number of features at each node is:

$$m = \left\lfloor \sqrt{v} \right\rfloor,$$

where $v$ is the number of features and $\left\lfloor \cdot \right\rfloor$ denotes the floor operator.
Now using the $m$ features at each split, pick the best variable/split point [see Eq. 3.3.1] and split the node into two daughter nodes. The process is repeated until a node has one class, which corresponds to a GDI of zero.

The random forest classifier creates multiple decision trees by using separate bootstrapped [110] samples from the data for each tree. Bootstrapping also helps to ensure that the trees developed are not correlated. Data not used to train a tree is known as being “out-of-bag”. This data is used to evaluate the performance of the classifier.

Bootstrapping is sampling a data set of independent data with replacement. For each bootstrap, $z^{*,b}$, bootstrapping is defined as:

$$z^{*,b} = (x_1^*, x_2^*, ..., x_N^*), \quad b = 1, 2, \ldots B,$$

where $b$ is the $b$th bootstrap data set for $N$ data points and $B$ is the total number of bootstrapped samples.

After forming the binary trees, a “vote” from each tree determining the class of the “out-of-bag” data is computed by taking into account only data samples that were not used in any decision trees. To form a decision, we let $\hat{C}_b(x)$ be the class prediction of the $b$th random forest tree, then the final classification prediction is:

$$C_{final}(x) = \text{majority vote} \left\{ \hat{C}_b(x) \right\}^B,$$

where $\text{majority vote}$ is the class that has the most “votes” from the random forest consisting of $B$ trees.
3.4 3D Optical Phase Code Authentication

Twenty speckle patterns were recorded from 3D optical phase code A, B, and C at distances of 30 mm, 70 mm, 110 mm, and 150 mm from the CCD sensor. In total, there were 12 classes, as described in Table 3.4.1, and 120 true class images. Feature extraction was then performed [see Eqs. 3.2.1-3.2.5] and the random forest classifier was used to authenticate these speckles. Fig. 3.4.1 below depicts an overview of the proposed authentication system. The random forest classifier was trained using ten speckles from each class while the other ten were used for testing. The random forest classifier training model was evaluated by calculating the out-of-bag error for the random forest using 100 trees, shown in Fig. 3.4.2. After 100 trees, the percent error was about 0.67%.

![Diagram](image)

Fig. 3.4.1. Overview of the 3D optical phase code authentication system.

Testing data was then inputted into the model using the other ten recorded speckles from each class. We note that the classifier automatically places the data into a class. To determine the reliability of a measurement, we also observe the score that the random forest calculates. The score is the percentage of “votes” from each binary tree to each class. There
was a 100% correct classification rate for each class. In addition, the number of votes to the correct class in the proposed classification scheme was on average 99.953%. Thus, the proposed classification scheme may be used to authenticate a unique 3D optical phase mask.

![Graph showing out-of-bag classification error using 100 trees which converges to about 0.0067.](image)

**Fig. 3.4.2.** Out-of-bag classification error using 100 trees which converges to about 0.0067.

<table>
<thead>
<tr>
<th>3D OPC A</th>
<th>3D OPC B</th>
<th>3D OPC C</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 mm</td>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td>70 mm</td>
<td>Class 4</td>
<td>Class 5</td>
</tr>
<tr>
<td>110 mm</td>
<td>Class 7</td>
<td>Class 8</td>
</tr>
<tr>
<td>150 mm</td>
<td>Class 10</td>
<td>Class 11</td>
</tr>
</tbody>
</table>
A test was also conducted to determine the performance of the classifier to speckles from 3D OPCs that do not fall into any classes, which we consider false class data. Fifty speckles were captured from different configurations of 3D OPCs, which were constructed using a process similar to those constructed in Fig. 3.2.1, and placed at arbitrary distances from the CCD sensor. Features were extracted from the recorded speckle pattern and inputted into the random forest classifier. If a speckle pattern did not have a 95% vote, it was assumed that the classifier was unable to decide the class of the speckle. An example of a false class speckle pattern is shown in Fig. 3.4.3. The false class pattern was created by placing 3D optical phase code C 100 mm from the CCD and recording the resulting speckle pattern. For this particular false class speckle, the classifier determined that the pattern belonged to class 8. Moreover, the score received only 84% of the votes thus we could not determine the veracity of the phase mask and conclude that it was not authentic.

Table 3.4.2. Classification results for the 12 class system for when 120 test images were used and 50 false class images

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td>True</td>
<td>120</td>
</tr>
<tr>
<td>False</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>False</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

Table 3.4.2. shows the confusion matrix of the classifier accounting for both true and false class data.

The accuracy of the classifier when the 3D OPC was shifted from their original positions was also evaluated. At distances of 70 mm, 110 mm, and 150 mm, 3D optical
phase code A was displaced up to +/- 10 mm from the original position by increments of 1 mm. As the 3D optical phase code was further from the CCD device, the classifier was less sensitive to displacement errors as shown in Fig. 3.4.4. We note that the classifier was able to correctly classify speckles for all displacements about 110 mm and 150 mm; however, at 70 mm there were misclassifications at – 9 mm and – 10 mm. Thus, a user must be mindful the distance the 3D optical phase code is from the CCD sensor.

Fig. 3.4.3. A false class speckle pattern obtained by placing 3D optical phase code C 100 mm from the CCD sensor.

Fig. 3.4.4. Effect of displacement of 3D optical phase code A on the class confidence score output from the random forest classifier for distances 70 mm, 110 mm, and 150 mm from the CCD sensor.
3.5 Conclusion

In conclusion, we present an authentication scheme using a transparent QR code containing a 3D optical phase mask to generate a 3D optical phase code (OPC). An advantage of a 3D optical phase mask over a 2D is that it is difficult to mathematically characterize a 3D code made of randomly scattering medium and/or to duplicate it physically. An optical authentication system was designed using 3 separate 3D OPCs which were placed 30 mm, 70 mm, 110 mm, and 150 mm from a CCD sensor. A 445 nm blue laser diode illuminated the 3D OPCs at each distance to generate a unique speckle pattern that was captured by the CCD. Feature extraction was then performed on the speckle pattern by calculating the mean, variance, skewness, kurtosis, and entropy. A multiclass random forest classifier was used to classify the recorded speckles at each distance. A 100% accuracy rate was achieved. Thus, we have shown we can use mean, variance, skewness, kurtosis and entropy of a speckle image combined with the random forest classifier to determine the authenticity of a 3D OPC. Overall, it is difficult to reproduce the 3D optical code from either the resulting speckle pattern or visual inspection. As a result, we can use this 3D optical phase code system to authenticate an object.
Part II: Integrated Circuit Authentication under Photon-Limited Conditions
Chapter 4

Integrated Circuit Authentication using Photon-limited X-ray Microscopy

A counterfeit integrated circuit (IC) may contain subtle changes to its circuit configuration. These changes may be observed when imaged using an x-ray; however, the energy from the x-ray can potentially damage the IC. We investigate a technique to authenticate ICs under photon-limited x-ray imaging. We model an x-ray image with lower energy by generating a photon-limited image from a real x-ray image using a weighted photon-counting method. Feature extraction is performed on the image using the speeded up robust features (SURF) algorithm. The IC is then authenticated by comparing the SURF features to a database of SURF features from authentic and counterfeit ICs. Experimental results with real and counterfeit ICs using an x-ray microscope demonstrate that an IC image captured using orders of magnitudes lower energy x-rays can still can be correctly authenticated. To the best of our knowledge, this is the first report on using photon-counting x-ray imaging to prevent potential damage when authenticating ICs [111].

4.1 Introduction

Integrated circuits (IC) are found in all electronic items ranging from computers and cell phones to air planes and radar systems. It is critical that these components are not
compromised, perform as expected, are not counterfeits, and can be authenticated. Attackers may modify the existing structure of the IC or insert hardware Trojan horses (HTH) into the IC to disable, destroy, or leak information that is [113][114][115]. Various methods are used to authenticate ICs such as measuring the power and current measurements of the ICs at different junctions [115] along with timing-based methods that measures the delay between a circuits’ input and output [116]. Visual or automated inspection of an IC can be performed using x-ray microscopy [117][118]. In [118], a real time classifier was developed for x-ray images using a stacked auto encoder for dimensional reduction followed by a deep neural network for classification. However, the energy of the x-ray used to capture these images may damage some IC chips [119][120][121].

We propose a method to authenticate ICs imaged using orders of magnitudes lower energy x-rays. Our objective is to investigate if ICs can be inspected and authenticated at much lower x-ray energies to prevent potential damage to the IC. In our experiments, both real authentic and counterfeit ICs are imaged using an x-ray microscope (Xradia). To investigate the IC inspection under very low x-ray energy, the x-ray microscope image must undergo photon-counting [16][29][46][103][122]. Photon-counting generates a photon-limited image that simulates the effect of x-ray imaging with substantially reduced energy. We perform this based on a weighted photon-counting method. The speeded up robust features (SURF)[123] from the photon-limited image are computed. SURF features from this image are then compared with a database of SURF features from authentic and counterfeit ICs. The shortest Euclidean distance is found between each SURF feature from
the test x-ray image with the SURF features in the database. The test image is then assumed to belong to the same class as the IC with which it has the most SURF features in common.

4.2 Photon-limited X-ray images

In our experiments, x-ray images of both authentic and counterfeit ICs were captured using the Xradia microscope, depicted in Fig. 4.2.1(a). The x-ray was used to image an IC at a voltage of 80 kV and power of 7W, which is shown in Fig. 4.2.1(b). Imaging at these parameters can potentially damage the IC. The voltage and power of the x-ray machine can be reduced to about 40 kV and 1W, respectively, and still able to image an IC as shown in Fig. 4.2.1(c). This reduction in power can help diminish the amount of damage caused by the x-ray. Unfortunately, the voltage and power of the x-ray machine we used cannot be further reduced to obtain photon-counting x-ray images. As a result, the reduction in voltage and power was modeled using photon-counting imaging allowing us to substantially reduce the energy of the x-ray [16][29][46][122][103]. Photon-counting was applied to the x-ray images acquired at 80 kV and power of 7 W to adjust the number of photons in the image. A model for simulating photon-limited images (PC-NW) is [16]:

\[
C(x, y) \sim \text{Poisson}(\lambda(x, y)), \quad \lambda(x, y) > 0, \quad \lambda(x, y) = \frac{N_p I(x, y)}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} I(x, y)},
\]

where \(C(x, y)\) is the number of photons detected at pixel \((x, y)\), \(\lambda(x, y)\) is the Poisson parameters, and \(\text{Poisson}(\lambda)\) is \((\lambda^k/k!)exp(-\lambda)\). Moreover, \(N_p\) is the expected number of
photons in the scene, \( N_x \) and \( N_y \) are the total number of pixels in the \( x \) and \( y \) directions, respectively, \( I(x, y) \) is the irradiance at coordinates \((x, y)\), and \( \sum_{x,y} \lambda(x, y) = N_p \).

Figure 4.2.1. (a) X-ray microscope used in the experiments (interior view). (b) An integrated circuit (IC) x-ray image obtained using 80 kV and 7 W with x-ray microscope. (c) The IC x-ray image obtained using 40 kV and 1 W.
Figure 4.2.2 shows an example of a real, authentic IC obtained using the x-ray microscope. Fig. 4.2.2(a) depicts the original 256 (W) × 512 (V) pixel x-ray image obtained using the x-ray microscope. A photon-limited x-ray image of the IC using 1.5e6 photons, or 11.4 photons/pixel, is shown in Fig. 4.2.2 (b) using the PC-NW model.

In Fig. 4.2.2 (b), we note that it is difficult to see the wires coming from the die of the IC to the rest of the circuit. To preserve the information about the wires at low radiation levels, we apply photon-counting based on a weighted photon-counting method (PC-W). To emphasize the wires, we simply multiply the number of photons detected by weights to give precedence to areas with more photons. We generate the weights based on the gamma distribution, which was shown to perform well:

\[
w(x, y) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \lambda(x, y) \right)^{\alpha-1} e^{-\beta(\lambda(x, y))},
\]

where we define \( \beta = \mu/\sigma^2 \) and \( \alpha = \left[ \frac{\mu^2}{\sigma^2} \right] \), \( \mu \) is the mean of \( \lambda(x, y) \) and \( \sigma^2 \) is the variance of \( \lambda(x, y) \), which is similar to acquiring the parameter estimation used for Bayesian estimation[124]. The final distribution becomes:

\[
C_{\text{weight}}(x, y) \sim \text{Poisson}\left( \lambda(x, y) \right) \tilde{w}(x, y),
\]

where \( \tilde{w}(x, y) \) are the normalized weights such that \( \sum_{x, y} \tilde{w}(x, y) = 1 \).

Figure 4.2.2(c) depicts the photon-limited image of Fig. 4.2.2(a) using the PC-W model and 11.4 photons/pixel. Figure 4.2.3 depicts counterfeit circuits of Fig. 4.2.2(a) along with their corresponding photon-limited images based on the PC-W model. In Fig.
4.2.3(a), the wires on the IC have been tampered with while in Fig. 4.2.3(c) the die has been changed. Figs. 4.2.3(d) and 4.2.3(d) show their corresponding photon-limited IC image using 11.4 photons/pixel using the PC-W model, respectively.

Figure 4.2.2. Example of a real authentic 256 (W) x 512 (V) pixels (a) integrated circuit. A photon-limited IC is generated based on the (b) photon-counting model with no weights and (c) photon-counting model using the weighted photon-counting method with $1.5 \times 10^6$ photons in the image, or 11.4 photons/pixel.
Figure 4.2.3. Example of a 256 (W) x 512 (V) pixel counterfeit images where (a) the wires have been tampered along and (c) the die has been changed. (b) and (d) depict their corresponding photon-limited IC image, respectively, using 1.5e6 photons, or 11.4 photons/pixel based on the weighted photon-counting method.
4.3 Photon-Counting authentication

The classification method chosen was feature matching using the speeded up robust feature (SURF)[123] algorithm based on the MATLAB image search using point features. Interest points are first found. These are distinct points of an image such as corners, blobs, and T-junctions. These features are also common amongst multiple images of a class. The descriptor is then created which is a feature vector that can describe the interest points. It is robust to noise and geometric and photometric deformations. The final step is to match the feature vectors with the feature vectors of images in a database associated with authentic or counterfeit ICs. This is performed by finding the distance between vectors such as by measuring the Euclidean distance. The SURF algorithm is a scale and rotation invariant detector based on Hessian matrix-based measurements and distribution based descriptor. Before the SURF feature is computed, an integral image [123] is first generated defined as:

\[
I_z(x, y) = \sum_{k \geq x, l \geq y} I(k, l),
\]

where \(I_z(x, y)\) denotes the sum of all the pixels to the left and above coordinate \((x, y)\).

The advantage of forming an integral image is the minimal operations needed to calculate the area of a rectangle, as shown in Fig. 4.3.1, by using the following equation:

\[
I_{\text{sum}}(x, y) = I_A(x, y) + I_D(x, y) - I_C(x, y) - I_B(x, y),
\]
where $I_A, I_B, I_C$ and $I_D$ is the sum of pixels in the image to the left and above points A, B, C, and D, respectively, as depicted in Fig. 4.3.1. Note that the integral image is padded with zeros on the top and left most side of the image. In addition, the numeric value to the diagonal left of the letters represent the sum of the pixels to the left and above that letter.

Fig.4.3.1 Constructing an integral image.
To find the interest points, the Hessian matrix, \( H(X,s) \), is computed where \( s \) is the standard deviation and determines the scale and \( X = (x,y) \) in the image \( I \). This matrix serves as the blob detection, or method at finding distinctive regions of an image such as bright spots. The hessian matrix is defined as:

\[
H(x, \sigma) = \begin{bmatrix}
L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\
L_{xy}(X, \sigma) & L_{yy}(X, \sigma)
\end{bmatrix}
\]

where \( L_{xx}(X, \sigma) \) is the convolution of the Gaussian second order derivative \( \frac{\partial^2}{\partial x^2} g(\sigma) \), where \( g(\sigma) \) is defined as:

\[
g(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y^2)/2\sigma^2}.
\]

More specifically,

\[
L_{xx}(X, \sigma) = I(x) * \frac{\partial^2}{\partial x^2} g(\sigma),
\]

where * denotes convolution. The same is applied for \( L_{yy}(X,s) \) and \( L_{xy}(X, \sigma) \).

In practice, \( L_{yy}(X,\sigma) \), \( L_{xy}(X, \sigma) \), and \( L_{xx}(X, \sigma) \) have to be discretized. In SURF, the second order Gaussians are approximated using box filters which take the mean of an area thus performing a smoothing operation. The integral image is used because the mean of a box on the center pixel can be rapidly computed using only 4 additions and one division by dividing by the sum of the pixels in a block by the total number of pixels in the box. Typically, a 9x9 box filter with scale, \( s \) (which corresponds to the Gaussian derivative \( s \)), of 1.2 is used.

The approximations \( D_{xx}, D_{yy}, \) and \( D_{xy} \) are used to denote the filter approximations in the \( x, y, \) and \( x-y \) directions, respectively. These are the discrete kernels corresponding to
\(L_{xy}(X, \sigma), L_{xy}(X, \sigma),\) and \(L_{xx}(X, \sigma)\). It has been shown in [123] that the determinant of the Hessian, \(\text{det}(H_{\text{approx}})\), is:

\[
\text{det}(H_{\text{approx}}) = D_{xx}D_{yy} - (0.9D_{xy})^2,
\]

4.3.5

where 0.9 is a constant weight.

When the determinant is at a maximum, the point is considered an interest-point. To make the feature extraction scale invariant, the box filter size changes such as 9x9, 15x15, 21x21, and 27 x 27. Note that the filter size increases by 6. This scale sequence is known as an octave and is advantageous due to allowing the image size remaining the same. For each new octave, the scaling should increase. For example, the second octave may be 15 x15, 27x27, 39x 39…. When changing the filter size, the scale changes accordingly. Thus, if \(s = 1.2\) and there is a scale of 3, \(s = 3.6\). Typically, 3 octave levels are used each with 4 different box filter sizes.

To localize the feature points, non-maximum suppression [125] is used to find the maximum pixel value across an edge. A non-maximum suppression in a 3 x 3 window is used for the 3 octaves. The maxima of the determinant of the Hessian are then interpolated in scale and image space.

To make SURF robust to rotation, the principal direction of the feature points is needed. To do this, Haar-wavelets are calculated in the \(x\) and \(y\) direction within a circular window of size \(6s\), where \(s\) is the scale for which the interest point was detected. The length of the wavelet used is \(4s\). Again, integral images are used to compute the Haar-wavelets response. These responses are then weighted with a Gaussian of \(s = 2.5s\) centered at the interest points.
This weighting helps such that the further the location of the wavelet is from the feature-point, the smaller the influence it has on determining the principal orientation of the feature. The dominant orientation is estimated by summing the horizontal and vertical wavelet responses within a rotating wedge which covers 60° in the wavelet response space. The largest local orientation vector is then chosen to describe the orientation of the interest.

Figure 4.3.2 (a) depicts the 100 SURF feature points and (b) depicts images of the Haar wavelets in the x and y directions. A visual example of calculating the (c) orientation of a SURF feature point which is used to (d) calculate the Haar wavelets in the x in y direction in a 4 x 4 rotated window.
point descriptor. Once the principal direction is selected, the axis is rotated to the principal
direction. Fig. 4.3.2(a) shows an example of 100 SURF features of an IC. Fig. 4.3.2(b)
depicts the Haar-wavelet response in the $x$ and $y$ directions. Fig. 4.3.2(c) shows an example
of a $60^0$ rotating window around an interest point and along with the orientation of the
interest point.

Extraction of the descriptors are then performed. A square region centered on the
interest point is constructed and oriented along the orientation of the interest point as shown
in Fig. 4.3.2(d). The square regions are then split into a $4 \times 4$ subarea of size $20s$, where $s$
is the scale where the feature was detected. Moreover, each subarea is calculated based on
$5 \times 5$ regularly spaced points. The Haar wavelet response is calculated; the wavelet
responses in the $x$ (horizontal) and $y$ (vertical) directions are denoted as $d_x$ and $d_y$,
respectively. The interest area is weighted with a Gaussian with $s = 3.3s$ centered on the
interest point giving robustness for deformations and translations. For each subarea, the
horizontal and vertical Haar wavelet responses are summed to form a first set of entries to
the feature vector. The absolute values of $|d_x|$ and $|d_y|$ are also summed in order to obtain
information about the polarity of the image intensity changes. Thus, each sub region has a
four-dimensional vector:

$$V = \left( \sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right).$$

Equation 4.3.6

Each interest point then has 64 feature points. Figure 4.3.2(d) shows an example of
calculating this feature vector for a SURF interest point. Moreover, the feature vector is
normalized such that it is a unit vector to achieve invariance to contrast.
From the images in the training class, a k-d tree is created [126] where k represents the number of nearest neighbors. When a new feature vector is introduced, a nearest neighbor search [127] is used to find the two closest feature vectors. This is done by measuring the Euclidean distance, $d$, between feature vectors:

$$d = \sqrt{\sum_{i=1}^{n} (X_{train} - X_{test})^2},$$  \hspace{1cm} (4.3.7)$$

where $n$ is the total number of features, $X_{train}$ is the data corresponding to SURF feature points from the database and $X_{test}$ is the SURF features from an interest point in the test image.

Using a 2-D tree, the shortest and second shortest Euclidean distance is computed. Note that the feature vectors contain 64 features. A ratio test is used to ensure that the smallest computed Euclidean distance is not an outlier using the following equation [128]:

$$d_1 < 0.8d_2,$$  \hspace{1cm} (4.3.8)$$

where $d_1$ is the smallest Euclidean distance and $d_2$ is the second smallest.

Moreover, we also wish to ignore matches where the distance is far. Thus a distance threshold is set at $d_1 > 0.25$ [129].

Each of the SURF feature vectors obtained from the test image is matched with the SURF features in the database based on the shortest Euclidean distance. The test image is said to belong to the same class as the image in the database with which it has the most SURF features in common.
Each of the SURF feature vectors obtained from the test image is matched with the SURF features [see Eq. 4.3.2] in the database based on the shortest Euclidean distance. The test image is said to belong to the same class as the image in the database with which it has the most SURF features in common.

4.4 Classification Results

A database was constructed using 7 true class (authentic ICs), 9 false class (counterfeit ICs), and 16 test images of real x-ray images of ICs obtained with the x-ray microscope at 80 kV and a power of 7 W similar to the images shown in Figs. 4.2.2(a), 4.2.3(a) and 4.2.3(c). The images were registered and aligned. The difference between an authentic IC and counterfeit IC can be either the wire configuration or the die. We note that photon-counting was applied to both the test images and the images used in the database.

Using the proposed photon-counting authentication scheme, Fig. 4.4.1 depicts the accuracy of the classifier when photon-counting was applied to the test images using the photon-counting model based on the weighted photon-counting method (PC-W) along with when no weights were used (PC-NW). Using the PC-NW model, 100% accuracy was achieved until 5e7 photons or 381.5 photons/pixel. However, using the PC-W model, it was possible to authenticate ICs down to about 1.5e6 photons in the scene, or 11.4 photons/pixel, with 100% accuracy. The area under the curve metric (AUC) was used to compare the performance of the classification using the PC-W and PC-NW models for when 11.4 photons/pixel were used in the scene. The area under the curve metric is defined as [126]:
\[ AUC = \frac{TP}{2(TP + FN)} + \frac{TN}{2(TN + FP)}, \]

where TP indicated true positive, FN is false negative, TN is true negative, and FP is false positive[126].

Table 4.4.1. shows the confusion matrix using the PC-W and PC-NW models with 11.4 photons/pixel. The AUC calculated for the PC-W model was 1 while the AUC for the PC-NW model about 0.53.

![Percent Accuracy vs. Number of Photons](image)

Fig. 4.4.1. Graph of percent accuracy vs. total number of photons for SURF feature matching using 16 test images. The images were counterfeit and authentic 256 (W) x 512 (V) pixel photon-limited. IC x-ray images using the photon-counting model based on the weighted photon-counting method (PC-W) and photon-counting model without weights (PC-NW). There were 16 images used for training which were photon-limited.
Table 4.4.1. Confusion Matrix for Photon-Limited IC Images with 11.4 photons/pixel Using the Photon-Counting Model Based on the Weighted Photon-Counting Method (PC-W) and the Photon-Counting Model Without Weights (PC-NW)

<table>
<thead>
<tr>
<th>True Class</th>
<th>PC-W Model</th>
<th>PC-NW Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted</td>
<td>Predicted</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>True Class</td>
<td>True Positive (TP) = 6</td>
<td>False Negative (FN) = 0</td>
</tr>
<tr>
<td>False Class</td>
<td>False Positive (FP) = 0</td>
<td>True Negative (TN) = 10</td>
</tr>
</tbody>
</table>

4.5 Conclusion

In summary, we have demonstrated an integrated circuit (IC) authentication scheme for photon-limited IC images with orders of magnitudes lower x-ray energies that are currently used in x-ray microscopes. Imaging an IC with an x-ray can potentially damage the circuit; however, this can be prevented as we have presented by substantially limiting the energy of the x-ray by orders of magnitude. We modeled an IC image acquired using an x-ray with reduced energy by applying a photon-counting technique based on a weighted photon-counting method to real x-ray microscope images of authentic and counterfeit ICs. By substantially reducing the number of photons, photon-limited images were generated. The speeded up robust features (SURF) of the image were then calculated. These features were compared to a database of SURF features from known authentic and counterfeit integrated circuits, which were not photon-limited, by finding the shortest Euclidean distance between feature vectors. The unknown IC image belonged to the same class as the image in the database with which it had the most SURF features in common. Although the quality of the photon-limited IC image was degraded, a 100% classification rate was still achieved up to about 11.4 photons/pixel using the photon-counting model based on the
weighted photon-counting method. Thus, it is possible to authenticate the ICs using an x-ray with substantially less energy while still successfully authenticating the IC. Future work includes exploring other classification approaches [73][130][131] along with further analysis of the photon-counting images acquired under reduced energy x-ray images [133][134]. In addition, we should experiment with a larger dataset to validate the robustness and accuracy of the proposed classification scheme.
Part III: Three-dimensional Imaging in Low Light Environments
Chapter 5

Three-dimensional Object Visualization and Detection in Low Light Illumination using Integral Imaging

Conventional two-dimensional (2D) imaging systems that operate in the visible spectrum may perform poorly in environments under low light illumination. In this work, we present the potential of passive three-dimensional (3D) integral imaging (II) to perform 3D imaging of a scene under low light conditions in the visible spectrum and without the need for a photon counting or cooled CCD camera. Using dedicated algorithms, we demonstrate that the reconstructed 3D integral image is naturally optimum in a maximum likelihood sense in low light levels and in the presence of detector noise enabling object visualization in the scene. The conventional 2D imaging fails due to the limited number of photons. Using 3D imaging, we demonstrate the potential for 3D detection of objects behind occlusion in a photon-starved scene. To the best of our knowledge, this is the first report of experimentally using II sensing under low illumination conditions for 3D visualization and 3D object detection in the presence of obscurations with a conventional image sensor [132].
5.1 Introduction

Imaging in poor illumination conditions is challenging using conventional image sensors that operate in the visible spectral range. This occurs due to the low flux of the scene along with noise that is in the camera itself including read noise, photon noise, and dark current. Thus, a two-dimensional (2D) imaging technique may not properly visualize a scene under these conditions. However, there is much interest in imaging in these environments including the fields of medical imaging [135], infrared imaging [136] and astronomy [137] etc. Visualizing an object at night can be accomplished using night vision imaging or long-wave infrared (LWIR) imaging [58]. Night vision imaging enhances the amount of light currently present. If no light is available, an infrared (IR) pulse laser source can be used. Moreover, LWIR detects the thermal radiation of a scene. However, the images suffer from poor resolution and a LWIR camera can be bulky and expensive.

It has been shown that using three-dimensional (3D) imaging techniques, passive photon counting imaging sensors that operate in visible spectrum can be used to accumulate enough photons to allow a scene to be visualized after 3D reconstruction [16][30][46][60][138]. Integral imaging is a technique that uses a lenslet array or a camera array [26][55][65][139][140][141] to capture multiple perspectives of a scene and produce a 3D reconstruction of a scene either computationally or optically. Synthetic aperture integral imaging (SAII) [55] is a 3D imaging technique that captures multiple perspectives of a scene by moving an imaging sensor or an imaging sensing array. The SAII technique requires the imaging sensor to move laterally or longitudinally in a plane parallel to the scene. At each camera position, a 2D image, known as an elemental image (EI), is acquired.
The captured images are processed to reconstruct a 3D image. In [30][46], SAII was used to visualize a scene in low illumination conditions by using SAII for image pick up and applying iterative denoising algorithms using total variation penalties on the 3D reconstructed image. Moreover, in [46] a 16-bit cooled CCD camera was used to image the scene under low illumination while in [30] a photon counting model was used to generate the photon-starved scene.

We investigate the potential of passive integral imaging to capture and reconstruct a 3D scene under low illumination without a photon counting imaging sensor or cooled CCD camera. Elemental images are captured under low illumination by a conventional low cost compact CMOS sensor to record a scene containing a person in the presence of occlusion. After image acquisition, 3D reconstruction is performed. Moreover, image denoising using total-variation regularization [142] is applied to the reconstructed image for scenes with significantly low signal-to-noise ratio (SNR). Face detection using the Viola-Jones object detection framework [143] is used on the 3D image reconstructed at the plane of the person. We demonstrate 3D integral imaging reconstruction is naturally optimum in low light levels and in presence of sensor noise. Also, while 3D object detection was successful in our experiment, 2D imaging failed due to the low light conditions.

5.2 Statistical Analysis of Low Illumination Scene

Acquiring the elemental images for SAII pickup process is shown in Fig. 5.2.1(a). The EI captured by a sensor, $E(x, y)$, can be modeled as $E(x, y) = i(x, y)r(x, y)$, where $i(x, y) (>0)$ is the illumination factor and $r(x, y)$ is the reflection component on the interval
The reflection component depends mainly on material properties, such as black valet \( r(x, y) = 0.01 \) or silver-plated metal \( r(x, y) = 0.90 \) [108].

The computational reconstruction of SAII can be regarded as the inverse process of the pickup process. As shown in Fig. 5.2.1(b), based on ray back projection and the pinhole model, pixels on each EI pass through their corresponding virtual pinholes and overlap in the 3D imaging space. The original information of the 3D scene on the reconstruction plane is in focus in the reconstructed image, while information not on the reconstruction plane appears blurred. For 3D reconstruction, the EIs are assumed to be read-noise limited images due to the low light levels [144]. The 3D reconstructed image using SAII, denoted as \( I(x, y; z) \), for read-noise limited images is:

\[
I(x, y; z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} E_{k,b} \left( x - k \frac{L_x \times p}{c_x \times M}, y - b \frac{L_y \times p}{c_y \times M} + \varepsilon \right),
\]

where \((x, y)\) is the pixel index, \(z\) is the reconstruction distance, \(O(x, y)\) is the overlapping number on \((x, y)\), \(K\) and \(B\) are the total number of EIs obtained in each column and row, respectively; \(E_{k,b}\) is the ideal EI in the \(k\)-th column and \(b\)-th row, \(L_x\) and \(L_y\) are the total number of pixels in each column and row, respectively, for each \(E_{k,b}\), \(M\) is the magnification factor and equals \(z / g\), \(g\) is the focal length, \(p\) is the pitch between image sensors, \(c_x\) and \(c_y\) are the size of the image sensor, and \(\varepsilon\) is the additive read noise due to the detector [46].
To estimate the pixel intensity in the 3D reconstructed image, we define
\[ \xi(x, y) = E(x, y) + \varepsilon \] for an EI. Since \( \varepsilon \) is Gaussian, we assume that \( \xi(x, y) \) is also Gaussian. More specifically, the probability density function of \( \xi(x', y') \sim N(\mu(x, y; z), \sigma^2(x, y; z)) \) where \( x' = x - k(L_x \times p) / (c_x \times M) \), \( y' = y - b(L_y \times p) / (c_y \times M) \).
\( M \), \( \mu(x, y; z) \) and \( \sigma(x, y; z) \) denotes the mean and standard deviation of pixel \((x, y)\) at reconstruction distance \(z\). To estimate the mean \( \mu(x, y; z) \), we maximize the log likelihood estimation by taking the derivative of the log likelihood of the probability density function in Eq. 5.2.2:

\[
\log \left[ l \left( \mu(x, y; z), \sigma^2(x, y; z) \big| \xi_{1,1}(x', y'), \ldots, \xi_{K,B}(x', y') \right) \right] = 5.2.2
\]

\[
- \frac{K \times B}{2} \log \left( \sigma^2(x, y; z) \right) - \frac{K \times B}{2} \log(2\pi)
\]

\[
- \frac{1}{2\sigma^2(x, y; z)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} (\xi_{k,b}(x', y') - \mu_x(x, y; z))^2,
\]

\[
\hat{\mu}(x, y; z) = \frac{1}{K \times B} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \xi_{k,b}(x', y'), 5.2.3
\]

where \( l(.) \) denotes the likelihood function.

Thus, 3D integral imaging reconstruction is naturally optimum in the maximum likelihood sense in the presence of noise due to low light conditions. The mean \( \hat{\mu}(x, y; z) \) at \((x, y)\) is simply the average of the observed pixels from the EIs. Moreover, the maximum likelihood estimator \( \hat{\mu}(x, y; z) \) takes the same form as the reconstructed image \( I(x, y; z) \) in Eq. 5.2.1.

Under low illumination, the illumination function \( r(x, y) \) is close to 0. Thus, the elemental images \( E(x, y) \) are substantially degraded. In addition, the imaging sensor itself contains noise making it more difficult to capture a scene with poor illumination including read noise, dark current shot noise, and reset noise. Fig. 5.2.2 depicts an example of a read-
noise dominant elemental image along with its corresponding histogram. The signal-to-noise ratio (SNR) of the captured EI can be defined as [46]:

Fig. 5.2.2. Read-noise dominant elemental image and its corresponding histogram.
\[ SNR = \sqrt{\frac{\langle g_o^2 \rangle - \langle N^2 \rangle}{\langle N^2 \rangle}}, \]

where \( \langle g_o^2 \rangle \) is the average power of the object region of the EI, and \( \langle N^2 \rangle = (\Phi_o + \Phi_b) Q_e t + D t + N_r^2 \) is the average power of the reference noise from the scene at the location of the object. Here, \( \Phi_o \) and \( \Phi_b \) are the photon flux of the object and background (photons/pixel/second), respectively, \( Q_e \) is the quantum efficiency (electrons/photons), \( t \) is the integration time (seconds), \( D \) is the dark current (electrons/pixel/second), and \( N_r \) is the read noise (electrons RMS/pixel).

In poor illumination conditions, read noise becomes much greater than photon noise, which is the square root of the signal, when a short exposure time is used [108]. Thus, an approximation for the SNR of a read-noise limited image and the approximate number of photons per pixel on the object, \( \gamma_{\text{num}} \), respectively, is [46]:

\[ SNR \approx \Phi_o Q_e t / N_r, \]

\[ \Phi_o t_e = N_{\text{photons}} \approx SNR \times N_r / Q_e. \]
5.3 Experimental Results

The SAII experiments, as shown in Figs. 5.3.1(a), were carried out under both good (high SNR) and low light (low SNR) illumination. The camera used was a Basler daA1920-30uc USB 3.0 camera, depicted in Fig. 5.3.1(b), having dimensions of 29 mm × 29 mm × 8.9 mm and aperture diameter of 16 mm. The camera pixel size is 2.2 μm × 2.2 μm with a sensor size of 4.22 mm × 2.38 mm. Moreover, the image size is 1920 (W) × 1080 (H) pixels and the exposure time used was approximately 0.1 seconds. The read noise is 6 electrons RMS/pixel, dark current is 6 electrons/pixel/sec, and quantum efficiency is 0.57 electrons/photon, which were obtained from the camera specifications. The number of EIs used was 36 using a 6 × 6 camera array with a pitch, p, of 40 mm. Figs. 5.3.2(a) and 5.3.2(b) depict the first elemental image containing a person behind occlusion, which has an SNR of 2.04 and 21.474 photons/pixel on the object, and the 3D reconstruction at z = 5 m where the person is in focus, respectively, with an SNR of 2.66 and 28 photons/pixel. These images were obtained with relatively higher illumination and serve as reference images for the low light scene experiments presented below. For SNR measurements, the object was assumed to be the face and the noise window was to the left (background noise) and the same size as the face.
Fig. 5.3.1. The (a) synthetic aperture integral imaging (SAII) used in the outdoor, low light experiment and the (b) USB board camera used in the experiment. \( z \) is the reconstruction distance, \( g \) is the focal length, \( p \) is the pitch between imaging sensors, and \( c \) is the sensor size.
Experiments were performed under lower light conditions, as shown in Fig. 5.3.2(a), which is denoted as scene 1. An external light source (i.e. cell phone light) was used to provide a limited illumination and the images were acquired on hilly terrain. The object
used was a person behind occlusion (i.e. tree branches), similar to the image shown in Fig. 5.3.3(a). We note that integral imaging has the ability to reconstruct the scene behind occlusion by integrating the rays that partially get through the occlusion from different angles. Fig. 5.3.3(a) depicts an EI taken under low light illumination conditions producing a read-noise limited image. The SNR of the EI is 0.646 with an estimated 6.8 photons/pixel on the person’s face. We note that the reference noise was a section of the image to the left of the face only consisting of noise. The 3D reconstruction is shown in Fig. 5.3.3(b) where the reconstruction distance is $z = 3.9$ m with a higher SNR of 0.942 and 9.916 photons/pixel corresponding to a 45.8% increase.

The experiment was repeated for extreme low light conditions for different object depths, as shown in Fig. 5.3.4. Fig. 5.3.4(a) shows the first elemental image corresponding to a scene having an SNR of 0.486 and 5.116 photons/pixel on the person’s face. Fig. 5.3.4(b) depicts the reconstructed image at $z = 4.3$ m with an SNR of 0.893, corresponding to an increase of 83.7%. Fig. 5.3.4(d) is an EI corresponding to a scene having an SNR of 0.272 and 2.863 photons/pixel on the person’s face. Fig. 5.3.4(e) depicts the reconstructed image at $z = 5.1$ m with an SNR of 0.796, which is a 192% increase.
Fig. 5.3.3. Experimental results under low light conditions for Scene 1: (a) read-noise limited EI with an SNR of 0.646, and (b) corresponding 3D reconstructed image at $z = 3.9$ m having an SNR of 0.942. The red box denotes a detected face using the Viola-Jones algorithm. The face was detected using 3D imaging while it was unable to be detected using 2D imaging.
Fig. 5.3.4. Experimental results under low light conditions for scene 2. (a) read-noise limited EI under low light conditions with an SNR of 0.486 along with (b) the corresponding reconstructed 3D image at $z = 4.3$ m having an SNR of 0.893. Moreover, (c) is the face after total-variation regularization denoising that was able to be detected. Experiments with a third scene (scene 3): (d) read-noise limited EI under low light conditions with an SNR of 0.272 and (e) the corresponding reconstructed 3D image at $z = 5.1$ m having an SNR of 0.796. Moreover, (f) is the face after total-variation regularization denoising that was able to be detected. The red boxes denote detected faces using the Viola-Jones algorithm.

Fig. 5.3.5(a) depicts a histogram of the object, that is the person’s face from the EI recorded in the experiment shown in Fig. 5.3.4 (a), which can be approximated as Gaussian. Fig. 5.3.5(b) depicts a histogram of the object from the 3D reconstructed image at $z = 4.3$ m. We observe that the histogram of Fig. 5.3.4 (b) has more pixels with similar intensity values compared to Fig. 5.3.4 (a). Thus, there is a reduction in entropy [108] in the reconstructed 3D images due to the reconstruction algorithm which is optimum in the maximum likelihood sense. Entropy measures the amount of uncertainty or randomness [108]:

94
\[ H = -\sum_{j=1}^{L} P(a_j) \log_2[P(a_j)], \]

where \( L \) is the total number of pixels, and \( P(a_j) \) is the probability of an event \( j \) occurring. The entropy of the object in the 2D EI shown in Fig. 5.3.3(a), 5.3.4(a) and 5.3.4(d) are 4.12 bits/pixel, 3.97 bits/pixel, and 3.97 bits/pixel, respectively. Moreover, the entropy of the corresponding 3D reconstructed images shown in Fig. 5.3.3(b), 5.3.4(b), and 5.3.4(e) are 2.43 bits/pixel, 2.17 bits/pixel and 2.06 bits/pixel. Table 5.3.1 presents all the metrics computed for Scenes 1, 2, and 3.

![Histogram](image)

(a)

![Histogram](image)

(b)

Fig. 5.3.5. Histogram of (a) object (i.e. the person’s face) from the elemental image [see Fig. 5.3.4(a)] with an entropy of 3.97 bits/pixel, and (b) the histogram of the object in the 3D reconstructed image shown in Fig. 5.3.4(b) with a reduced entropy of 2.17 bits/pixel.
Table 5.3.1 Scene metrics. $z =$ reconstruction distance; EI = elemental image; 3D Rec. = 3D reconstructed image; SNR = signal-to-noise ratio; inc. = increase; dec. = decrease.

<table>
<thead>
<tr>
<th>Scene</th>
<th>SNR (EI)</th>
<th>Entropy (bits/pixel)</th>
<th>SNR (3D Rec.)</th>
<th>Entropy (bits/pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scene 1 (z=3.9 m)</td>
<td>0.646</td>
<td>0.942</td>
<td>45.8</td>
<td>4.12</td>
</tr>
<tr>
<td>Scene 2 (z=4.3 m)</td>
<td>0.486</td>
<td>0.893</td>
<td>83.7</td>
<td>3.97</td>
</tr>
<tr>
<td>Scene 3 (z=5.1 m)</td>
<td>0.272</td>
<td>0.796</td>
<td>192</td>
<td>3.97</td>
</tr>
</tbody>
</table>

### 5.4 Face Detection

Face detection was performed on the 3D reconstructed low light images using the Viola–Jones framework [143]. A 3D reconstructed image is inputted into the detection framework. If no face is detected in the scene, a sliding window is used to search for a face. The window size is the estimated face pixel size [145] padded by 75 pixels in each direction. The window is then slid horizontally and/or vertically by an amount equal to half the window width and height, respectively. At each window, total-variation regularization using an augmented Lagrangian approach [142] is applied using a regularization parameter of $\mu = 100$. At each window, the Viola–Jones framework is applied, and the minimum face size in the detector is specified according to the expected size of the face, in pixels, [145] to eliminate false positives. Once a face has been detected, the box coordinates are stored. Moreover, if boxes in a single reconstruction plane overlap, the box containing the face with the lowest variance from the original reconstructed scene is assumed to be the best box. Lastly, if the face is detected on different depths, the face corresponding to the lowest variance is assumed to be the best reconstruction distance of that face.
The classifier was able to detect a face in the 3D reconstructed images shown in Fig. 5.3.3(b), as denoted by the red box, without using a sliding window. The detection system works well on the 3D reconstructed image while the face was not detected in the 2D elemental image shown in Fig. 5.3.3(a). Moreover, the classifier was applied to 13 different scenes (not shown in the Letter due to space limitations) having different SNR conditions, including the scenes shown in Figs. 5.3.4(b) and 5.3.4(e). If there were enough photons in the scene to reconstruct a face with sufficiently high SNR, the accuracy was 100% with a false positive rate of 0%. For scene 2 and scene 3 shown in Fig. 5.3.3, the face detection system was able to detect the 3D reconstructed face using the proposed algorithm with the sliding window. Figures 5.3.4(b) and 5.3.4(e) display the location of the detected face in the 3D reconstructed image. Figures 5.3.4(c) and 5.3.4(f) depict the denoised face, which were found using sliding windows of size 285 × 285 and 264 × 264 pixels, respectively. Face detection failed on the 2D EIs shown in Figs. 5.3.4(a) and 5.3.4(d), respectively.

5.5 Conclusion

In summary, using 3D integral imaging and a conventional image sensor such as a CMOS camera, we have demonstrated the potential for 3D object visualization and detection under poor illumination conditions. We have shown that 3D integral imaging (II) implemented with a low-cost CMOS imaging sensor may be capable of imaging and reconstructing a scene under poor illumination, that is without the need for a photon counting or cooled CCD camera. The 3D reconstruction formed from the 2D EIs is optimum in the maximum likelihood sense in low light conditions and in the presence of
detector noise, and may visualize objects in the scene whereas the same objects may not be visible using the 2D images. The Viola-Jones face detection algorithm was used to detect a face in the 3D reconstructed scene, which was behind occlusion and could not be detected in the 2D EIs. Future work includes applying the proposed scheme digital holography and polarimetric imaging [62], and exploring other classification approaches [147]. While SAII was used, other techniques may be [59].
Chapter 6

Learning in the Dark: 3D Integral Imaging Object Recognition in very low illumination conditions using Convolutional Neural Networks

We propose a framework for three-dimensional (3D) object recognition and classification in very low illumination environments using convolutional neural networks (CNNs). 3D images are reconstructed using 3D integral imaging (InIm) with conventional visible spectrum image sensors. After imaging the low light scene using 3D InIm, the 3D reconstructed image has a higher signal-to-noise ratio than a single 2D image, which is a result of 3D InIm being optimal in the maximum likelihood sense for read-noise dominant images. Once 3D reconstruction has been performed, the 3D image is denoised and regions of interest are extracted to detect 3D objects in a scene. The extracted regions are then inputted into a CNN, which was trained under low illumination conditions using 3D InIm reconstructed images, to perform object recognition. To the best of our knowledge, this is the first report of utilizing 3D InIm and convolutional neural networks for 3D training and 3D object classification under very low illumination conditions.
6.1 Introduction

Imaging a scene in low illumination conditions using conventional image sensors that operate in the visible spectrum is difficult as the captured images become read-noise dominant. Thus, signal-to-noise ratio (SNR) suffers resulting in poor scene visualization in addition to making object recognition a difficult task. There is much interest in a broad range of fields to image in low light conditions such as remote sensing [148], underwater imaging [149], night vision[150][151], etc. Image sensors that are designed for imaging in low light conditions include electron-multiplying CCD cameras (EM-CCD)[150][151], scientific CMOS (sCMOS) cameras [148] or night vision cameras. However, both the EM-CCD and sCMOS cameras are expensive and bulky. In particular, the EM-CCD needs to be cooled to around -55º C prior to operation. Night vision operates by amplifying the number of photons in the scene. If too few photons are available, an active near infrared source is required to illuminate the scene. Infrared cameras are effective in low light conditions. However, they have lower resolution compared with visible range cameras and may require bulkier and more expensive optics.

Passive cameras for 3D imaging using three-dimensional (3D) integral imaging (InIm) [55] have been reported [26][152][153][154][155]. In 3D InIm, an array of cameras or a single moving camera may be used to capture a scene, with each camera obtaining a unique perspective of the scene known as an elemental image (EI). Using the acquired EIs, a 3D image can be computationally or optically reconstructed. Integral imaging has been investigated in low illumination conditions. In [16], a photon-counting model was used to simulate photon-limited images from EIs that captured a 3D scene under sufficient
illumination. Computational 3D InIm reconstruction was performed using photon-limited EIs. It was shown that the 3D InIm reconstruction produces the maximum likelihood estimate of objects that lie on the corresponding 3D reconstructed depth plane. Thus, the 3D reconstructed image has higher SNR compared with a single 2D image. In [46], a 16-bit cooled camera was used to obtain EIs of objects under photon-starved conditions. After 3D reconstruction and denoising using total-variation denoising, object visualization was achieved whereas it was not possible using a single 2D image. In [132], 3D InIm was used to obtain EIs of an outdoor scene containing an object behind occlusion under low illumination conditions. With a single 2D image, face detection was not possible in the experiments. After computational 3D InIm reconstruction, object detection was successful. However, object classification in low light levels was not possible in this approach.

We show for the first time that it is possible not only to detect, but also to classify 3D objects in a scene in very low illumination conditions by 3D InIm. The novelty of the manuscript stems from the unique approach to the 3D training of the CNN classifier. We train the CNN classifier by using denoised 3D reconstructed images acquired using elemental images obtained under various low illumination conditions. By using 3D training data under these illumination conditions, the CNN is able to perform face recognition as it has been trained to recognize the face under non-optimal illumination conditions. Thus, the novelty lies in enabling a novel 3D object recognition approach for object recognition in low light levels which may not have been possible using conventional 2D approaches.

We use low cost passive image sensors that operate in visible spectrum. The EIs are read-noise dominant. 3D InIm is naturally optimal in the maximum likelihood sense for
read-noise dominant images as this follows a Gaussian distribution. Upon 3D InIm reconstruction, SNR increases resulting in improved image visualization. The scene is then denoised with total-variation regularization using an augmented Lagrange approach (TV-denoising) [142]. Regions of interest are obtained to detect faces in the scene [143], which are then inputted into a pre-trained convolutional neural network for facial recognition [156][157]. We demonstrate by experiments that the 3D InIm system trained in the dark with CNN was able to successfully perform face detection and classification in low light levels.

6.2 Three-Dimensional Integral Imaging in Low Illumination Conditions

3D InIm is a 3D imaging technique that uses a lenslet array, an array of cameras, or a moving camera to capture different perspectives of a scene, known as elemental images. The 3D InIm captures both intensity and angular information. Fig. 6.2.1(a) depicts the integral imaging pickup stage. Once the EIs have been acquired, the scene can be reconstructed, as shown in Fig. 6.2.1(b), by back-propagating the captured light rays through a virtual pin hole to a particular depth plane a distance \( z \) away. Fig. 6.2.1(c) depicts the chief ray, \( R_i \), from the object surface in 3D space \((x,y,z)\) at location \((x_0,y_0,z_0)\) with azimuth angle \( \theta \) and zenith angle \( \phi \) being imaged by the \( i \)-th lens located at \((x_i,y_i,z_i)\) and arriving at the sensor plane at \((\tau, \psi)\). Using the acquired elemental images, 3D InIm reconstruction can be performed optically or computationally. Fig. 6.2.2 depicts the synthetic aperture integral imaging (SAII) pick-up and reconstruction stage. Computational 3D InIm reconstruction is implemented as follows:
\[
I(x, y; z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} E^{k,b} \left( x - k \frac{L_x \times p_x}{c_x \times M} , y - b \frac{L_y \times p_y}{c_y \times M} \right),
\]

(6.2.1)

where \((x, y)\) is pixel index, \(z\) is reconstruction distance, \(O(x, y)\) is the overlapping number on \((x, y)\), \(K\) and \(B\) are the total number of elemental images obtained in each column and row, respectively; \(E^{k,b}\) is the elemental image in the \(k\)th column and \(b\)th row, \(L_x\) and \(L_y\) are the total number of pixels in each column and row, respectively, for each \(E_{k,b}\), \(M\) is the magnification factor and equals \(z/g\), \(g\) is the focal length, \(p_x\) and \(p_y\) is the pitch between image sensors, \(c_x\) and \(c_y\) are the size of the image sensor.

A captured image can be defined as \(E(x,y) = I(x,y)r(x,y)\) where \(I(x,y)>0\) is the illumination factor and \(r(x,y)\) is the reflection coefficient between 0 and 1 [108]. As the scene illumination decreases, the illumination factor diminishes. Moreover, read-noise becomes greater than the scene signal hindering adequate scene visualization. Thus, the image becomes read-noise dominant. Read-noise results from on-chip sensor noise, is additive, and can be modeled as a zero-mean Gaussian distribution. Using Eq. 6.2.1, the 3D InIm reconstruction with read-noise is:
Fig. 6.2.1. Integral Imaging. (a) pickup, and (b) 3D reconstruction stages. c = sensor size, p = pitch, f = focal length, z = distance. (c) Parameters details in (a). R_i = chief ray, ℓ_i = the i-th lens; θ = azimuth angle; ϕ = zenith angle.

Fig. 6.2.2. Synthetic aperture integral imaging (SAII) pick-up and reconstruction stage.
\[ I(x, y; z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \left( E^{k,b}(x', y') + \varepsilon_r^{k,b}(x', y') \right), \]

\[ = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} E^{k,b}(x', y') + \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \varepsilon_r^{k,b}(x', y'), \]  
(6.2.2)

where \( \varepsilon_r^{k,b}(x', y') \) is zero mean additive white Gaussian noise (i.e. read noise) for the elemental image in the \( k \)th column and \( b \)th row at location \((x', y')\), \( x' = x - k(L_x \times p_x) / (c_x \times M) \) and \( y' = y - k(L_y \times p_y) / (c_y \times M) \).

Taking the variance of Eq. 6.1.2, the variance of the noise component for a fixed \( z \), assuming that noise is wide sense stationary, is:

\[ \text{var} \left( \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{b=0}^{B-1} \varepsilon_r^{k,b}(x', y') \right) = \frac{1}{O(x, y)} \sigma^2, \]  
(6.2.2)

where \( \text{var}(.) \) is variance and \( \sigma^2 \) is the variance of the read-noise.

As the number of overlapping images increases, the variance or noise power, of the read noise decreases. It was shown that integral imaging reconstruction is naturally optimal in the maximum-likelihood sense for read-noise limited images as the distribution is approximately Gaussian. Without photon counting devices to measure the flux density, the SNR of the image is estimated as [46]:

\[ SNR = \sqrt{\frac{\langle g_o^2 \rangle - \langle N^2 \rangle}{\langle N^2 \rangle}}, \]  
(6.2.3)

where \( g_o \) is the average power of the object region in the EI, \( \langle N^2 \rangle \) is average noise power defined as \( \langle N^2 \rangle = (\Phi_o + \Phi_b)Q_e + D + n_i^2 \), where \( \Phi_o \) and \( \Phi_b \) are the photon flux of the object and background (photons/pixel/second), \( D \) is dark current (electrons/pixel/second), \( Q_e \) is the quantum
efficiency (electrons/photons), \( t \) is exposure time (seconds), \( n_r \) is read noise (RMS electrons/pixel/second), and \( <> \) denotes mean ensemble, respectively.

The number of photons per pixel (\( N_{\text{photons}} \)) can be estimated, assuming dark current noise is negligible and the exposure time is sufficiently short, as:

\[
\Phi_o \ t = N_{\text{photons}} \approx SNR \times n_r / Q_e ,
\]

where \( N_{\text{photons}} \) is the estimated number of photons.

### 6.3 Experimental Results

A synthetic aperture integral imaging experiment was conducted using Allied Vision Mako-192 camera with 86.4 mm x 44 mm x 29 mm camera dimensions. The sensor is an e2v EV 76C570 and a CMOS sensor type. The F/# is F/1.8 with focal length of 50 mm, pixel size of 4.5 um x 4.5 um, sensor size of 7.2 mm (H) x 5.4 mm (V), and image size of 1600 (H) x 1200 (V). The camera read-noise is 20.47 electrons/pixel/sec and the quantum efficiency at 525 nm is 0.44 electron/photons. A gain of 0 dB was used in the experiments. The InIm setup consisted of 72 elemental images using 3 x 24 array with 10 mm (H) x 80 mm (V) a pitch and exposure time of 0.015 s.

The experimental setup for low illumination conditions consisted of a 3D integral imaging set up with 6 subjects located a distance 4.5 m away from the camera array. Experiments were conducted for each subject under different illumination conditions resulting in different SNR levels. The illumination conditions were altered by adjusting the intensity of the light source. Fig. 6.3.1(a) depicts the elemental image [reference image]
with an SNR of 10.41 dB [i.e. good illumination] and Fig. 6.3.1(a) shows the 3D reconstructed image at \( z = 4.5 \) m with an SNR of 12.39 dB. Prior to 3D reconstruction, the elemental images were registered and aligned due to the experimental conditions (e.g. unbalanced camera array). Fifty bias frames were taken and averaged for each camera and subtracted from the elemental images. SNR was computed by assuming \( <g_o^2> \) is the object (i.e. the person’s face) while \( <N^2> \) is an area of scene that is completely dark. The elemental images acquired using the 3D InIm under low illumination conditions are shown in Fig. 6.3.1(b-f), which is in order of decreasing illumination and dominated by read-noise. In Fig. 6.3.1(b), the SNR was -1.20 dB with approximately 40.53 photons/pixel on the object. The person captured is still visible. In Fig. 6.3.1(c), the SNR decreases to -9.13 dB with 16.26 photons/pixel. The average power of object is lower than the noise power for the images shown in Fig. 6.3.1(d-f). As a result, SNR cannot be computed as \( \left\langle N^2 \right\rangle > \left\langle g_o^2 \right\rangle \) resulting in an imaginary number in Eq. 6.2.4.

3D reconstructed images at \( z = 4.5 \) m are shown in Fig. 6.3.2, with the 3D reconstructed images corresponding to the elemental images shown in Fig. 6.3.1. In Fig. 6.4.1 (b-f), the SNR increases to 8.93 dB, 0.96 dB, -5.28 dB, -9.67 dB, and -12.38 dB, respectively. Moreover, the corresponding photons/pixel for Fig. 6.3.2(b-f) is 130.02, 51.95 photons/pixel, 25.34 photons/pixel, 15.27 photons/pixel, and 11.18 photons/pixel, respectively. Fig. 6.3.3(a) depicts a graph of the SNR [see Eq. 6.2.3] of the elemental images and the corresponding 3D reconstructed images a \( z = 4.5 \) m as a function of illumination. Illumination levels 1 to 17 correspond to the scene light levels used in the experiments with 1 corresponding to the highest illumination level. The SNR of the 3D
reconstructed images is higher than that of the 2D EIs. We note that the SNR could not be computed for EIs with SNRs below -21.35 dB as the noise became greater than the signal. Fig. 6.3.3(b) depicts a graph displaying SNR (in dB) as a function of the number of photons/pixel. Overall, the 3D reconstructed images have a higher number of photons/pixel relative to their corresponding 2D EI.

Fig. 6.3.1. Sample elemental images with corresponding SNR and photon/pixel. The SNR of the images shown in Fig. 6.3.1(d-f) cannot be computed as the average power of the object regions is less than that of the background. N/A = not applicable.
Fig. 6.3.2. Three-dimensional (3D) reconstructed images at $z = 4.5$ m corresponding to the elemental images shown in Fig. 6.3.1. The SNR and photons/pixel can be computed for low light levels.

Fig. 6.3.3. SNR vs illumination. (a) Graph of SNR (in dB) as a function of decreasing illumination for both 3D reconstructed (Recon.) images taken at $z = 4.5$ m and elemental images. (b) SNR (dB) as photons/pixel on the object increases for 3D reconstructed image and elemental images. Illumination levels 1 to 17 correspond to the scene light levels used in the experiments with 1 corresponding to the highest illumination level. The photons/pixel for each SNR computed is shown in (b).

Additional experiments were carried out to evaluate the advantages of 3D InIm in low light conditions when compared with increasing the exposure time of a single camera and
recording multiple 2D elemental images using a single camera perspective followed by taking the average of the images. To evaluate image quality, we define the following metric:

\[
SNR_{\text{contrast}} = \frac{\mu_{\text{obj}}}{\sigma_{\text{noise}}},
\]

(6.3.1)

where \(\mu_{\text{obj}}\) is the mean of the object and \(\sigma_{\text{noise}}\) is the standard deviation of the background noise, which was an area of the image containing very low pixel values.

A 3D InIm experiment was conducted using the experimental parameters described above. Fig. 6.3.4(a) depicts the reference image while Fig. 6.3.4(b) depicts the 3D reconstructed image at \(z = 4.5\) m, with corresponding \(SNR_{\text{contrast}}\) of 31.5 dB and 33.76 dB, respectively. A low light condition experiment was then conducted. First, the scene was captured using a single image, but using three different exposure times. Fig. 6.3.4(a-c), depicts the captured images having an exposure time of 0.010 s, 0.015 s and 0.203 s under low light conditions, respectively. The \(SNR_{\text{contrast}}\) of the images shown in Fig. 6.3.4(a) and Fig. 6.3.4(b) cannot be computed as the object region intensity is less than that of the background. The \(SNR_{\text{contrast}}\) of Fig. 6.3.5(c) is 8.213 dB. Another set of experiments was carried out to capture 72 images from a single perspective along with a 3D InIm experiment. As shown in Fig. 6.3.6(a), 72 images from a single perspective were taken using an exposure time of 0.015 s and averaged while in Fig. 6.3.6(b) the 3D reconstructed image at \(z = 4.5\) m is shown. The \(SNR_{\text{contrast}}\) are 6.38 dB and 16.702 dB, respectively. The experiment was then repeated once more using an exposure time of 0.010 s. Fig. 6.3.6(c) depicts the average of 72 images obtained from a single perspective while Fig. 6.3.6(d) depicts the 3D reconstructed image at \(z = 4.5\) m. The \(SNR_{\text{contrast}}\) are 2.152 dB and 15.94 dB, respectively. Thus, by capturing both intensity and angular information, image contrast
and visualization using 3D InIm image reconstruction is improved compared with taking the average of multiple images captured using a single perspective or increasing the exposure time and capturing a single image. One reason is that 3D InIm segments out the object of interest from the background.

![Fig. 6.3.4 The (a) elemental image and (b) 3D reconstructed image at \( z = 4500 \) mm. The SNR contrast\(^\text{contrast}\)s are 31.5 dB and 33.76 dB, respectively.](image1)

![Fig. 6.3.5. 2D captured images having an exposure time of (a) 0.010 s, (b) 0.015 s and (c) 0.203 s under low light conditions. The SNR contrast\(^\text{contrast}\) of the images shown in (a) and (b) cannot be computed as the object region intensity is less than that of the background. The SNR contrast\(^\text{contrast}\) of (c) is 8.213 dB.](image2)
Fig. 6.3.6. (a) Average of 72 elemental 2D images, and (b) the 3D InIm reconstructed image at $z = 4.5$ m using an exposure time of 0.015 s for each elemental image. The SNR$_{contrast}$ is 6.38 dB in (a) and 16.702 dB in (b), respectively. (c) Average of 72 elemental 2D images and (d) the corresponding 3D InIm reconstructed image at $z = 4.5$ m using an exposure time of 0.010 s for each elemental image. The SNR$_{contrast}$ is 2.152 dB in (c) and 15.94 dB in (d), respectively.

6.4 Object Classification Using Convolutional Neural Networks

A Convolutional Neural Network (CNN) [126] was then trained on low illumination data for face recognition. An advantage of deep learning over other machine learning algorithms (e.g. Support Vector Machines or Random Forest Classifier) is that feature extraction is not needed. However, deep learning increases the computational complexity. In addition, deep learning requires a sufficiently large training set. The novelty of the CNN is that the training images were 3D reconstructed images of faces after TV-denoising obtained under different illumination conditions. The customized CNN employed in the experiments used larger filters in the convolutional layers as these
performed well on images obtained under low light conditions. Training in photon-starved environments helps improve the classifier’s ability to discern between different subjects in low illumination conditions. To illustrate the need for learning in the dark, normalized correlation was used to demonstrate the difficulty in discriminating faces under low illumination conditions. Fig. 6.4.1(a) shows a 3D reconstruction reference image at $z = 4.5$ m after TV-denoising obtained using EIs under an SNR of 10.41 dB. This image was correlated with 3D reconstructed images after TV-denoising whose EIs were obtained under an SNR of -12.75 dB, shown in Fig. 6.4.1(b) (true class object) and 6.4.1(c) (false class object), with correlation values of 0.58 and 0.48, respectively. Note that 1 indicates the images are perfectly correlated, and 0 indicates no correlation. Thus, it is difficult to discriminate objects under low light conditions without training the classifier information about what object appears in low light.

![Correlation between 3D reconstructed image at $z = 4.5$ m after TV-denoising using EIs (a) obtained under SNR of 10.41 dB as a reference. This 3D image was correlated with 3D reconstructed images after TV-denoising whose EIs were obtained under an SNR of -12.75 dB for (b) true class object, and (c) false class object; with correlation values of 0.58 and 0.48, respectively. Classification is difficult.](image)

A CNN was trained to perform facial recognition using data from the 3D InIm reconstructed data. A data set was collected consisting of 6 different subjects, and 17
different illumination conditions acquired using the 3D InIm. The images were then computationally reconstructed over distances of 4 m to 5 m with a step size of 50 mm where the true object distance was 4.5 meter. Fig. 6.4.2 below depicts examples of the training images used. The dataset was then split into training and testing whereas 4 randomly chosen illumination conditions having SNRs of approximately -1.41 dB, -8.322 dB, -8.971 dB, and -12.75 dB was not used to train the classifier (test set) and the other 13 illumination conditions were used for training. Thus, there were 24 test scenes. The training images were grayscale images of size 256 x 256 pixels and were perturbed by adding additive Gaussian noise with mean 0 and standard deviation of 0.01, 0.05 and 0.1, rotated -1, -0.5, 0.5 and 1 degrees, and translated -5, -3, 3, and 5 pixels in both the x- and y-directions generating a total of 29,232 images. The data was then denoised using total-variation regularization using an augmented Lagrange approach with a regularization parameter of 20000. Fig. 6.4.2 depicts examples of denoised 3D reconstructed training images acquired at various SNRs, reconstruction depths, additives noise and rotations. The CNN consisted of: a convolution layer [13 x 13, 20 filters], rectified linear unit layer (ReLU), 2 x 2 max pooling, convolution layer [11 x 11, 20 filters], ReLU, 2 x 2 max pooling, fully connected layer, and a SoftMax layer [6 outputs]. For training, stochastic gradient descent was used with a learning rate of .0001 and a maximum of 10 epoch used along with the cross-entropy metric to evaluate model performance. In total, the model took approximately 4 hours to train on a high performance computer utilizing a GPU Tesla K40m running CUDA 8.0 and implemented using MATLAB.
For classification, only regions of the 3D reconstructed image consisting of information from all 72 elemental images were considered to reduce the size of the input image. The image was then denoised using total-variation regularization using an augmented Lagrange approach with a regularization parameter of 20000 [142]. Afterwards, the Viola-Jones face detector [143] was used to find regions of interest. The regions were then inputted into the CNN classifier. This process was repeated over all $z$. If the same face appeared in the same region and detected over multiple depths, the estimated object reconstruction depth corresponded to the face with the highest mean intensity value. The rationale is that this reconstruction depth contained the most object information (i.e. strongest signal). More specifically, the object region can be modeled as signal plus additive noise whereas incorrect depths can be considered as noise. This is not the only approach and other approaches may be considered for future work. Note that faces were not detected for the Viola-Jones classifier for EIs with SNRs below approximately -21.36 dB. Table 6.4.1. summarizes the results. The proposed 3D system had a 100% accuracy. Figure 6.4.3 below depicts an overview of the classification scheme.
Fig. 6.4.3. Overview of training CNN classifier and acquiring test data

Table 6.4.1 Confusion matrix for face recognition using the CNN trained under low illumination conditions with 3D reconstructed images using 4 test scenes for each of the 6 subjects.

<table>
<thead>
<tr>
<th>Actual</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class 2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class 3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Class 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Class 6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

6.5 Conclusion

In conclusion, we present a 3D InIm system trained in the dark to classify 3D objects obtained under low illumination conditions. Regions of interest obtained from 3D
reconstructed images are obtained by denoising the 3D reconstructed image using total-variation regularization using an augmented Lagrange approach followed by face detection. The regions of interest are then inputted into a pre-trained Convolutional Neural Network (CNN). The CNN was trained using 3D InIm reconstructed under low illumination after TV-denoising. The EIs were obtained under various low illumination conditions having different SNRs. The CNN was able to recognize the 3D reconstructed faces after TV-denoising with 100% accuracy. Using a single 2D elemental image, regions of interest cannot even be extracted for low illumination conditions. Future work includes more dynamic scenes, utilizing different algorithms to improve image quality and classification in different scene conditions [140] and improving the data set size to create a more robust classifier.
BIBLIOGRAPHY


