
Theses and Dissertations

Summer 2017

A wavelet-based framework for efficient processing of digital imagery with an application to helmet-mounted vision systems

Jaclyn Ann Hoke
University of Iowa

Follow this and additional works at: <https://ir.uiowa.edu/etd>



Part of the [Electrical and Computer Engineering Commons](#)

Copyright © 2017 Jaclyn Ann Hoke

This dissertation is available at Iowa Research Online: <https://ir.uiowa.edu/etd/6435>

Recommended Citation

Hoke, Jaclyn Ann. "A wavelet-based framework for efficient processing of digital imagery with an application to helmet-mounted vision systems." PhD (Doctor of Philosophy) thesis, University of Iowa, 2017.

<https://doi.org/10.17077/etd.thro601t>

Follow this and additional works at: <https://ir.uiowa.edu/etd>



Part of the [Electrical and Computer Engineering Commons](#)

A WAVELET-BASED FRAMEWORK FOR EFFICIENT PROCESSING
OF DIGITAL IMAGERY WITH AN APPLICATION TO HELMET-
MOUNTED VISION SYSTEMS

by
Jaclyn Ann Hoke

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Electrical and Computer Engineering
in the Graduate College of
The University of Iowa

August 2017

Thesis Supervisor: Associate Professor Thomas Schnell

Copyright by
JACLYN ANN HOKE
2017
All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Jaclyn Ann Hoke

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Electrical and Computer Engineering at the August 2017 graduation.

Thesis Committee:

Thomas Schnell, Thesis Supervisor

David R. Andersen

Er-Wei Bai

Jon G. Kuhl

Andrew Kusiak

To my boys – Robbie and Daniel

Imagination is more important than knowledge.
Knowledge is limited. Imagination encircles the world.

--Albert Einstein

ACKNOWLEDGMENTS

I would like to thank my family for their never-ending support through this process from the time I started (and received) my Masters through the completion of this journey to finish my Doctorate. Your words of encouragement, trips to visit, and the couple occasional times you had to talk me off the “ledge” of potentially throwing in the towel were critical to my success. I could not have achieved this goal without you.

I would also like to thank all those at work who supported me in various ways. First, my “work dads”, David W. Jensen and Steven Koenck, thank you for so many reasons. You gave me a shot by hiring me for the internship that sparked my passion for image processing and led to many fascinating projects in the first couple years of my career. You never hesitated to be an ear when I needed to discuss work, school, and life in general and continued to provide moral support and motivation even after I moved on to other groups at work. Thank you to the managers I had while I was going through this process for allowing me the flexibility to work and go to school at the same time: Alex Postnikov, Scott Zogg, and Christina Conway. And also a huge thanks to Jason Wenger and Jason Ramapuram for supporting the software implementation.

ABSTRACT

Image acquisition devices, as well as image processing theory, algorithms, and hardware have advanced to the point that low Size-Weight-and-Power, real-time embedded imaging systems have become a reality. To be practical in a fielded application, an image processing sub-system must be able to conduct multiple, often highly complex tasks, in real-time. The design and construction of such systems have to address technical challenges, including real-time, low-latency processing and fixed-point algorithms in order to leverage lowest-power computing platforms. Further design complications stem from the reality that state-of-the-art image processing algorithms take very different forms, greatly complicating low-latency implementations. This dissertation presents the design and preliminary implementation of an image processing sub-system that minimizes computational complexity and power consumption by eliminating repeated transformations between processing domains. Specifically, this processing chain utilizes the LeGall 5/3 wavelet as the basis for applying multiple algorithms within a single domain. The wavelet processing chain is compared, in terms of image quality, computational cost, and power consumption, to a benchmark processing chain comprised of algorithms intended to produce high quality image results. Image quality is assessed through a subject matter expert evaluation. Computational cost is analyzed theoretically and empirically, and the power consumption is derived from the execution times and characteristics of the processing devices. The results demonstrate significant promise, but several areas for additional work have been identified.

PUBLIC ABSTRACT

Image acquisition devices, as well as image processing theory, algorithms, and hardware have advanced to the point that low Size-Weight-and-Power, real-time embedded imaging systems have become a reality. The design and construction of such systems have to address many technical challenges, including the reality that low-latency implementations are complicated by state-of-the-art image processing algorithms taking very different forms. This dissertation presents the design and preliminary implementation of an image processing sub-system that minimizes computational complexity and power consumption by eliminating repeated transformations between processing domains. Specifically, this processing chain utilizes the LeGall 5/3 wavelet as the basis for applying multiple algorithms within a single domain. The wavelet processing chain is compared, in terms of image quality, computational cost, and power consumption, to a benchmark processing chain comprised of algorithms intended to produce high quality image results. The results demonstrate significant promise, but several areas for additional work have been identified.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS.....	xiii
CHAPTER 1 INTRODUCTION	1
Motivations	1
Contributions	2
Organization	3
CHAPTER 2 HISTORICAL BACKGROUND OF THE APPLICATION	6
Helmet Mounted Displays	6
Characterizing Helmet Mounted Displays	7
History of Helmet Mounted Displays.....	9
History of Night Vision Goggles	11
Challenges and Applications	14
CHAPTER 3 TECHNICAL BACKGROUND	18
Wavelet Transforms.....	18
Continuous Wavelet Transform	19
Discrete Wavelet Transform.....	20
Real vs Complex.....	24
Candidate Algorithms	27
Registration.....	28
Noise Reduction	32
Fusion	38
Dynamic Range Compression	41
Compression	44
Summary.....	46
CHAPTER 4 FRAMEWORK DESIGN.....	47
Assumptions	47
DWT Processing Chain	48
Wavelet Selection.....	49
Algorithm Selection.....	51
Benchmark Processing Chain	62
Implementation	66
Framework Implementation	66
Graphical User Interface.....	69
CHAPTER 5 RESULTS, ANALYSIS AND DISCUSSION	72
Image Quality Analysis	72
Image Results	74
Assessment	84
Discussion.....	86
Computational Complexity and Power Consumption	88

Computational Complexity Analysis	89
Power Consumption Analysis	94
Discussion.....	96
Algorithm Selection and Ordering.....	98
CHAPTER 6 SUMMARY AND FUTURE WORK	101
Summary.....	101
Future Work.....	102
REFERENCES	105
APPENDIX A SUPPLEMENTARY IMAGE DATA	118
Multi-Focus Fusion Results.....	118
Comparison of Adaptive Scaling of High-Frequency Sub-bands	119
Images Used in DRC SME Evaluation.....	120
APPENDIX B DT-CWT WAVELET FRAMEWORK	123

LIST OF TABLES

Table 1: Summary of image registration literature reviews.....	30
Table 2: LeGall 5/3 Filter Coefficients.....	49
Table 3: Votes received by each framework during subjective assessment of results	85
Table 4: Average processing time of the DWT-based framework	93
Table 5: Average processing time of the benchmark framework.....	93
Table 6: Power consumption of the DWT-based framework.....	95
Table 7: Power consumption of the benchmark framework.....	95

LIST OF FIGURES

Figure 1: Scene-to-eye diagram for a digital system	6
Figure 2: F-35 Gen III helmet.....	11
Figure 3: Clip-on NVG [27]	11
Figure 4: Typical NVIS I2 tube and optics.....	12
Figure 5: FoV of the human visual system and HMDs (shown in bold)	15
Figure 6: (a) DWT Filter Block Diagram and (b) Multi-level Filter Bank.....	23
Figure 7: Two-Level 2D DWT	24
Figure 8: Two-Level "Single-tree" 1D CWT Filter Bank.....	25
Figure 9: Three-Level DT-CWT Filter Bank.....	26
Figure 10: Representative Processing Chain	27
Figure 11: Salt and pepper noise with variance of 0.05.....	33
Figure 12: Gaussian noise with mean = 0 and variance of 0.05	34
Figure 13: Speckle noise with variance of 0.05	34
Figure 14: Brownian noise.....	35
Figure 15: High-level flow for (a) spatial fusion and (b) transform-based fusion.....	39
Figure 16: Compression Block Diagram	45
Figure 17: Example hard threshold operator for $\tau = 3$	52
Figure 18: Example soft threshold operators for $\tau = 3$	53
Figure 19: Visible haloing artifacts in maximum selection multi-focal fusion	55
Figure 20: Area-based multi-focal fusion algorithm minimizes artifacts but has lost detail.....	55
Figure 21: Two views of how corresponding pixels between passes are used for the area-based activity measure computation	58
Figure 22: Sample scene with (a) original 16-bit representation and (b) 8-bit representation after a simple DRC approach that shows hidden detail	59

Figure 23: Example plot of the resulting $y_c[n]$ for $y_{LL}[n] = [1:10,000]$ when $\alpha_1 = 256$, $\alpha_2 = 2048$, $\max(C_{LL}) = 30196$, $\beta_1 = 8$, $\beta_2 = 16$, and $\beta_3 = 256$	61
Figure 24: (a) 3x3 pixel and (b) 5x5 pixel filter masks of the median filter algorithm.....	63
Figure 25: Algorithm flow for DWT-based processing chain.....	67
Figure 26: Algorithm flow for benchmark processing chain.....	68
Figure 27: Preliminary Qt 4 GUI.....	70
Figure 28: Expanded Qt 4 GUI.....	71
Figure 29: Test Scene 1 – Mid input images with (a) near, (b) mid, and (c) far focus.....	75
Figure 30: Output from benchmark framework after applying additional histogram equalization.....	76
Figure 31: Sample output of DWT-based framework after applying additional histogram equalization.....	77
Figure 32: Sample output of DWT-based framework with no histogram equalization.....	78
Figure 33: Sample output of DWT-based framework with modified β values in DRC.....	79
Figure 34: Oversaturated output of DWT-based framework from modified α values in DRC.....	80
Figure 35: Sample Diorama results from (a) benchmark chain and (b) wavelet chain.....	81
Figure 36: Sample Dilbert results from (a) benchmark chain and (b) wavelet chain.....	82
Figure 37: Sample Matrix results from (a) benchmark chain and (b) wavelet chain.....	83
Figure 38: DWT-based framework results with CLAHE applied pre-processing and simple scaling of wavelet coefficients for DRC.....	87
Figure 39: Execution time in milliseconds for DWT-based and benchmark frameworks.....	92
Figure A.1: Example (a) far focus, (b) mid focus, (c) and near focus inputs with (d) fused results.....	118
Figure A.2: DRC with (a) a single scalar and (b) adaptive scaling of detail coefficients.....	119
Figure A.3: Dilbert (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results.....	120

Figure A.4: Diorama (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results	121
Figure A.5: Matrix (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results	122
Figure B.1: DT-CWT (a) HL Orientations, (b) HH Orientations, (c) LH Orientations, (d) LL approximation coefficients	123
Figure B.2: Sample image results after applying the DT-CWT processing chain.....	124

LIST OF ABBREVIATIONS

ANVIS	Aviator Night Vision Imaging System
AWMF	Adaptive Weighted Median Filter
CDF	Cohen-Daubechies-Feauveau
CLAHE	Contrast Limited Adaptive Histogram Equalization
CPU	Central Processing Unit
CRT	Cathode Ray Tube
CWT	Complex Wavelet Transform
DASH	Display and Sight Helmet
DCT	Discrete Cosine Transform
DRC	Dynamic Range Compression
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding with Optimal Truncation
FAA	Federal Aviation Administration
FFT	Fast Fourier Transform
FSD	Filter-Subtract-Decimate
GHz	GigaHertz
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HDR	High Dynamic Range
HE	Histogram Equalization
HH	High-High sub-band
HIDSS	Helmet Integrated Display Sight System
HL	High-Low sub-band
HMD	Helmet Mounted Display
HMDS	Helmet Mounted Display System
HUD	Heads Up Display
IHADSS	Integrated Helmet and Display Sighting System
IHS	Intensity-Hue-Saturation
IQA	Image Quality Assessment
IR	Infrared
JHMCS	Joint Helmet Mounted Cueing System
JPEG	Joint Photographic Experts Group
JSF	Joint Strike Fighter

LH	Low-High sub-band
LL	Low-Low sub-band
MF	Median Filter
MSE	Mean Squared Error
MTFA	Modulation Transfer Function Area
NASA	National Aeronautics and Space Administration
NVG	Night Vision Goggles
PCA	Principal Component Analysis
PSNR	Peak Signal to Noise Ratio
ROLP	Ratio-of-low-pass
SME	Subject Matter Expert
SQRI	Square Root Integral
SVD	Singular Value Decomposition
SWAP	Size-Weight-and-Power
TV	Total Variation
VPU	Vision Processing Unit
VR	Virtual Reality
WMF	Weighted Median Filter

CHAPTER 1 INTRODUCTION

Motivations

Image acquisition devices, as well as image processing theory, algorithms, and hardware have advanced to the point that low Size-Weight-and-Power (SWAP), real-time embedded imaging systems have become a reality. To be practical in a fielded application, an image processing sub-system must be able to conduct multiple, often highly complex tasks, in real-time, including many (or all) of the following:

- Signal conditioning
- Noise reduction
- Registration and panoramic stitching of multiple sensor inputs
- Super-resolution
- Multi-focal fusion (also known as Hands-Free Focus)
- Multi-spectral fusion
- High dynamic range compression
- Contrast enhancement
- Compression for image\video storage or transmission

The design and construction of such systems have to address technical challenges, including real-time, low-latency processing and fixed-point algorithms in order to leverage lowest-power computing platforms.

Further design complications stem from the reality that the state-of-the-art algorithms for the above list of image processing tasks take very different forms, greatly complicating low-latency implementations. For example, the optimal signal conditioning algorithm might be performed in the spatial domain, optimal noise removal might utilize a wavelet-based algorithm, and the optimal registration algorithm might be performed in the Fourier domain. These transformations between different processing domains can easily begin to dominate the total computational cost of an image processing sub-system.

The objective of this work is to investigate the feasibility of performing most, or even all, of the image processing tasks for a given application in a single domain, namely the wavelet domain. Since their introduction in the mid-1980s, wavelets have become a standard tool for signal and image processing. The multi-resolution aspect of wavelet processing provides a tradeoff between the spatial domain and the frequency (Fourier) domain. Also, the linear computational complexity (i.e., faster than a Fast Fourier Transform (FFT)) and fixed-point implementations lend themselves to computation- and latency-sensitive applications. This research is aimed at establishing a structured framework for the wavelet-based design of computationally efficient processing chains for digital imagery, quantifying the impact on computational cost and power consumption.

Helmet-mounted vision systems, and more explicitly digital vision systems capable of both day and night operations, are the primary focus. The manufacture of microchannel plates for image intensifier tubes used in night vision imaging systems is a complex, and often error-prone, process. A system built upon digital technologies could provide a lower-cost solution with nearly equivalent, if not improved, capabilities. Day- and night-capable digital vision systems still have not become a complete reality as the performance of the current low-light digital sensors is insufficient without the application of advanced image processing algorithms, and the required algorithms have been difficult to implement in a real-time manner.

The methods used in this research are formulated for wider applicability such that, given a selection of algorithms capable of achieving a desired outcome, the framework will allow this wavelet approach to be tailored to a specific application.

Contributions

Specifically, this dissertation makes the following contributions to advance the state-of-the-art in efficient digital imagery processing chains:

- Develop and assess a wavelet-based framework for digital processing chains that enable complex, real-time applications such as helmet-mounted vision systems. Design of the processing chain includes a review and selection of an appropriate wavelet, as well as the development of an understanding of how the various techniques for processing wavelet coefficients impact the final image results. The resulting image data was assessed qualitatively through a subjective analysis conducted by a set of subject matter experts (SMEs) with operational experience relevant to the target application or expertise in relevant areas of image processing.
- Demonstrate the viability of the wavelet-based framework by characterizing the computational savings and associated power savings for processing digital imagery captured in both laboratory and operationally relevant environment. Digital vision systems are inherently a real-time application that often requires an already heavily burdened end user to carry the power source (e.g. dismounted soldier carrying pounds of batteries). The study benchmarks the wavelet-based processing chain against a baseline comprised of image processing algorithms aimed at producing high-quality image results.
- Provide an understanding of how this framework can be tailored for wider applicability through guidelines for selecting and ordering algorithms in the processing chain in terms of the impact on image results, computational efficiency and power consumption.

Organization

Chapter 2 describes the general background of the application and problem space motivating this research. The chapter starts with a historical overview of helmet-mounted display (HMD) systems and night vision technology, predominantly from the military perspective. The chapter outlines the technical and non-technical challenges that continue

to drive additional research in this area and concludes with a presentation of the wide variety of applications for helmet-mounted or head-worn devices. The intent of this chapter is to present the motivation for the broad applicability of this research beyond the target application context and to illustrate the importance of characterizing the impacts of algorithm selection and ordering.

Chapter 3 reviews the technical background of wavelet transforms and the candidate processing algorithms. The first part of the chapter is intended to concisely present wavelet transforms. A brief overview of the continuous wavelet transform is provided; however, given the discrete nature of digital imagery, emphasis is placed on the discrete wavelet transform (DWT). It includes a comparison of the real-valued and complex-valued wavelet transform. The second part of Chapter 3 describes the candidate processing algorithms considered for inclusion in the image processing chain. This review focuses primarily on wavelet-based approaches for each algorithm, but also includes recent non-wavelet techniques.

Chapter 4 describes the processing framework design, including the documentation of several assumptions that influenced design decisions. It describes the wavelet selected for the DWT-based processing chain and presents the algorithms selected for computationally efficient enhancement of digital imagery. The chapter also presents a benchmark processing chain used to support the analysis of the DWT-based processing chain. A brief rationale is provided for each algorithm in both processing chains. The chapter concludes with the implementation details of the wavelet-based and benchmark processing chains.

Chapter 5 presents the research results. It provides a description of the approaches used to quantify the results in terms of image quality, computational cost and power consumption and summarizes the results of each of these analyses. The chapter also provides a discussion of the relevance of these results to the state of the art in the field. It concludes with a set of guidelines that should be considered when tailoring the

framework for applications beyond the target application in helmet-mounted vision systems.

Chapter 6 summarizes the research presented in this dissertation. It provides a brief recap of the primary contributions and identifies several areas that would benefit from additional research efforts. The appendices provide supporting information not included in the main text. Appendix A includes supplementary image data examined during the processing chain design to select and evaluate the results of individual algorithms. Appendix B presents a short summary of a preliminary evaluation of an alternative wavelet for the processing chain.

CHAPTER 2 HISTORICAL BACKGROUND OF THE APPLICATION

Given that the military has been at the forefront of the development and application of HMDs and night vision devices, and that the intended application of this research is in that domain, this historical overview presents primarily the military perspective. This chapter also summarizes the technical and non-technical challenges that continue to drive additional research in this domain. The wide varieties of applications for helmet-mounted or head-worn devices are presented; illustrating additional motivation for continued HMD-related research.

Helmet-Mounted Displays

Warfighters have worn protective head gear for thousands of years, and it has evolved to provide both protection and functionality. The integration of advanced electronic functionality into head-worn equipment can enhance soldier capabilities during the execution of tactical missions, providing enhanced situation awareness that varies widely depending on the tactical mission and operational environment. Just as there are many different types of tactical missions, there are different types of helmet systems used to best enhance the soldier's effectiveness. New systems are implemented to provide specific advanced capabilities or performance enhancement over existing systems, but new functionality always comes at a cost that requires an analysis of the tradeoffs between overall system performance and human sensory system performance.

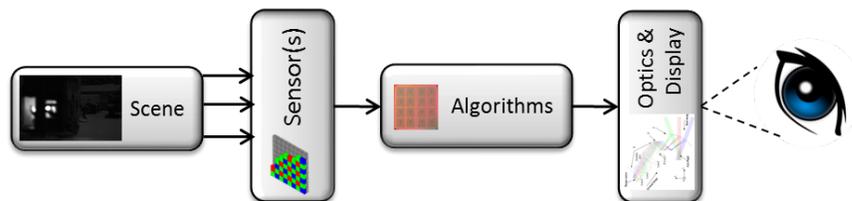


Figure 1: Scene-to-eye diagram for a digital system

Figure 1 shows the high-level flow, from the scene to the eye, for presenting information to the user of a typical digital system.

Characterizing Helmet-Mounted Displays

What is an HMD? Melzer and Moffit [1] described an HMD as a system that minimally consists of “an image source and collimating optics in a head mount.” An aviation-specific extension to this description was presented by Rash [2] to include a coupling system that utilized head and/or eye position and motion to slave one or more aircraft systems, most often a head-directed sensor. Stepping back from the aircraft-centric characterization, Manning and Rash [3] provided a general method for describing HMD systems comprising the same basic building blocks, including:

- A *mounting platform* that serves as an attachment point and stabilizer to maintain the alignment between the user’s eyes and the viewing optics
- An *image source* that generates information imagery optically presented to the user. Technological advances have enabled a wide range of potential image sources for the varying applications for HMDs.
- *Relay optics*, consisting of various optical elements (such as lenses) and a beam splitter (combiner), which transfers the imagery to the eye(s).
- A *head tracker* that enables the presentation of spatially-referenced symbology or outside information provided by a sensor or synthetic database. A head tracker is optional if the HMD is used only to present status information to the user.

With an understanding of the basic building blocks of an HMD, the methods for categorizing the systems can be discussed. These categorizations reflect many of the design tradeoffs that must be considered during the design and development of an HMD. There are three primary methods for classifying HMDs: 1) the method by which the imagery is presented to the eye, 2) the mode by which symbology or imagery is presented, and 3) the optical design.

Shontz and Trumm [4] selected a categorization based on the method by which the imagery is presented to the eyes. The systems are characterized by presentation to one eye or both and by whether the display is see-through or occludes vision. Examples include the one-eye, see-through type like the Integrated Helmet and Display Sighting System (IHADSS) [5][6] fielded on the AH-64 Apache helicopter and the two-eye, see-through type such as the Thales TopOwl[®] [7]. Additional types include one-eye, occluded systems and two-eye, occluded configurations that encompass most night vision goggles (NVGs).

The second classification is based upon the mode by which symbology or imagery is presented [8] and uses the terms *monocular*, *biocular*, and *binocular*. *Monocular* refers to an HMD that presents information to a single eye, such as the previously mentioned IHADSS. *Biocular* refers to a system that presents two images from a single sensor or multiple sensors where the images are identical. *Binocular* refers to a system that presents two images from two sensors displaced in space. Binocular systems can provide perspective to the user and can overcome field of view (FoV) limitations via partial overlap of the images. Examples include the Helmet Integrated Display Sight System (HIDSS) [9] and the Aviator's Night Vision Imaging System (ANVIS) [10].

The third approach characterizes HMDs by optical design. The most predominant type within this scheme uses reflective and refractive lenses to relay imagery with a final element called a “combiner” positioned in front of the eye [11]. Visor-projected designs [12] relay the imagery optically and project it onto the visor, where it is reflected into the user's eye. This is a lower-weight approach that provides an improved center of mass (CoM), but it is highly susceptible to image degradation in high-vibration environments. Designs that use holographic optical elements [13] allow for low-weight and compact designs but are susceptible to humidity and temperature degrading the hologram. Waveguide technology, such as the BAE Q-Sight[™] [14] or SBG Labs DigiLens[®] [15], uses holographic optics embedded within a visor or glasses to direct imagery to the eye. This

technology provides a low-weight approach that can be used with existing NVGs; however, it is new enough that the disadvantages are still being examined. Another approach within this classification is to use lasers that scan the imagery directly onto the retina [16], offering reduced weight and improved CoM. Scanning complexity, susceptibility to high-vibration environments and safety concerns have severely limited the use of these types of designs.

History of Helmet-Mounted Displays

The official history of HMDs started during World War I between 1915 and 1917. Albert Pratt was awarded several patents for an “Integrated Helmet Mounted Aiming and Weapon Delivery System” for a marksman [17]. Since this time, various militaries across the world have pursued the development, application, and fielding of a variety of helmet-mounted technologies. In aircraft, the development of head-up displays (HUDs) began to mitigate the safety concerns associated with pilots having to look down at displays to obtain status information. As beneficial as HUDs proved to be, their forward fixed position limited the usefulness in an environment where constant head motion is required. This factor provided the motivation for mounting a display on the head (or helmet).

One of the first sighting HMDs ever fielded was the gimbaled gun, deployed in the U.S. Army AH-1G Huey Cobra [18]. The Navy introduced a system into the F-4J and F-4N Phantom fixed-wing jets coupled with radar and the AIM-9H Sidewinder missiles [19]. Both of these systems were fielded in the early 1970s. The first complete visually coupled system was introduced and fielded for operation use in the 1980s. The IHADSS was fielded in the U.S. Army AH-64 Apache and has been used in both day and night missions for over three decades [20]. In 1994, the U.S. Air Force and U.S. Navy initiated the first joint office project developing the Joint Helmet Mounted Cueing Systems (JHMCS), a system that would enhance pilot situation awareness and provide head-out control of sensors and targeting systems. The system has a modular design that

accommodates the use of an image intensified night vision module that can be reconfigured in flight [21]. Many more HMD programs have developed systems for fixed-wing, rotary wing, mounted and dismounted soldier, simulation and training, and medical applications.

Outside the United States, two of the most significant systems were developed by the Russians and the Israelis. The Russians developed a helmet-mounted sight attached to the ZSh-5 series helmet to support the Vympet R-73/AA-11 Archer carried by the MiG-29 Fulcrum and the Su-27 Flanker [22]. While this system had relatively limited functionality, it greatly improved the close combat capabilities [23] and was sold to air forces in India, Iraq, North Korea, Libya, Syria, Iran, Yugoslavia, and possibly Cuba [24]. The advantages this system provided during the Cold War resulted in a dramatic increase of HMD programs in the West. The Israelis began development of the Display and Sight Helmet (DASH) series of HMDs in the 1970s. The DASH 3, combined with the Python 4 had “fire and forget” capabilities and helmet-sight guidance. This system is highly significant because it was one of the original “embedded” designs, with the complete optical and position sensing package built into a standard helmet form factor, and provided a significant basis for the JHMCS [25]. Several other systems have been developed in countries like France and the United Kingdom but have not had the same level of impact on continued research and development activities.

The most recent, and most advanced, system is the Joint Strike Fighter (JSF) Helmet Mounted Display System (HMDS) developed by Rockwell Collins, Inc., shown in Figure 2 above. The HMDS provides one of the first fully integrated day/night systems, capable of projecting night vision imagery directly onto the visor. It is a biocular system that provides a virtual HUD, enabling the JSF to be the first fighter aircraft in over 50 years without a HUD [26]. Each F-35 helmet is custom-built for the pilot to ensure accurate alignment of helmet symbologies.

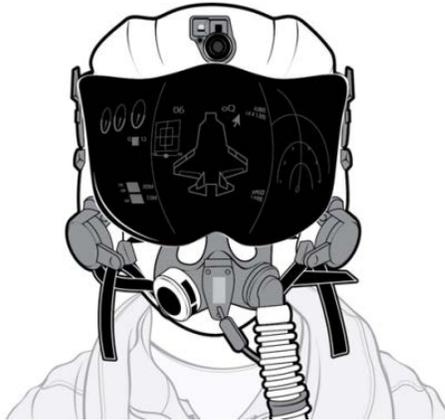


Figure 2: F-35 Gen III helmet¹

History of Night Vision Goggles

Night vision imaging systems are light image intensification (I^2) devices that amplify the ambient light, including visual light and near Infrared (IR). Specifically, the term Night Vision Goggles (NVGs) refers to an I^2 device that is head-worn and most often a binocular design, affixed to a helmet in a clip-on fashion, as shown in Figure 3.

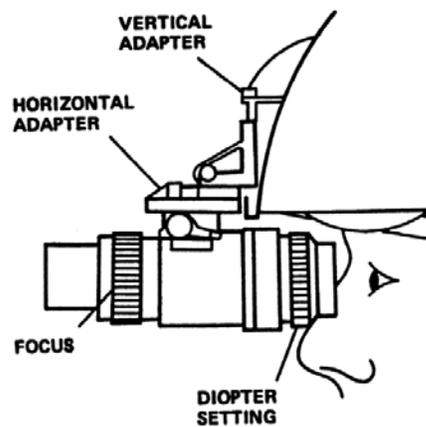


Figure 3: Clip-on NVG [27]

¹ Illustration by Peter Sucheski – January/February 2016 issue of *Popular Science*

Development of night vision devices began in the late 1930s for use by the German army and devices were tested in the early 1940s mounted on Panther tanks [28]. At the same time, the United States was developing infrared sighting devices [29]. The systems developed by both the Germans and the U.S. were “active” devices that utilized infrared illumination sources, which could be detected by opposing forces. This limited the utility of active night vision technology.

The 1950s saw considerable research in night vision technologies [30] with a wide range of applications. In the late 1950s, the U.S. Army began experimenting with T-6A infrared driving binoculars [10]. Around this same timeframe, a team at Bendix Research Laboratories was researching the first continuous-channel electron multiplier, a key step in the development of the microchannel plate [31]. The microchannel plate is one of the foundational elements of passive night vision upon which future night vision devices would be built.

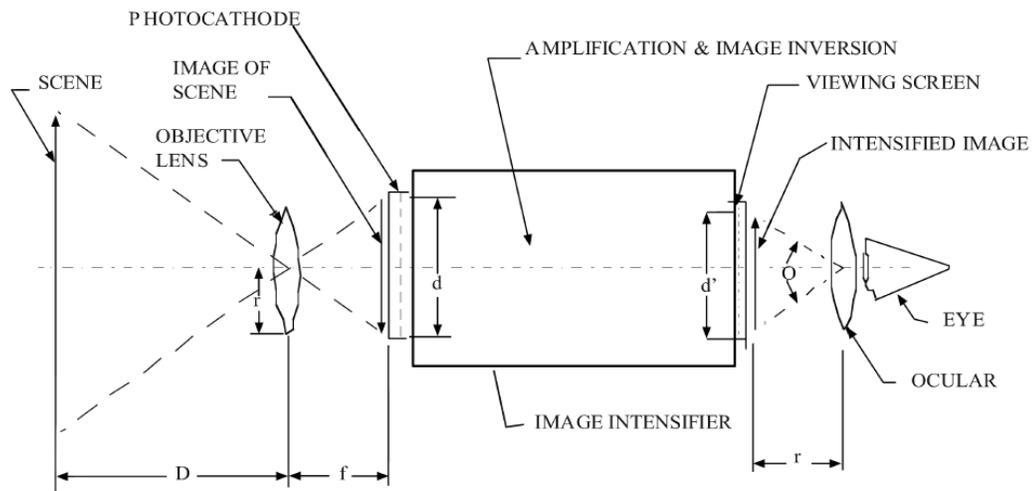


Figure 4: Typical NVIS I² tube and optics²

² *The Avionics Handbook*, Chapter 7 – Night Vision Goggles

Gen I I² tubes were developed in the 1960s, but were too large for a head mount. It was in the late 1960s that Gen II I² tubes were developed, enabling small, lightweight packaging suitable for head-mounted applications. The U.S. Army Night Vision and Electro-Optics Laboratory began using Gen II technology to develop NVGs for foot soldiers and experimenting with them in aviation during the late 1960s and early 1970s. The U.S. Air Force began to use SU-50 Electronic Binoculars in the 1970s, and the Army adopted Gen II AN/PVS-5, both for aviators. In the late 1970s, ANVIS [10] development began, utilizing the first Gen III I² tubes.

The 1980s saw the introduction of two versions of ANVIS into military aviation, and a critical requirement for compatible lighting was identified. The Army released Aeronautical Design Standard ASD-23 [32]. This was followed shortly by a tri-service specification, MIL-L-85762 [33] and the updated MIL-L-85762A that expanded the requirements to accommodate various types of night vision imaging systems [34][35]. A series of nighttime accidents involving NVGs led to a 1989 Congressional hearing [36] and a review of the safety and appropriateness of their use in aviation [27][37][38]. The results deemed that the older AN/PVS-5 NVGs were insufficient and that newer ANVIS were necessary. The 1990s and beyond saw an increased desire for the integration of HUD symbology with NVGs, spurring further research that ultimately led to the development of the AN/AVS-9, Panoramic NVGs, and Low Profile NVGs over the next several decades.

In the mid-1990s, the Federal Aviation Administration released several studies on the use of NVGs in civil aviation [39]-[42]. The reports found that there were several key issues that must be addressed before NVGs could be successfully used: restriction to Gen III devices, modification of cockpit light, modification of interior light, modification of exterior lighting, development of appropriate training programs, and updates to regulations and standards. Additional civil applications for NVGs include forestry\wildlife observation, law enforcement, EMS helicopters [43], and the luxury end

of the automotive industry [44] [45]. A very recent development generating significant interest for both military and civil applications is a prototype contact lens that enhances night vision. University of Michigan scientists placed a thin strip of graphene that reacts to photons between layers of glass, making dark images appear brighter [46]. This technology has a long way to go before it will serve as an NVG replacement, but it will likely spur renewed night-vision-related research efforts.

Challenges and Applications

Though HMDs provide increased situational awareness to the wearer, several key technical and non-technical challenges still exist and continue to drive additional research in this domain. HMDs that interfere with normal perceptual processes may degrade situation awareness, and if the system requires an unacceptable compromise, it will not be used. Design and implementation are limited by physical constraints, safety concerns, and technical challenges. Total system weight must be limited to avoid creating neck strain for the wearer. Safety during pilot ejections limits weight and form-factor; occluded vision has limited the exploration of non-transparent displays in the cockpit and in vehicles; and cognitive tunneling and display clutter have driven several studies into symbology designs and placement.

Many of the currently fielded HMDs and NVGs have limited instantaneous FoV, as shown in Figure 5. The total binocular FoV (both eyes) for the human visual system is 200° horizontal by 130° vertical [47], but in order to meet ejection safety, head-supported weight, and CoM requirements, a 40° horizontal by 30° vertical FoV is typical. Increased FoV often comes at the expense of increased head-supported weight.

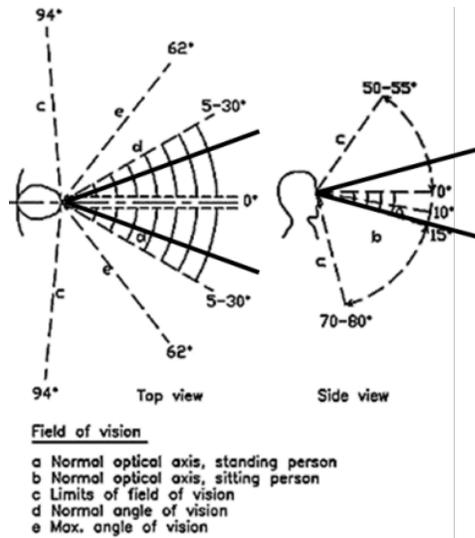


Figure 5: FoV of the human visual system³ and HMDs (shown in bold)

Warfighters have reported depth perception as a major deficiency, although it is likely the result of inadequate motion perception cues caused by the limitations in peripheral vision. Analog I² systems are further limited by their performance in weather conditions, such as dense fog. This provides significant motivation for digital imaging technology or the fusion of I² imagery with imagery obtained in wavelengths that are not affected by these weather conditions, such as certain wavelengths in the IR spectrum. It is important to note, however, that digital sensor technology by itself cannot yet provide the solution due to a sensor gap in extreme low-light conditions, driving the need for a coupled solution based on digital imaging technology and advanced image processing algorithms.

Despite these challenges, the plethora of potential applications continues to drive a need for the “Ultimate Display” proposed in 1965 by Ivan Sutherland [48]. Sutherland proposed a system in which all-powerful computers could generate graphics of objects that would behave exactly, in all sensory modes, as their real-world counterparts.

³ Human visual system image from *Accessibility for the Disabled – A Design Manual for a Barrier Free Environment*

Realization of the ultimate system is still in the future, but an increasing number of the required technologies are coming to fruition. Thus, the number of fielded systems remains relatively small due to cost, lagging technology, ergonomic shortcomings [49], a lack of applications that excite users, lack of awareness about the potential benefits, and a “visceral dislike” [50] of wearing a monitor on one’s head.

The military originally led in the application of HMDs, primarily for navigation and situational awareness applications, including helmet-mounted sights, head-position sensing for line-of-sight designation and targeting when coupled with sensors and weapons, distributed aperture sensor systems for a “windowless cockpit,” and even an operator interface for remotely piloted vehicles. Another highly popular application is simulation, training, and mission rehearsal [51] seeking to provide an alternative to large dome simulators, simulating and integrating entire environments within a single display. Other military applications include security monitoring and maintenance and inspection.

Commercial and consumer applications have seen an explosion in recent years. Commercial applications use the basic concept of an HMD as a head-up mode of presenting information, but the cost/benefit and user acceptance in this area has been relatively slow to come. The National Research Council of Canada flight-tested the use of HMDs for Instrument Approach Procedures [52]. NASA has been researching the use of head-worn displays in aviation for HUD equivalency in commercial and business jets [53], and several companies have begun to advertise future products, such as the Thales TopMax [54]. Aviation training applications, such as virtual-reality-based inspection [55], have been explored.

Uses in the medical industry have increased, including viewing pre-operative images (ultrasound, x-ray, MRI) as if looking through the patient at the internal organs for surgical planning [56] and performing surgery through natural body openings or through small incisions with the surgeon viewing indirectly through a remotely operated camera. In most medical applications, the imagery is viewed on a monitor at a distance,

but HMDs could allow increased hand-eye coordination, situation awareness and flexibility. An example system was demonstrated for this application, using computer-generated graphics integrated into HMD imagery [57]. Additional applications include molecular studies [58], virtual reality dynamic anatomy [59], airway management visualization and training for paramedics [60], and telepresence [61][62].

Consumer applications are even further burdened by the need to be extremely low cost, as users will not pay large sums of money to wear something that is uncomfortable. One of the most common consumer applications is personal gaming using head-worn displays. A 2006 entry into this market was the Trimension HMD, which was touted as the “next level of realism by offering greater immersion inside the game via an HMD” [63]. Since the entrance of the Oculus Rift VR headset [64] into the gaming market, there has been a veritable explosion of offerings. Google Glass [65] has also taken consumer applications to new heights, allowing the user to access the internet and social media, navigate, take photos and video, play music, use “hands free” interfacing with phones (e.g. texting and calling), and access a multitude of applications [66].

CHAPTER 3 TECHNICAL BACKGROUND

This chapter provides a brief overview of wavelet transforms. Given the discrete nature of digital imagery, a high-level discussion of the continuous wavelet transform is provided while a more detailed discussion of the DWT is presented. The real-valued and complex-valued wavelet transforms are compared. A review of candidate algorithms for the proposed processing chain and current techniques is provided, with an emphasis on wavelet-based methods.

Wavelet Transforms

Historically, Fourier transforms have dominated signal processing utilizing sinusoidal basis functions to characterize a signal through its frequency components. These sinusoidal signals cover the entire temporal domain, meaning Fourier analysis is inherently a global transform that cannot analyze local or transient properties. The windowed Fourier transform was developed to overcome this limitation; however, it utilizes windows of identical shape and size across the entire signal, and thus still cannot analyze transient structures.

Unlike the Fourier transform, wavelet transforms are based on wave-like oscillations of finite duration, meaning they begin at zero, increase, and then decrease back to zero. This characteristic enables analysis of both spatial and frequency information, rendering wavelets an incredibly powerful tool for signal processing. It is well known that wavelets were first mentioned in an appendix to Alfred Haar's thesis, and the Haar sequence was further detailed shortly after in a 1910 paper [67]. There are several differing views on the further history of wavelet theory with foundations in mathematics, quantum physics, and digital signal processing.

Within the field of digital signal processing, much of the foundation is attributed to the work of Stephane Mallat in the 1980s [68]. Ingrid Daubechies used Mallat's work to construct her now famous group of orthonormal basis functions that are considered one

of the cornerstones of wavelet processing [69]. Since that time, the mathematics behind wavelet theory has been studied in significant detail, is largely considered complete, and has now entered the refinement stage of generalizations and extensions. Further research lies in the application of wavelet processing to specific technical challenges in various domains.

Continuous Wavelet Transform

A wavelet function is defined as $\psi \in L^2(\mathbb{R})$, satisfying the conditions of zero mean and normalization, in the sense of the square norm, to one. The mathematical representations of these conditions are $\int_{-\infty}^{\infty} \psi(t) dt = 0$ and $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$, respectively. Wavelet-based processing has found success due to the ability to create a family of wavelets of differing time-widths from the “mother wavelet” function $\psi(t)$ through scaling and translation:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad \text{Eq 1}$$

where a is the scaling parameter controlling the compression, b is the translation parameter controlling the time location of the wavelet, and $a, b \in \mathbb{R}$ with the exception that $a \neq 0$.

Given a continuous, one-dimensional (1D) signal, $x(t)$, the continuous wavelet transform decomposes that signal using the scaled and translated wavelets by applying the following equation:

$$Wx(a, b) = \langle x, \psi_{a,b} \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt, \quad \text{Eq 2}$$

where $*$ is the operation of complex conjugate. For a wavelet transform to be invertible, it must satisfy the admissibility condition $0 < C_{\psi} < \infty$ with the admissibility constant defined as

$$C_{\psi} = \frac{1}{2} \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\zeta)|^2}{|\zeta|} d\zeta, \quad \text{Eq 3}$$

and $\hat{\psi}$ is the Fourier transform of ψ . This allows for recovery of the original signal through

$$x(t) = \int_0^\infty \int_{-\infty}^\infty \frac{1}{a^2} Wx(a, b) \frac{1}{\sqrt{|a|}} \tilde{\psi}\left(\frac{t-b}{a}\right) db da, \quad \text{Eq 4}$$

with $\tilde{\psi}(t)$ being the dual function of $\psi(t)$, sometimes defined as $\tilde{\psi}(t) = C_\psi^{-1}\psi(t)$, which satisfies

$$\int_0^\infty \int_{-\infty}^\infty \frac{1}{a^3} \psi\left(\frac{t_1-b}{a}\right) \tilde{\psi}\left(\frac{t-b}{a}\right) db da = \delta(t - t_1). \quad \text{Eq 5}$$

Discrete Wavelet Transform

While it is straightforward to mathematically describe the continuous wavelet transform, both the signal and wavelet functions must have closed forms, rendering it difficult or computationally impractical to apply. Instead, the DWT is used, with *discrete* referring to discrete sets of dilation and translation factors and discrete sampling of the signal. A commonly used method to discretize a and b was presented in literature [70], expressing the parameters as $a = a_0^j$ and $b = kb_0 a_0^j$. j is a value that affects the scaling of the wavelet transform, and k is a value that affects the translation of the wavelet. It is important to note that the translation distance of a wavelet varies with respect to the scale; thus the continuous domain parameter b accounts for the scaling factor to complete the discretization. These parameters are substituted into the wavelet function to obtain

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j} t - kb_0), \quad \text{Eq 6}$$

and the discrete version of the wavelet transform becomes

$$d_{j,k} = \langle x, \psi_{j,k} \rangle = a_0^{-j/2} \int_{-\infty}^\infty x(t) \psi(a_0^{-j} t - kb_0) dt. \quad \text{Eq 7}$$

Typically, the dyadic sampling method is used, setting $a_0 = 2$ and $b_0 = 1$, resulting in $a = 2^j$ and $b = k2^j$. Discrete child wavelets of the “mother wavelet” $\psi(t)$ are obtained through shifting and scaling by powers of two, and the wavelet function on an orthonormal basis is

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \quad \text{Eq 8}$$

and the orthonormal wavelet transform is given by

$$\langle x, \psi_{j,k} \rangle = 2^{-j/2} \int_{-\infty}^{\infty} x(t)\psi_{j,k}(2^{-j}t - k) dt. \quad \text{Eq 9}$$

The formal approach for constructing these orthonormal bases is provided in Mallat's work [68].

The idea behind multi-resolution analysis is to use the limit of successive approximations of function $x(t)$, each a smoother version of the function that corresponds to different resolutions. Once the initial resolution, J , has been selected, any $x(t) \in L^2(\mathbb{R})$ can be expressed as

$$x(t) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(t) + \sum_{j=J}^{\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t) \quad \text{Eq 10}$$

with detail or wavelet coefficients $\{d_{j,k}\}$ and the approximation or scaling coefficients $\{c_{j,k}\}$ defined using scaling function $\phi_{j,k}(\cdot)$ by

$$c_{j,k} = 2^{-j/2} \int_{-\infty}^{\infty} x(t)\phi_{j,k}(2^{-j}t - k) dt. \quad \text{Eq 11}$$

It is important to note that a recursive relationship exists between the scaling coefficients and the wavelet coefficients at successive resolutions. Using the dilation equation presented in [71], the relationship is expressed as

$$c_{j,k} = \sum_{l \in \mathbb{Z}} g_l c_{j-1, 2k-l} \quad \text{Eq 12}$$

$$d_{j,k} = \sum_{l \in \mathbb{Z}} h_l c_{j-1, 2k-l} \quad \text{Eq 13}$$

where g_l are the coefficients of a low-pass filter and h_l are the coefficients of a high-pass filter. The equations denote the approximation and detail coefficients, respectively, at level j and illustrate how they can be obtained from the approximation coefficients at resolution $j-1$. The outputs of the corresponding high- and low-pass filters are down-sampled by a factor of two to generate the sequences of coefficients. When iterating the opposite direction through the levels, the approximation coefficients at level $j-1$ are computed from the coefficients at level j through

$$c_{j-1,k} = \sum_{l \in \mathbb{Z}} g_l c_{j,2k-l} + \sum_{l \in \mathbb{Z}} h_l d_{j,2k-l}, \quad \text{Eq 14}$$

and the output of the corresponding filters are up-sampled by inserting a zero between every two samples to generate the sequence. This recursive relationship allows the DWT to be computed through a pyramid algorithm.

Thus far, the discretized wavelet transform has still been discussed in the context of decomposing a continuous signal $x(t)$. For a discrete signal $x[n]$, the DWT can be obtained by passing it through a set of filters and calculated through convolution. The signal $x[n]$ is passed through a low-pass filter with impulse response $g[n]$, providing the approximation coefficients

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k]. \quad \text{Eq 15}$$

The signal is simultaneously decomposed using a high-pass filter with impulse response $h[n]$, to obtain the detail coefficients

$$y[n] = (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]. \quad \text{Eq 16}$$

This process removes half the frequencies from the signal, allowing the results to be down-sampled according to Nyquist's rule. Thus, the output can be down-sampled by 2, to produce

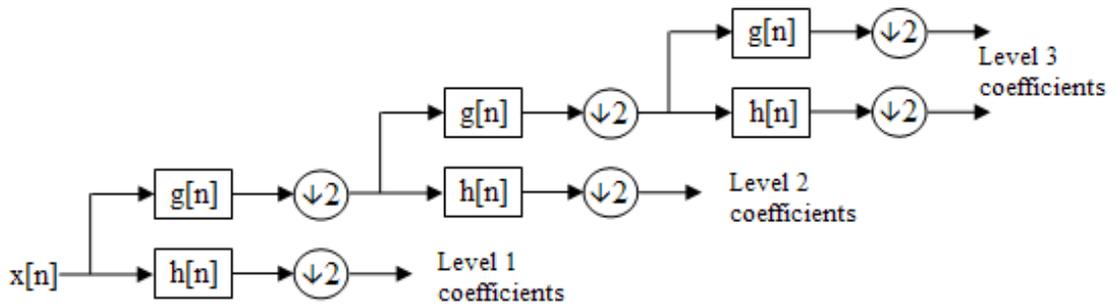
$$y_{low} = (x * g) \downarrow 2 \quad \text{Eq 17}$$

$$y_{high} = (x * h) \downarrow 2. \quad \text{Eq 18}$$

g and h must be a related pair known as a quadrature mirror filter for this decomposition to be successful. The decomposition process is often repeated to increase the resolution of the analysis. The approximation coefficients serve as the input to subsequent levels of decomposition. This filtering process is often represented using block diagrams such as those shown Figure 6.



(a)



(b)

Figure 6: (a) DWT Filter Block Diagram and (b) Multi-level Filter Bank

Finally, for two-dimensional (2D) signals such as digital imagery, the DWT is most often applied in a separable fashion to each dimension. Each 2D filter has separable horizontal and vertical components that are used to process the rows of the image first, followed by the columns. Down-sampling is applied to both dimensions of the image. After a single level of decomposition, four different frequency sub-bands are obtained that represent the different combinations of the filtering operations with the low-pass and high-pass filters on the rows and columns. The four sub-bands are low-low (LL), low-high (LH), high-low (HL), and high-high (HH) and represent the approximation, horizontal detail, vertical detail, and diagonal detail, respectively. Multiple levels of decomposition can be applied until the desired resolution is reached for the intended

application. The approximation coefficients in the LL sub-band from level $j-1$ serve as inputs to the filtering process of level j . The filter bank representation of this process and resulting frequency bands are shown in Figure 7.

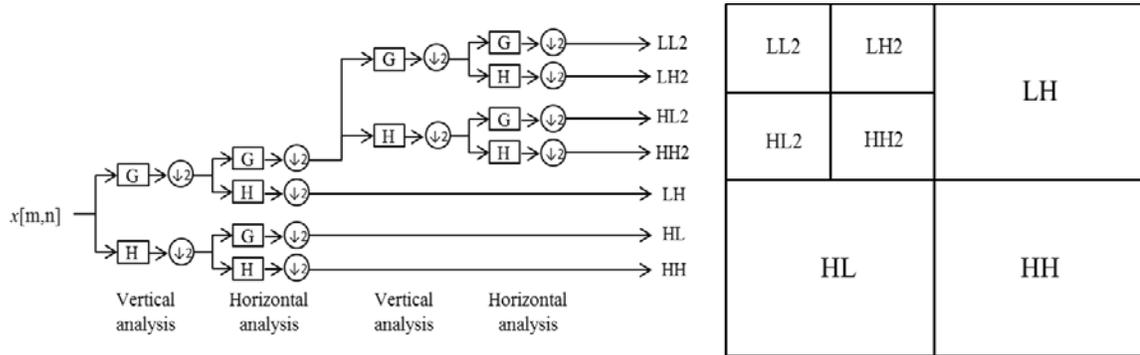


Figure 7: Two-Level 2D DWT

Real vs Complex

The previous discussion focused on wavelet transforms with real-valued filters, which have been successfully used in many image processing applications; however, the constraint of being real-valued leaves these transforms subject to several shortcomings. First, with the exception of the Haar wavelet, real-valued orthogonal wavelet filters cannot be symmetric, and symmetry is an expected property of many applications. Next, the real-valued transform is shift-variant, meaning that small shifts in an input signal will result in a dramatically different set of wavelet coefficients. Shift-variance is the result of frequency aliasing caused by subsampling and dramatically complicates processing tasks such as image registration. Increased frequency selectivity through the use of longer filter lengths can reduce the impact of this limitation, but this comes at the cost of increased computational complexity and decreased ability to analyze local properties of the signal. Finally, implementation of the 2D transform through separable wavelets

suffers from poor directional selectivity for diagonal features. The spectrum of the 2D separable wavelet is the convolution of the spectra of the 1D wavelets. Since the 1D wavelet is real-valued, its spectrum must be two-sided, and the spectrum of 2D wavelets must have support in all four quadrants. This results in an inability to distinguish between 45° and -45° features.

Real-valued biorthogonal wavelets can overcome the symmetry limitation but suffer from their own limitations. Thus, the complex wavelet transform (CWT) was developed to address the shortcomings of the real-valued wavelet transform [72]. A generalization of the standard wavelet transformation for complex numbers leads to the 1D filter bank description of the CWT in Figure 8. Complex wavelets in this sense could support approximate shift invariance and better directional selectivity but could not be perfectly reconstructed. Experimentation with complex factorizations that could be reconstructed came at the expense of shift invariance. This requires a tradeoff depending on context and application that can be impractical for many uses.

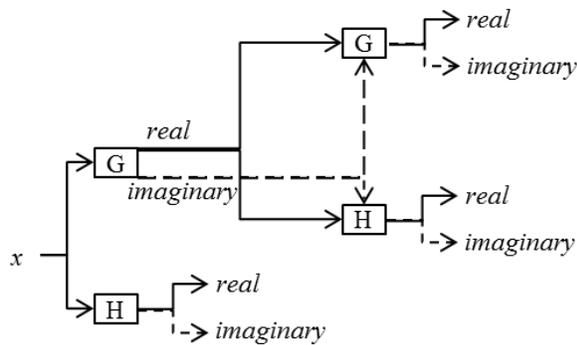


Figure 8: Two-Level "Single-tree" 1D CWT Filter Bank

The concept of a “dual-tree” complex wavelet transform (DT-CWT) was introduced to overcome this limitation [73]. The DT-CWT of a 1D signal is computed by

running the signal through two filter banks, as shown in Figure 9, with the output of the first being treated as the real coefficients and the output of the second being treated as the imaginary coefficients. The scaling and wavelet filters are designed such that each is a half-sample shift of the other and the infinitely iterated system creates complex wavelets that are Hilbert transforms of the other, meaning they are 90° out of phase with each other. In fact, the wavelets can be designed in such a way that the real wavelet has even symmetry and the imaginary wavelet has odd symmetry. Employing two filter banks computes an output that is oversampled by a factor of two, resulting in coefficients that are less aliased and have increased shift-invariance.

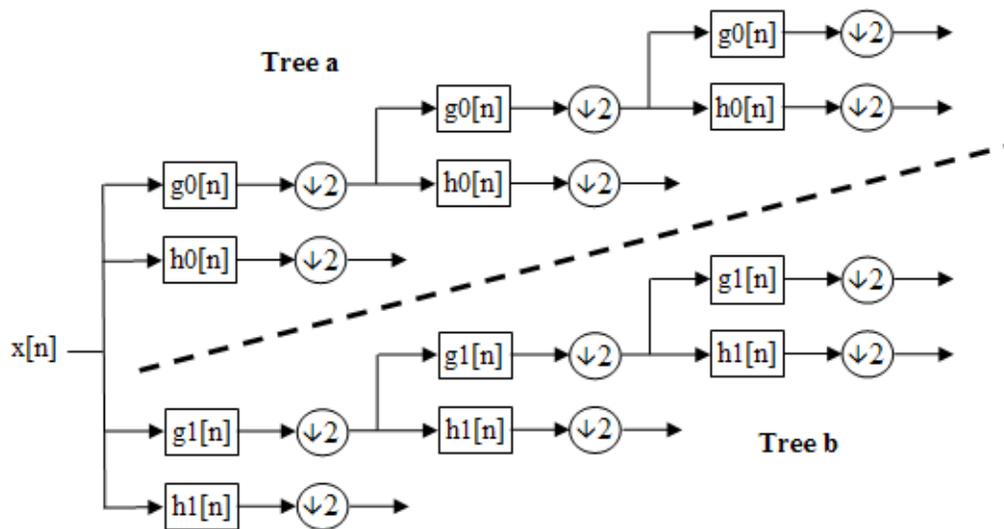


Figure 9: Three-Level DT-CWT Filter Bank

Similar to the 2D real-valued wavelet, the 2D complex wavelet has a separable implementation. Unlike the real-valued wavelet, however, the 2D dual-tree complex wavelets are not directly separable. Instead, they are constructed through the combination of several separable wavelets. This construction results in increased

directional selectivity. The conventional 2D DWT has only three angularly selective sub-bands of 0, 45, and 90 degrees. The 2D DT-DWT has six angularly selective sub-bands of 15, 45, 75, 105, 135, and 165 degrees, as well as a spectrum structure that can reduce the frequency aliasing and related artifacts due to decimation. These factors provide performance that is essentially equivalent to the steerable pyramid transform [74] used in many computer vision algorithms and has led to many DT-CWT-based image processing algorithms.

Candidate Algorithms

Figure 10 expands the *Algorithms* block of the high-level flow and depicts the basic digital processing chain applied to the target helmet-mounted vision system. A review of the recent techniques, focused primarily on wavelet-based approaches, for each of the candidate algorithms is presented here. This algorithmic overview will, when possible, be a narrative description of the processes and refrain from the use of mathematical equations. For certain algorithms, such as fusion, the abundance of active research has produced significant volumes of available resources. In these cases, more recent developments will be presented, and the reader will be referred to available literature reviews and surveys.

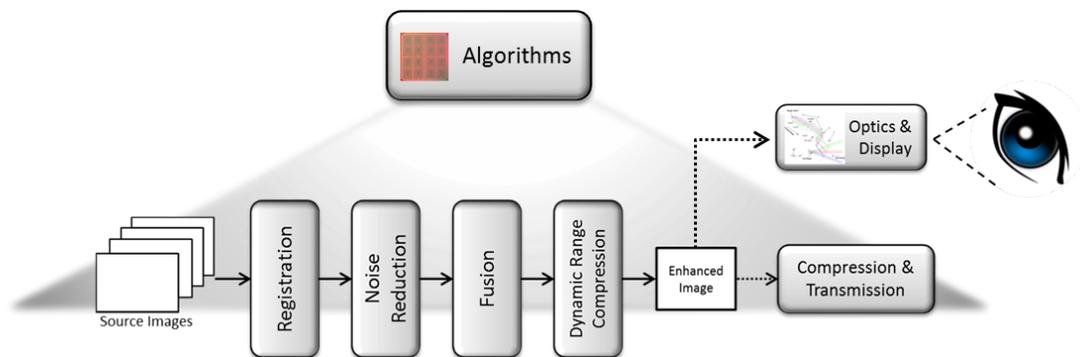


Figure 10: Representative Processing Chain

Registration

Registration is the process of establishing correspondence between, and geometrically aligning, multiple images of the same scene or object taken from differing viewpoints, at different times, or from different sensors. It is a foundational step for many algorithms, including fusion, change detection, target tracking, and the well-known application of creating photo-mosaics. For the intended application of this research, image registration is a key component of providing increased FoV for the end user through mosaicking multiple sensors and increased situational awareness in a variety of operational environments through multi-modal fusion.

At a high level, registration algorithms can be divided into two classes: area-based (also known as pixel-based) and feature-based. Fundamentally, these algorithms are comprised of four primary steps:

- A *feature detection* process that locates salient or distinctive features in an image. A few examples of descriptive features are regions, edges, line intersections, or corners.
- A *feature matching* step that establishes the relationship between features detected in the input imagery. The success of the matching step is highly dependent upon the type of features and the quality of the feature descriptors extracted during the detection step.
- A *transformation estimation* step that calculates the type and parameters of the geometric mapping used to align the images. This alignment is calculated directly from the feature correspondences determined in the previous step. As a result of this dependence on the correspondence, the accuracy of feature detection and reliability of feature matching have a significant impact on the resulting transformation.
- *Image transformation and resampling* utilizes the geometric mapping to project the images into a common reference frame and interpolates image values. The

interpolation and resampling step is required to calculate the image data value due to the fact that the transformation often results in non-integer projected pixel locations. There are many interpolation techniques that require a tradeoff between computational complexity and the resulting image quality.

Area-based methods do not directly detect features within the image, choosing instead to merge the detection and matching steps by using pixel windows of a pre-determined size to estimate the correspondence between the image frames. These approaches are highly sensitive to changes in intensity and tend to be unsuitable for multi-modal image registration.

Many state-of-the-art image registration methods are based upon multi-resolution approaches using either wavelet transformations or pyramid decompositions. In general, these techniques construct a series of sub-images of varying resolutions. Transform estimation begins at the coarsest scale and progresses to the finer scales, making corrections at each level. Initial investigations into this hierarchical strategy were driven by a desire to reduce computational complexity. These decompositions greatly decrease the search space at each level, reducing the computational load. Depending on the image content and type of motion being estimated, when combined with appropriate similarity measures, these techniques exhibit near top performance.

Registration is one of the oldest, and most profusely researched, topics in the area of computer vision. In fact, a 2003 survey of image registration methods stated, “*in the last 10 years more than 1000 papers were published on the topic of image registration*” [75], showing how prolific researchers are in this area. Given the vast body of knowledge, numerous surveys and overviews have been written, and several are summarized in Table 1. The reader is referred to these summaries for an understanding of currently available techniques. However, it is important to note that this is not a comprehensive list of image registration techniques.

Table 1: Summary of image registration literature reviews

Author(s) [Ref]	Year	Summary
Brown [76]	1992	A survey of registration methods from three research areas: computer vision & pattern recognition, medical image analysis, and remotely sensed data processing. Presents theory, methods, and characteristics of each approach to be considered when selecting or designing an algorithm for a particular application.
Maintz and Viergever [77]	1998	A survey (publications between 1993 and 1998, refers to prior reviews for methods pre-1993) focused on medical image registration techniques. Presents nine criteria for classifying registration methods and reviews the methods within the context of these criteria. Presents research trends.
Zitova and Flusser [75]	2003	An attempt to provide a comprehensive review of image registration techniques regardless of area of application. Focuses primarily on techniques post-1992 and refers to Brown's work for methods published pre-1992. Classifies the methods according to area-based versus feature-based and according to four steps of the registration process. Summarizes the trends in research and a projected outlook for future topics.
Szeliski [78]	2006	A comprehensive tutorial reviewing alignment and stitching algorithms. Provides the theoretical foundation of basic motion models underlying the registration process, describes pixel-based and feature-based algorithms, and describes the approaches for blending the images together to create seamless mosaics. Concludes with an overview of open research problems.

Table 1 Continued

Author(s) [Ref]	Year	Summary
Deshmukh and Bhosle [79]	2011	An overview of the theoretical aspects of image registration and a survey of registration techniques. Discusses the types of spatial transformations that can be recovered and divides the approaches into pixel-based, feature-based, contour-based, Mutual Information, frequency domain, wavelet-based, and Hotelling transform-based algorithms. Summarizes a comparative study of six methods described in the paper and concludes that the wavelet-based approaches offer some of the better performance.
Sotiras, Davatzikos, and Paragios [80]	2013	An overview of techniques for deformable image registration in medical imaging, where the term <i>deformable</i> refers to the fact that the images are related through a spatially varying deformation model. Algorithms in this class involve three components: a deformation model, an objective function, and an optimization method. Previous works are discussed within the context of these three components.
Mani and Arivazhagan [81]	2013	Presents a review of publications related to medical image registration. Intended as an introduction for researchers new to the field, an overview of researchers in the field, and a reference for researchers searching for literature on a specific application. Classifies the methods based on the same nine criteria proposed by Viergever and presents the publications within this context.

Noise Reduction

Image noise can be defined as the random variation in intensity or color information in imagery and is often introduced during acquisition due to imperfections in image sensors and during transmission. According to the literature [82], there are two typical sources of noise: 1) dark current noise that responds to changes in temperatures and 2) photon noise that depends on the number of electrons sampled during the imaging process and tends to be insensitive to temperature. Noise in images degrades the visual quality and affects subsequent processing when utilizing the imagery for the intended tasks, thus the advent of denoising and restoration algorithms.

Since their introduction, thousands of approaches have been developed for removing the various types of noise that can corrupt an image, and these algorithms have been the subject of numerous reviews. The selection of an existing algorithm or design of a new algorithm for reducing noise should carefully consider the type of noise most likely to affect the imagery to which the method is being applied. Thus, for sake of completeness, noise processes are briefly reviewed, including sample images originally presented in the literature [81].

- *Additive noise* is described as

$$w(x, y) = s(x, y) + n(x, y) \quad \text{Eq 19}$$

where $s(x,y)$ is the original signal, $n(x,y)$ is the noise introduced into the signal, and $w(x,y)$ is the corrupted image, with the index (x,y) representing the pixel location.

- *Multiplicative noise* is described as

$$w(x, y) = s(x, y) \times n(x, y) \quad \text{Eq 20}$$

where again $s(x,y)$ is the original signal, $n(x,y)$ is the noise introduced into the signal, and $w(x,y)$ is the corrupted image, with the index (x,y) representing the pixel location.

- *Salt and Pepper noise* is a type of impulse noise, also referred to as intensity spikes, that is attributed to errors in data transmission, malfunctioning pixel elements in a sensor, faulty memory locations, or timing errors during digitization. Salt and pepper noise has only two possible values, a and b , with the probability of each value occurring being less than 0.1. For example, in an 8-bit image, the values of the noise components are often 0 for pepper noise and 256 for salt noise.

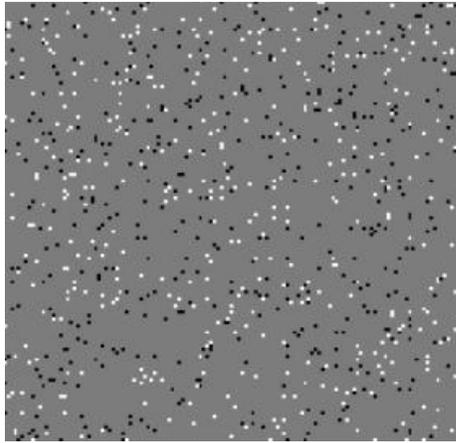


Figure 11: Salt and pepper noise with variance of 0.05

- *Gaussian noise*, as the name implies, has a Gaussian distribution and is evenly distributed across the image. Each pixel value within the noise-corrupted image is the sum of the original signal and a random Gaussian distributed noise value. This type of noise has a bell-shaped probability distribution described by the function

$$F(g) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(g-m)^2/2\sigma^2} \quad \text{Eq 21}$$

where g is the gray level, m is the mean of the function, and σ is the standard deviation.

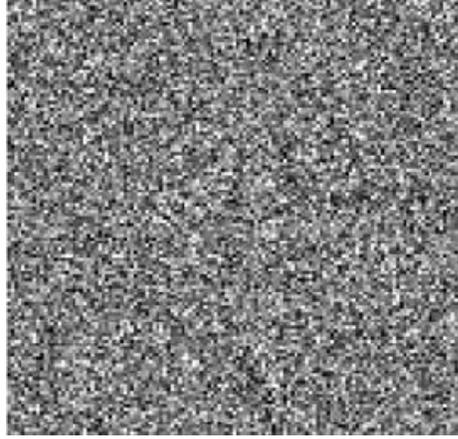


Figure 12: Gaussian noise with mean = 0 and variance of 0.05

- *Speckle noise* is a type of multiplicative noise that is common in coherent imaging systems, particularly several medical imaging modalities and Synthetic Aperture Radar, and is attributed to random interference with the returns. This type of noise has a gamma distribution described by the function

$$F(g) = \frac{g^{\alpha-1}}{(\alpha-1)!a^\alpha} e^{-\frac{g}{a}} \quad \text{Eq 22}$$

where $a^2\alpha$ is the variance and g is the gray level.

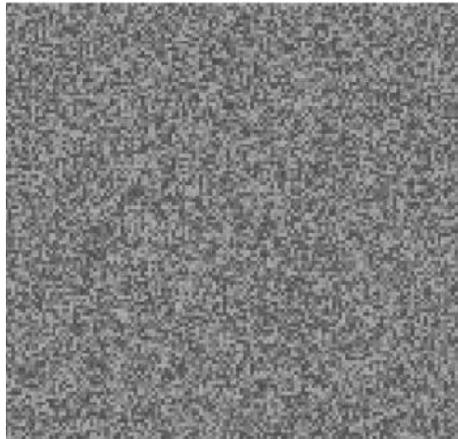


Figure 13: Speckle noise with variance of 0.05

- *Brownian noise* is considered to be fractal or $1/f$ noise and has the mathematical model of fractional Brownian motion. This is a non-stationary stochastic process with a normal distribution that is obtained by integrating white noise.

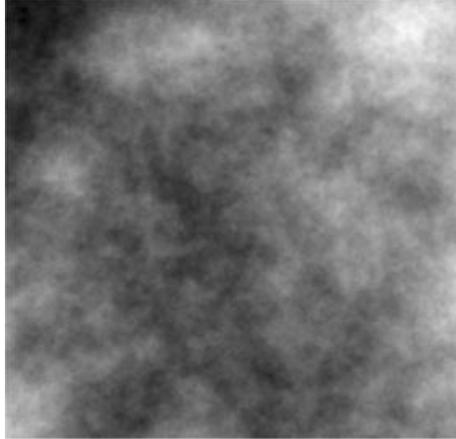


Figure 14: Brownian noise

The available literature reviews of denoising techniques take several forms. Some [83]-[85] are broad surveys of available techniques. Several overviews focus on specific classes of algorithms, such as these reviews of methods based on principal component analysis (PCA) [86], wavelet- and multi-resolution-based methods [87] [88], and several variations of median filter methods [89]. Others choose to focus on techniques that have proved to be well suited for removing specific types of noise, such as a review of techniques for removing impulse noise [90]. The reader is referred to Baudes et al. [91] for an in-depth review of algorithms available prior to 2005. These reviews provide the foundation for the following discussion of denoising methods.

Linear and nonlinear filtering techniques use convolution and the principle of a moving window. A well-known approach is the mean filter, which reduces the local variation in intensities of adjacent pixels through smoothing. Using a window of a

predetermined size (such as 3x3), the center pixel value is replaced with the average of the intensity values within the window. A simple variant of this technique is the Least Mean Square adaptive filter that varies a weighting scheme applied to the filter values by incorporating a local mean estimator. This approach better handles abrupt changes in intensity values. Both approaches tend to over-smooth the image, resulting in blurring or the loss of edges. Anisotropic diffusion is a nonlinear filtering method typically used for removing Gaussian noise that derives its name from similarities to heat diffusion equations. This technique uses partial differential equations to smooth the image while retaining awareness of flat regions and edge regions. Several diffusion models have been proposed to control the level of smoothing within the various image regions [92] [93].

A special category of nonlinear filters that adhere to the moving window principle are median filters (MF). The standard MF sorts the pixels within a window of predetermined size in numerical order and replaces the center pixel with the middle value of the sorted list. The Weighted MF (WMF) and Adaptive Weighted MF (AWMF) have various weights associated with each filter element, with the AWMF allowing the weights to adapt according to the noise content. The switching MF sought to overcome the smoothing and image degradation of typical MF designs by including a noise detection mechanism to identify “corrupt” pixels. Pixels that are identified as “corrupt” are filtered, while “uncorrupt” pixels remain unmodified. Directional MFs incorporate an impulse detector to identify noisy pixels and directional weighting based on gray level differences to allow robust denoising whether operating on an area within an image that is an edge, a flat region, or a line. Finally, Adaptive MFs employ methods to determine the local noise content and adapt the size of the filter based on the level of corruption.

Other techniques include total variation, bilateral filtering, the Field of Experts framework, and dictionary-based methods. Total variation (TV) is based on the view that noisy images have a larger discrete gradient than noise-free images, meaning noisy images are grainy and noise-free images are smooth. Denoising using TV calculates a

solution to the unconstrained minimization problem with a given Lagrange multiplier subject to noise constraints. Bilateral filtering is a non-iterative approach that combines local image values non-linearly, blending the two approaches of domain filtering and range filtering. The Field of Experts framework models image priors based on Markov random fields and denoise the image using an iterative gradient-descent approach on a negative log-likelihood term. Dictionary-based methods are patch-based algorithms that usually assume additive white Gaussian noise. These approaches denoise the patches separately and insert them into the recovered image, averaging overlapping patches.

There are numerous PCA-based techniques for image denoising. A spatially adaptive PCA denoising method was developed to operate directly on the color filtering array, preserving color edges and details. Another approach decomposes the image using adaptive principal components for denoising. Several PCA-based dictionary approaches exist. The first utilizes a hard threshold in image-specific orthogonal dictionaries that are learned from the image through PCA-based strategies. A second method uses small dictionaries for patch-based denoising that are learned through an adaptation of PCA for Poisson noise. PCA domain coefficient shrinkage with local pixel grouping preserves fine image structures while block-wise PCA computed through singular value decomposition (SVD) reduces the image root mean squared error. PCA has also been combined with the non-local means algorithm, using PCA to project image neighborhood vectors onto lower-dimensional subspace to improve computational performance.

Wavelet-based techniques operate under the following framework: decompose the image using the wavelet transform, modify the wavelet coefficients, and reconstruct the denoised image using the inverse wavelet transform. Most wavelet techniques modify the coefficients through a threshold operation. Hard threshold and soft threshold methods compare the coefficients to a predetermined τ and are considered Universal Threshold operations since a single τ is applied to the wavelet coefficients. VisuShrink [94] adheres to a hard threshold, but instead of using a predetermined threshold, τ is set proportional to

the standard deviation of the noise within the image. SureShrink [95] uses a level-dependent soft threshold to minimize mean squared error. BayesShrink [96] uses a sub-band-dependent soft threshold to minimize Bayesian risk.

Recent non-threshold-based wavelet approaches have used Weiner filters [97] or median filters [98] to modify the coefficients. Wavelet-based methods are still considered to be among the top performers for denoising. Techniques are not limited to the DWT, but also include the CWT and extensions to related processes such as Contourlets, Curvelets, and Ridgelets.

Fusion

Data fusion is not a new concept and, in fact, dates back to the 1950s and 1960s. Broadly defined, image fusion is the process of combining multiple input images into a single output image that provides more information than is present in any of the inputs individually. While fusion has typically been applied in the field of remote sensing for object recognition, classification, and change detection [99], the military applications have not gone unnoticed for both aviators [100] and ground troops. Fusion has also seen a significant surge in applications in medical image processing, as can be seen in a recent survey of the state of the art [101].

Fusion can take many forms, but the two that are of the most interest to the current research are multi-modal and multi-focal fusion. Multi-modal fusion seeks to combine imagery obtained from differing sensor modalities, such as visible and IR, often to overcome the limitations or disadvantages of each sensor. For example, visible cameras are relatively cheap and capable of producing high-quality images in nominal operating conditions, but the image quality degrades rapidly in dark, shadowed, foggy, cloudy, rainy, or smoke-filled environments. IR sensors of various wavelengths can overcome many of these environmental conditions; thus, a pairing of the two through fusion has long been desired for applications such as ground soldier vision systems.

Multi-focal fusion seeks to overcome the limitations in the depth of focus of current imaging systems by generating an “all-in-focus” image from input images with differing focus depths. This is a difficult problem due to the ambiguity in defining focus, especially in regions without texture, and removing artifacts that often appear along the edges of objects. The vast majority of the approaches in the following discussion apply to both multi-modal and multi-focal fusion algorithms.

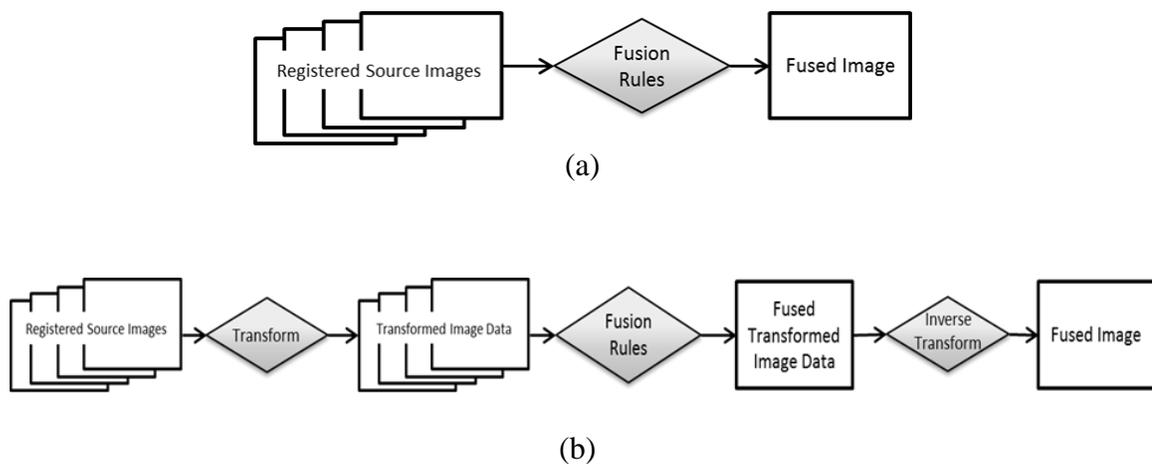


Figure 15: High-level flow for (a) spatial fusion and (b) transform-based fusion

There are many different types of fusion algorithms, often divided into two categories: those operating directly with the imagery in the spatial domain and those that utilize a transformation to operate on an alternative domain representation of the imagery, such as the frequency domain. The high-level processes for spatial fusion and transform-based fusion algorithms are depicted in Figure 15 to illustrate the differences. Spatial approaches are relatively simple and efficient; however, they are highly susceptible to registration errors and often distort spectral signals. Multi-resolution approaches, introduced in the 1980s, seemed to reduce the problems associated with traditional spatial approaches, performing well spectrally. In these techniques, a wavelet-based or pyramid-

based transform is used as a framework for decomposing the imagery. A fusion decision map is used to combine the data sets, and the inverse transform is applied to recover the fused image. The disadvantages of multi-resolution approaches include blocking artifacts and large memory requirements for pyramid representations, and the poor approximation of directional features and ringing artifacts for wavelet representations. Alternative wavelet transforms to the critically sampled DWT can be used to reduce the artifacts; however, this comes at the cost of increased memory requirements and computational complexity.

The simplest spatial fusion rule is to average the input images, but this approach reduces the contrast of the unique features contributed by the source images. More sophisticated spatial approaches to image fusion include PCA [102]-[104], intensity-hue-saturation transformation (IHS) [103], and high-pass filtering. IHS is not considered a transform-based approach because it is a “color space” transformation, not a domain transformation. Two well-known multi-resolution approaches are the Laplacian and Gaussian pyramid [105], which convolve the image with a blurring matrix and down-sample to form the next level of the pyramid. The Laplacian pyramid is formed through differencing each level of the pyramid and separating the image into the low- and high-frequency components. Once the pyramids are formed, the process is reversed, and at each level the pyramids are combined by taking the maximum from the source images. Variants include the filter-subtract-decimate (FSD) pyramid [106], the ratio-of-low-pass (ROLP) pyramid [107], and the contrast pyramid [108]. Wavelet transformations offer a more compact way to fuse images and are based on the primary concept of injecting the detail information, usually present in the high-frequency data, from each source image into the final fused image. A significant body of literature exists on wavelet-based approaches [101], and more recent research can be found in the literature [109]-[112]. In 2011, an Army Research Laboratory technical report compared 13 algorithms for multi-modal fusion, including spatial approaches, pyramid approaches, and several currently

available wavelet-based algorithms. The report found that wavelet-based approaches were among the top performers, and that the CWT approach offered performance in the top three across a wide variety of experiments and environmental conditions [113]. Multi-resolution approaches have been extended to use Ripplets [114], Curvelets [115], Contourlets [116], Slantlets [117], etc.

Additional approaches that do not fall neatly into the spatial or transformation classification include graph-based methods [118][119] and the related region-based method [120] that formulates fusion as an optimization problem, a nearest neighbor approach that measures pixel sharpness using the neighbor distance computed from the oriented distance in differential geometry [121], an SVD method [122], and the increasingly popular self-adaptive neural network approaches [123]-[125]. As is common in image processing, each approach has advantages and disadvantages. Combining multiple techniques into so-called hybrid approaches has shown to produce promising results [126]-[128].

Dynamic Range Compression

The human visual system can perceive five orders of magnitude of luminance at once and is capable of gradually adapting to over nine orders of magnitude [129]. Advances in high dynamic range (HDR) imaging [130] can capture orders of magnitude in excess of those perceptible to the human eye. Each pixel in an HDR image uses more bit depth than most displays are capable of rendering. Thus, it is necessary to apply a process called dynamic range compression (DRC), which compresses the data into a smaller dynamic range by reducing the bit depth. The goal of DRC is to perform this reduction in bits per channel while maintaining or, ideally, enhancing the visibility of the details. Early approaches simply displayed a subset of the dynamic range of an image by truncating to a set of maximum and minimum values. More recent approaches either seek to maximize the reproducibility of values based on the target display medium or

utilize models of how the human visual system perceives a scene to generate results that preserve realistic color and contrast.

The term *DRC* has frequently been used interchangeably with two other terms in the field of image processing: *contrast enhancement* and *tone mapping*. *Contrast enhancement* refers to the process that improves visual quality by enhancing the amount of color or gray differentiation between features in an image, allowing the features to be more visible by making the best use of the dynamic range of the display device. It is important to note that the key difference is that contrast enhancement algorithms do not inherently include a reduction in the number of bits per channel used to represent the image data. *Tone mapping* is the process of mapping one set of colors to another in order to approximate the appearance of HDR in a presentation medium that has a more limited dynamic range. Tone mapping and DRC are considered by many researchers to refer to the same process, and are often used interchangeably in peer-reviewed publications. The term *DRC* is inclusive of both terms from this point forward within this work.

Literature on DRC has been the subject of several fairly extensive reviews in the late '90s and early 2000s. The reader is referred to these sources [131]-[134] for an understanding of the research conducted prior to those dates. A recent comparison of various techniques for contrast enhancement in color images is found in the literature [135]. Historically, contrast enhancement and DRC algorithms have been divided into two categories, spatial domain methods and transform-based methods; however, those lines have become increasingly blurred as newer algorithms leverage techniques from both methods to compensate for individual weaknesses. The following review of recent techniques will include a third category described as hybrid methods.

Spatial domain methods include global histogram modifications such as gamma and logarithmic curve mappings [136], contrast stretching, and histogram equalization (HE). HE is a very popular technique that generates a mapping that redistributes pixel values within the histogram to stretch the peaks and compress the troughs. HE often

results in too much enhancement and contouring artifacts, driving the development of several HE-based techniques to address these issues. Lal and Chandra [137] provide an in-depth literature review of HE-based algorithms. Other spatial domain algorithms are based on filtering techniques, such as a content-adaptive bilateral filtering technique that allows both spatial similarity and range similarity parameters to adapt to the content, minimizing the haloing artifacts introduced by many filters [138].

A unique category of spatial domain methods are those that are biologically inspired. Retinex (Retina and Cortex) theory was first presented [139] to model how the eye perceives light intensities. The subsequent family of algorithms developed based on this model is applicable to contrast enhancement, DRC, shadow removal, and several medical imaging applications. An overview of Retinex implementations can be found in the literature [140]. Retinex-based approaches tend to perform very well across a variety of lighting conditions but tend to come at a significant computational cost and often introduce halo artifacts. An adaptive, multi-scale Retinex approach was developed to minimize halo artifacts [141]. Other biologically inspired techniques are based on the biological adaptation mechanism known as “shunting inhibition” and the center-surround network of cells [142] and a multi-scale luminance adaptation transform based on visual brightness functions [143].

Transform-based methods for DRC have been explored for several decades. Transform-based methods are well suited for these applications due to their ability to locate edge feature information. Wavelet techniques allow simultaneous insight into spatial and frequency characteristics. Many wavelet-based approaches exist [144], and, due to similarities in techniques, are often combined with noise reduction algorithms [145]-[150]. Other transform-based approaches include Curvelets [151], homomorphic filtering that operates in the frequency domain using the Fourier Transform [152], and an approach that uses linear and S-curve mappings on the low-frequency and high-frequency image components [153].

HE methods are simple and fast but can increase background noise and induce contouring artifacts. Biologically inspired methods have been widely accepted by the user community but are computationally costly. Transform-based methods tend to introduce blocking or haloing artifacts. To overcome these limitations, researchers have explored combinations of approaches and have seen promising results. Wavelet techniques have been combined with stochastic resonance to enhance very dark images [154]. Wavelet techniques have also been combined with HE [155] for enhancing medical imagery, with the SVD [156], and with center-surround Retinex [157]. More recently, multiple spatial techniques, particularly global tone synthesis reproduction and local histogram projection with contrast stretching, were combined in a block-wise algorithm that has shown significant results [158]. Another hybrid approach that combined multiple spatial techniques blended a modified sigmoidal function with local Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance images using inter-pixel contextual information [137].

Compression

Storing digital imagery requires significant amounts of memory, and transmitting these images requires large transfer time and/or significant bandwidth. Thus, data compression was developed to increase the practicality of storage and transmission of imagery by reducing the size of the image while preserving as much data as possible. The two most widely accepted approaches for image compression are standards created and released by the Joint Photographic Experts Group (JPEG), JPEG and JPEG2000. At a high level, both compression algorithms are based upon the same functional blocks, depicted below in Figure 16, but differ in the details of the three process steps of the “coder.”

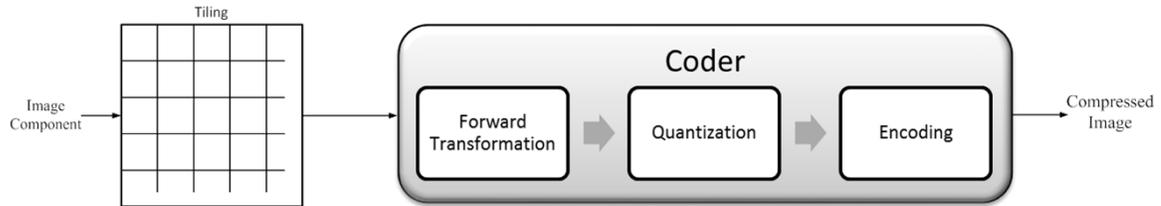


Figure 16: Compression Block Diagram

The foundation of the original JPEG compression algorithm is the discrete cosine transform (DCT) [159]. After the image is tiled into 8x8 pixel blocks, the DCT is used to convert the 2D spatial representation of each block into a frequency representation. Once the blocks have been transformed into the frequency domain, the data is quantized by reducing the high-frequency information. The value used in the quantization process controls the compression ratio. Finally, the quantized coefficients are sequenced in a “zig-zag” order and entropy coded using either Huffman coding [160] or arithmetic coding [161]. Details of these processing steps for DCT-based coding can be found in Gregory Wallace’s overview of the JPEG standard [162].

The JPEG2000 coder is based upon the DWT and, similar to JPEG compression, works on image tiles. JPEG2000 differs from JPEG in that the image is tiled into rectangular, non-overlapping blocks that are compressed independently as though they are separate images [163]. The JPEG2000 standard supports both a reversible integer-based DWT and a nonreversible DWT. Both wavelets come from the Cohen-Daubechies-Feauveau (CDF) family of biorthogonal wavelets [164]; particularly, the integer-based CDF 5/3 wavelet, also known as the LeGall 5/3 wavelet [165], and the CDF 9/7 wavelet. Once the blocks have been transformed, the wavelet coefficients are quantized to reduce the precision and bit depth required to represent the information. This is most often accomplished by dividing the magnitude of the coefficient by a scalar and rounding down. Finally, the quantized data is packet partitioned [166] to obtain

code-blocks that are independently entropy coded using a context-dependent binary arithmetic coding of bit planes called Embedded Block Coding with Optimal Truncation (EBCOT) [167].

The quantization step in both JPEG and JPEG2000 compression algorithms results in an inherently lossy process. However, JPEG2000 compression produces integer coefficients when utilizing the LeGall 5/3 wavelet, meaning the quantization step is essentially 1 and the image can be perfectly recovered. Additionally, The DCT-based JPEG standard compresses the 8x8 pixel blocks individually without reference to adjacent blocks, resulting in the “blockiness” artifacts often associated with JPEG compressed images. The shape of the wavelets used in JPEG2000 compression allows for better rendering and clearer edges within the image but can occasionally result in “ringing” artifacts around the periphery of an image [168]. These are just two of the numerous advantages provided by JPEG2000 compression.

Summary

This discussion has presented an overview of wavelet transforms and their applicability to low-SWAP, real-time embedded imaging systems. It reviewed the literature that suggests wavelet-based algorithms exist for the candidate algorithms in the desired processing chain and, in many cases, are among the top performers. Based on these findings, this thesis presents a test of the hypothesis that a wavelet-based framework for digital processing chains can enable complex, real-time applications such as helmet-mounted vision systems. The relevant information and conclusions from the existing body of literature are incorporated into the research design description that follows.

CHAPTER 4 FRAMEWORK DESIGN

Image acquisition devices, as well as image processing theory, algorithms, and hardware have advanced to the point that low-SWAP, real-time embedded imaging systems have become a reality. Much of the latency budget can be consumed during sensor integration and display scan. Thus, to be practical in a fielded application, an image processing sub-system must conduct multiple, often highly complex, tasks with minimal latency. The design and construction of such systems has to address technical challenges, including real-time, low-latency processing and fixed-point algorithms in order to leverage lowest-power computing platforms. Further design complications stem from the reality that the state-of-the-art algorithms often take very different forms. The transformation of image data between different processing domains can easily dominate the computational cost of an image processing sub-system.

This chapter presents the research design. It documents the assumptions that influenced design decisions and introduces the DWT-based processing chain designed for computationally efficient enhancement of digital imagery. It describes the benchmark processing chain designed to produce high-quality imagery that is used to support a comparative analysis of the DWT-based processing chain. The chapter concludes with the implementation details for both processing chains and the Graphical User Interface (GUI) that was developed to streamline implementation and analysis.

Assumptions

The candidate algorithm review focused on five potential algorithms: registration, noise reduction, fusion (multi-focus and multi-modal), DRC, and compression. The scope of the processing chains implemented for this research, however, is limited to denoising, fusion, and DRC. The driving application and entity utilize methods for image registration based on the mathematical characterization of the optics, camera geometry, and display system. This calibration process is used for existing HMDs and tends to

result in highly accurate alignment, while imagery-based registration techniques tend to be computationally complex and suffer the effects of parallax. Therefore, image registration algorithms are not included in either processing chain.

Compression will also be excluded from the processing chain at this time. Compression is only desirable and required if the imagery were to be transmitted from the HMD to allow others to view the imagery. Given the increasingly “net-centric” nature of today’s battlefield, this is not unlikely; however, the wavelet-based JPEG2000 compression scheme is well known and documented. Furthermore, JPEG2000 compression would have equal contribution to the computation and power calculations for both processing frameworks, with the possible exception of an additional wavelet transform for the benchmark chain.

Discussion of the various image processing algorithms has predominantly focused on algorithms that operate on spatial representations of image data and transformed representations, such as the frequency domain (or the spatial and frequency representation enabled by wavelet transforms). Another important group of algorithms are those that operate across the temporal domain, meaning processing sequences of image data captured over time by utilizing information from prior frames. Temporal processing is ideally suited for addressing artifacts that are not correlated across frames, and has shown some promise in areas such as denoising [169] [170]. However, temporal processing requires acquisition, storage, and retrieval of multiple frames of data, as well as the need to estimate and compensate for motion that occurs between frames. These limitations render temporal processing algorithms unsuitable for the low-latency and low-power requirements of the target application.

DWT Processing Chain

The first objective of this research was to develop and assess a wavelet-based framework for digital processing chains that enable complex, real-time applications such

as helmet-mounted vision systems. Since their introduction in the mid-1980s, wavelets have become a standard tool for signal and image processing. The multi-resolution aspect of wavelet processing provides a tradeoff between the spatial domain and the frequency domain. Also, the linear computational complexity and fixed-point implementations lend themselves to computation- and latency-sensitive applications, such as those explored in this work.

Wavelet Selection

The advantages and shortcomings of real-valued wavelet transforms, as well as complex-valued wavelet transforms were presented in Chapter 3. Despite the increased shift invariance, increased directional selectivity, and decreased likelihood of aliasing artifacts, the CWT was eliminated from considerations. The computational complexity and an inability to process the data “in-place” (meaning additional data storage and data access during algorithm execution) outweighed the benefits for the target application.

Several real-valued wavelets were considered when designing the processing chain, including the CDF 9/7 and LeGall 5/3 wavelets supported by the JPEG2000 compression standard. After consideration that included several factors, such as computational complexity and a desire for fixed-point implementation, the LeGall 5/3 was selected.

Table 2: LeGall 5/3 Filter Coefficients

k	Analysis Filter Coefficients		Synthesis Filter Coefficients	
	Low-pass	High-pass	Low-pass	High-pass
0	6/8	1	1	6/8
±1	2/8	-1/2	1/2	-2/8
±2	-1/8			-1/8

This is the real-valued wavelet that allows for lossless JPEG2000 compression, and is a biorthogonal wavelet constructed of symmetric filters. The coefficients of these synthesis and analysis filters used to perform the convolution-based implementation of the transform are presented in Table 2. The analysis filters have low-pass filter with length 5 and high-pass filter with length 3, hence the name of the wavelet. For the inverse transform, filter lengths are reversed.

The LeGall 5/3 wavelet lends itself to highly efficient implementation through a lifting scheme. The basic idea of lifting is to use a complementary pair of filters that allow for perfect reconstruction, and has shown the ability to provide significant reduction in memory and computational complexity. Details of the lifting scheme can be found in the literature [171] [172], but it has been demonstrated that the output of the lifting approach is identical to the output of the convolution-based implementation.

Due to the relatively simple structure of the LeGall 5/3 filters, the lifting implementation of the filter banks is quite straightforward. The decomposition of the signal, performed using the analysis filter bank, is implemented utilizing the following equations:

$$y[2n + 1] = x[2n + 1] - \left\lfloor \frac{x[2n] + x[2n+2]}{2} \right\rfloor \quad \text{Eq 23}$$

$$y[2n] = x[2n] + \left\lfloor \frac{y[2n-1] + y[2n+1] + 2}{4} \right\rfloor \quad \text{Eq 24}$$

where $x[n]$ represents the input signal, $y[n]$ represents the output (the resulting wavelet coefficients), and $\lfloor \cdot \rfloor$ is the floor operator.

The floor operation is effectively a quantization step, but it is interesting to note that, despite this quantization, the inverse transform is able to perfectly recover the original signal. As previously described in Chapter 3, the equations are applied to the rows in the image data first, followed by the columns. Subsequent levels of decomposition are applied to the resulting approximation coefficients.

The reconstruction of the signal is performed via the inverse transform using the synthesis filter bank. The inverse transform is implemented with the following equations:

$$x[2n] = y[2n] - \left\lfloor \frac{y[2n-1] + y[2n+1] + 2}{4} \right\rfloor \quad \text{Eq 25}$$

$$x[2n + 1] = y[2n + 1] + \left\lfloor \frac{x[2n] + x[2n+2]}{2} \right\rfloor \quad \text{Eq 26}$$

where $y[n]$ again represents the wavelet coefficients, $x[n]$ represents the recovered image data, and $\lfloor \cdot \rfloor$ is again the floor operator.

Algorithm Selection

The DWT-based processing chain performs denoising through an adaptive soft thresholding algorithm, fusion using an area-based fusion technique, and dynamic range compression using a self-designed method. These algorithms were selected, or designed, based on two key factors aimed at minimizing computational complexity and power consumption. The algorithms utilize simple mathematical operations that do not decompose into a computationally intractable problem, and processing can be accelerated through parallelizable implementations.

Photon noise tends to be the dominant type of noise in visual spectrum and infrared cameras [173]-[175], especially in low light conditions where the number of available electrons is inherently limited. This poor illumination during image acquisition results in uncorrelated noise with a Gaussian distribution. This Gaussian white noise will be common in the target operating environment, and there are several approaches that have demonstrated success in removing this type of noise. Denoising in the wavelet domain is based upon the premise that noise is spread throughout the small-magnitude coefficients while the image information is contained in the large-magnitude coefficients.

As described in Chapter 3, many wavelet-based algorithms are based upon some sort of threshold, such as the hard and soft threshold operators. The hard threshold

algorithm takes a “keep or kill” approach, setting the coefficients with absolute value below the threshold to zero and leaving all coefficients above the threshold unmodified. The hard threshold approach creates a discontinuity that can be clearly seen in Figure 17. This discontinuity can lead to haloing or blocking artifacts within the resulting denoised image.

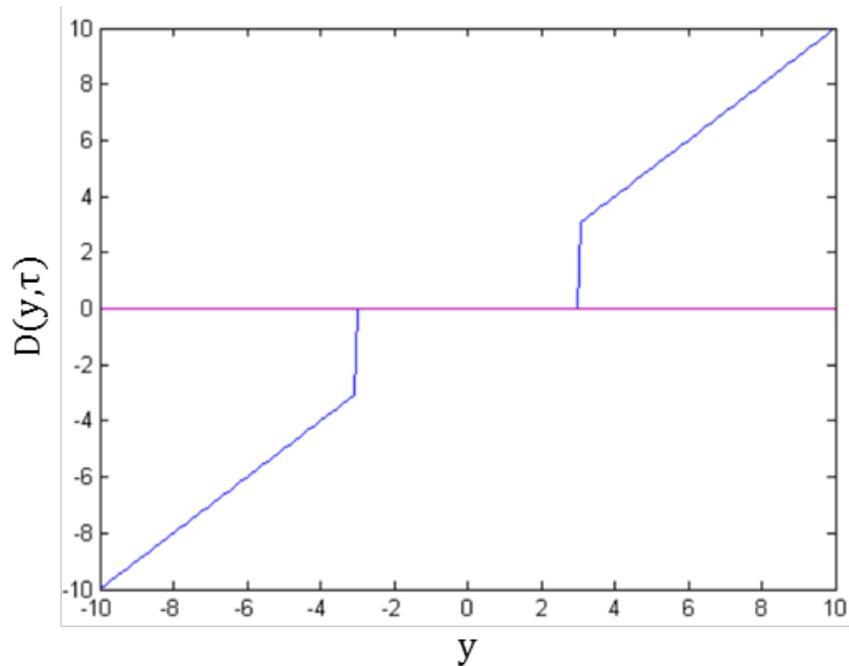


Figure 17: Example hard threshold operator for $\tau = 3$

Alternatively, the soft threshold algorithm, shown in Figure 18, “shrinks” the wavelet coefficients with absolute values above a threshold and sets those coefficients below the threshold to zero, effectively removing the noise component. There is a significant probability that subsequent processing steps would further emphasize artifacts induced by applying a hard threshold. Therefore, the soft threshold approach was selected for inclusion in the DWT-based processing framework.

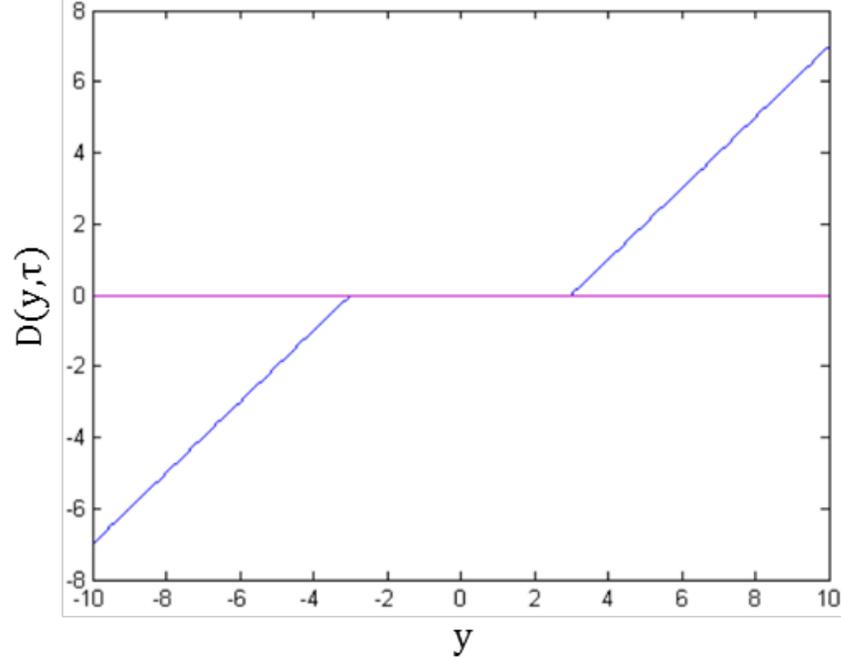


Figure 18: Example soft threshold operators for $\tau = 3$

The soft threshold operator is mathematically defined as

$$D(y[n], \tau) = \begin{cases} y[n] + \tau, & y[n] < -\tau \\ 0, & |y[n]| \leq \tau \\ y[n] - \tau, & \tau < y[n] \end{cases}, \quad \text{Eq 27}$$

where $y[n]$ represents the n^{th} wavelet coefficient in the image and τ is the threshold value. Choosing the appropriate threshold value is one of the most challenging aspects of performing image denoising in the wavelet domain. Many algorithms apply a single global threshold to the entire set of wavelet coefficients, while others apply thresholds that adapt to the coefficients in varying manners. The selected algorithm uses a sub-band adaptive threshold modified from the literature [95][176]. This adaptive threshold is

$$\tau = \delta_{mad} \sqrt{2 \log N} \quad \text{Eq 28}$$

where N is the number of pixels in the sub-band (i.e., HH) and δ_{mad} is the median absolute deviation. δ_{mad} is calculated using the following equation:

$$\delta_{mad} = \frac{\text{median}|C_i|}{0.6745} \quad \text{Eq 29}$$

where C_i can be the LH, HL, or HH coefficients for the i^{th} -level of decomposition. The scalar value in the denominator rescales the numerator to be a suitable estimate for the standard deviation for Gaussian white noise.

Visible spectrum sensors, including I^2 devices, have known limitations in operating conditions such as fog and smoke, while certain IR wavelengths can provide usable information in these operating conditions. Additionally, current generations of I^2 NVGs require the user to manually focus the scene. Soldiers have indicated that in the field they routinely forgo adjusting the focus, choosing instead to select a “good enough” focus depth and leave it unaltered for the duration of the operation. Auto-focus techniques are inadequate, as there is no way to identify the important focus depth, and they are unlikely to handle the constantly changing scene content introduced by frequent motion. Hence, it is straightforward to understand the significant role that multi-modal and multi-focus fusion play in the target application.

Similar to denoising, fusion in the wavelet domain is based on the premise that large-magnitude coefficients correspond to more acute changes in intensity and thus the salient feature data in the image. It is logical to assume that taking the per-pixel coefficient with the largest absolute value provides the best fusion result. However, features within an image are typically larger than a single pixel, and as a result, area-based fusion rules often produce better-quality images. An analysis confirmed this conclusion. The simple maximum selection shown in Figure 19 produces significant detail in the form of readable text at many depths in the image, but being highly sensitive to noise has introduced halving artifacts. The area-based algorithm selected for the processing chain, shown in Figure 20, minimizes the artifacts but does not produce as much detail in the text portion of the images⁴.

⁴ Images collected by Mr. Steven Koenck, Pr Systems Engineer at Rockwell Collins, Inc.

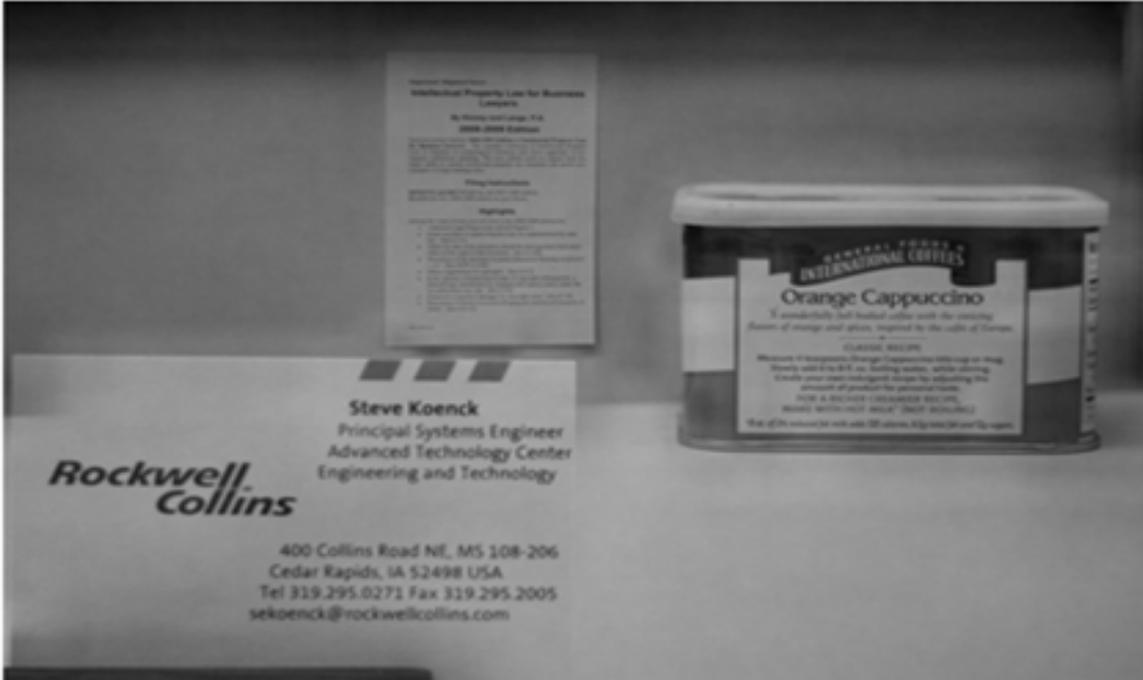


Figure 19: Visible haloing artifacts in maximum selection multi-focal fusion



Figure 20: Area-based multi-focal fusion algorithm minimizes artifacts but has lost detail

The fusion algorithm used in the processing chain is a modified version of the area-based algorithm presented in the literature [177] and is suitable for both multi-modal and multi-focal fusion applications. The fusion rule in the processing chain can be described as

$$y[n]_f = \begin{cases} \overline{y_i[n]}, & \text{if } y[n] \in LL \text{ sub-band} \\ \max_A(y_i[n], q_i[n]), & \text{otherwise} \end{cases}, \quad \text{Eq 30}$$

where $y_i[n]$ represents the n^{th} coefficient from image i and $q_i[n]$ represents the quality measure associated with the n^{th} coefficient from image i . The rule means that the approximation coefficients from all input images are averaged, while the detail coefficients subject to a modified maximum operator, \max_A , that utilizes an quality measure function $q(\cdot)$ to ensure that as much of the dominant features within the local area are included in the fused image.

The quality measure function for an individual image initializes each element of $q[n]$ with the absolute value of the detail coefficients $y[n]$ of the LH, HL, and HH sub-bands from all passes of the DWT. For each image setting,

$$q[n] = |y[n]|. \quad \text{Eq 31}$$

Then, for each subsequent pass of the DWT, from highest to lowest resolution, the calculation loops through the elements of the quality measure array (using proper bounds to ensure that only the entries associated with the appropriate pass are computed) and updates the values using equation

$$q \left[\frac{r}{2} \cdot w + \frac{c}{2} \right] = q \left[\frac{r}{2} \cdot w + \frac{c}{2} \right] + |y[r \cdot w + c]|. \quad \text{Eq 32}$$

r and c are the indices used (with proper bounds) to iterate through the image height and width values, and w is the width of the image. Indices are integers; thus, the quality measure for a lower-resolution pass is an aggregate of the absolute value of the coefficients from the corresponding 2x2 window from the higher-resolution pass before it. Once the quality measure has been calculated through all J passes of the DWT, it is

normalized based on the number of passes, and the values from the lower-resolution passes are propagated into the higher-resolution passes using

$$q[r \cdot w + c] = \frac{(q[r \cdot w + c] + q[\frac{r}{2} \cdot w + \frac{c}{2}])}{2}. \quad \text{Eq 33}$$

r and c are again the indices used (with proper bounds to ensure only the entries associated with the appropriate pass) to iterate through the image height and width values, and w is the width of the image. The quality measure for each element of a 2x2 window of the current pass will use a single quality measure value from the lower-resolution pass behind it during this propagation.

This process is further described within the context of the coefficients from three passes of the DWT depicted in Figure 21. The quality measure array is initialized with the absolute value of the coefficients shown for each pass. Then, for pass $p = 2$, the quality measure for the gray coefficient aggregates the absolute value of the 2x2 window of gray coefficients shown in pass $p = 1$. The single entry shown in pass $p = 3$ aggregates the four coefficients shown in pass $p = 2$. This recursive relationship equates the quality measure for the single coefficient shown in pass $p = 3$ to an aggregate of the detail coefficients from the four elements in pass $p = 2$ and the sixteen elements in pass $p = 1$.

The quality measures are then normalized, in the context of this example, by dividing by three. Propagating the values back up through the passes means that each quality measure value associated with the 4 elements in pass $p = 2$ becomes the average of the element's value and the quality measure value associated with the element from pass $p = 3$. Likewise, each quality measure value associated with the 4 gray elements shown in pass $p = 1$ becomes the average of the element's value and the quality measure value of the gray element from pass $p = 2$.

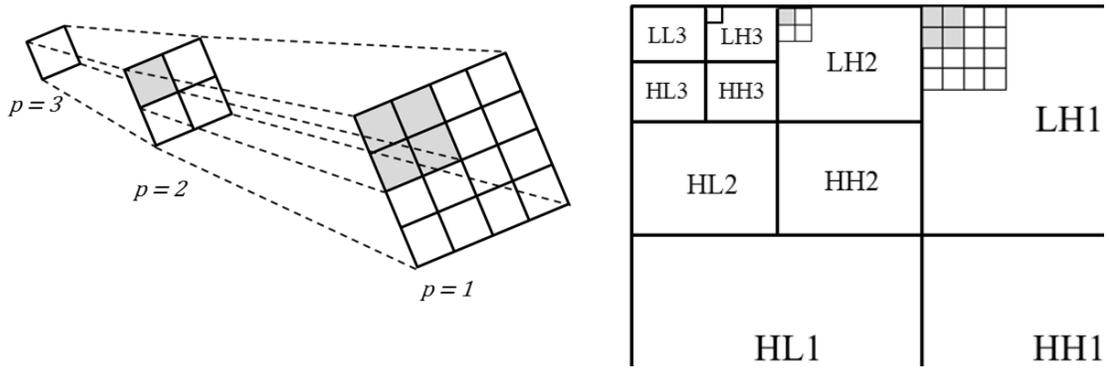


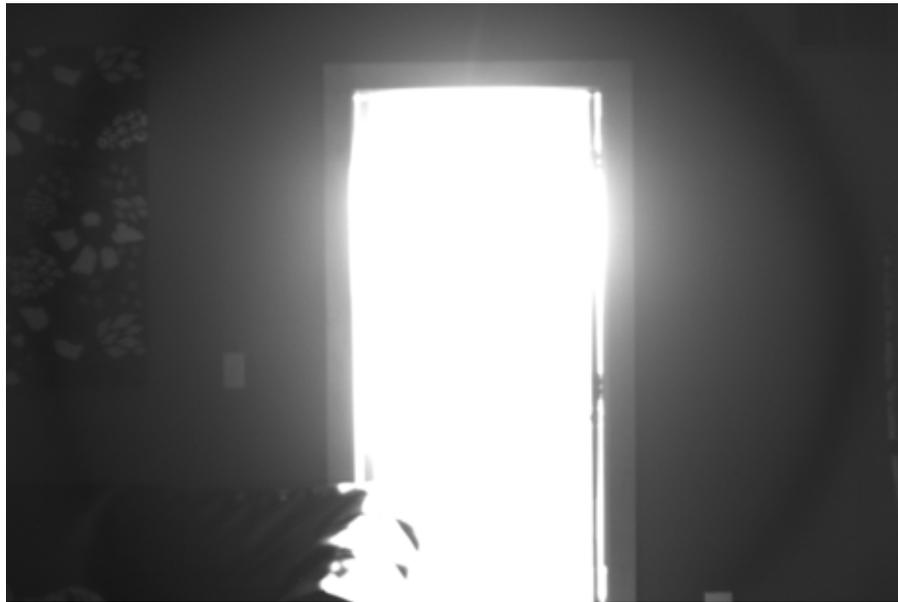
Figure 21: Two views of how corresponding pixels between passes are used for the area-based activity measure computation

The quality function is computed in this manner for all i images. Once this process is complete, max_A selects the detail coefficient $y_i[n]$ from the image that has the largest quality measure value $q_i[n]$ associated with it. This method results in the selection of spatially coherent blocks of coefficients, because the aggregation and propagation process often leads to the selection of blocks of four or sixteen coefficients mostly from a single image. Sample results for multi-focus fusion of three source images obtained from this modified algorithm are shown in Figure A.1 located in Appendix A.

DRC algorithms seek to overcome the physical limitations of display hardware and present images in a manner similar to how the scene is perceived by a human while preserving, or enhancing, important details used to comprehend the scene content. DRC is an essential processing step for this research as source imagery in the target application, as well as broader applications, may be obtained by imagers that natively capture data using different bit-depth representations. These techniques are also necessary for the intended application because the imaging system will encounter operational environments with extremes in dynamic range that cannot be displayed without losing critical detail, such as the example shown in Figure 22.



(a)



(b)

Figure 22: Sample scene with (a) original 16-bit representation and (b) 8-bit representation after a simple DRC approach that shows hidden detail

Global DRC algorithms that apply an identical mapping to every pixel, or an adaptive mapping based on global image content, are computationally efficient and often simple to implement. Local algorithms that may apply different mappings to a pixel based on regional image characteristics tend to produce higher-quality images and better compression, but at a significant increase in computational cost. Low-latency applications, including the target application, require a fast algorithm that produces realistic images across a wide variety of environments, demanding a tradeoff between computational efficiency and image quality. While many state-of-the-art wavelet-based algorithms exist [144] and yield high-quality results, they utilize additional combinations of computationally expensive processing steps such as bilateral filtering, DCT, SVD, and HE. Consequently, this processing chain includes a self-designed DRC algorithm.

The dynamic range of an image is controlled by the approximation coefficients (the LL sub-band) in the DWT representation, while the high-frequency sub-bands (LH, HL, and HH) contain the edge and feature information. Therefore, manipulating the approximation coefficients will compress the dynamic range and enhance contrast within the image. Detail within the final image can then be enhanced by manipulating the high-frequency sub-bands. These concepts were used to design a wavelet-based DRC algorithm that utilized only simplistic mathematical operations.

The non-zero approximation coefficients are manipulated using the piecewise linear function defined in Equation 34.

$$y_c[n] = \begin{cases} \beta_2 + \frac{(y[n]-\alpha_2) \cdot (\beta_3 - \beta_2) + (\beta_3 - \beta_2 - 1)}{\max(|C_{LL}|) - \alpha_2}, & \alpha_2 < y[n] \\ \beta_1 + \frac{(y[n]-\alpha_1) \cdot (\beta_2 - \beta_1) + (\beta_2 - \beta_1 - 1)}{\alpha_2 - \alpha_1}, & \alpha_1 < y[n] \leq \alpha_2 \\ \frac{\beta_1 y[n] + (\beta_1 - 1)}{\alpha_1}, & 0 < y[n] \leq \alpha_1 \\ \frac{\beta_1 y[n] - (\beta_1 - 1)}{\alpha_1}, & -\alpha_1 \leq y[n] < 0 \\ -\beta_1 + \frac{(y[n] + \alpha_1) \cdot (\beta_2 - \beta_1) - (\beta_2 - \beta_1 - 1)}{\alpha_2 - \alpha_1}, & -\alpha_2 \leq y[n] < -\alpha_1 \\ -\beta_2 + \frac{(y[n] + \alpha_2) \cdot (\beta_3 - \beta_2) - (\beta_3 - \beta_2 - 1)}{\max(|C_{LL}|) - \alpha_2}, & y[n] < -\alpha_2 \end{cases} \quad \text{Eq 34}$$

In this equation $y[n]$ are the input coefficients; α_1 and α_2 are breakpoint values in the segments representing the domain (input) of approximation coefficients; $\max(|C_{LL}|)$ is the largest absolute value approximation coefficient; and β_1 , β_2 , and β_3 are breakpoint values in the segments representing the range (output) of approximation coefficients. Identifying the optimal value for the breakpoints, α and β , presents a significant challenge, as it has empirically shown to be scene dependent. An example plot of the first 10,000 positive values of this piecewise linear function is shown in Figure 23.

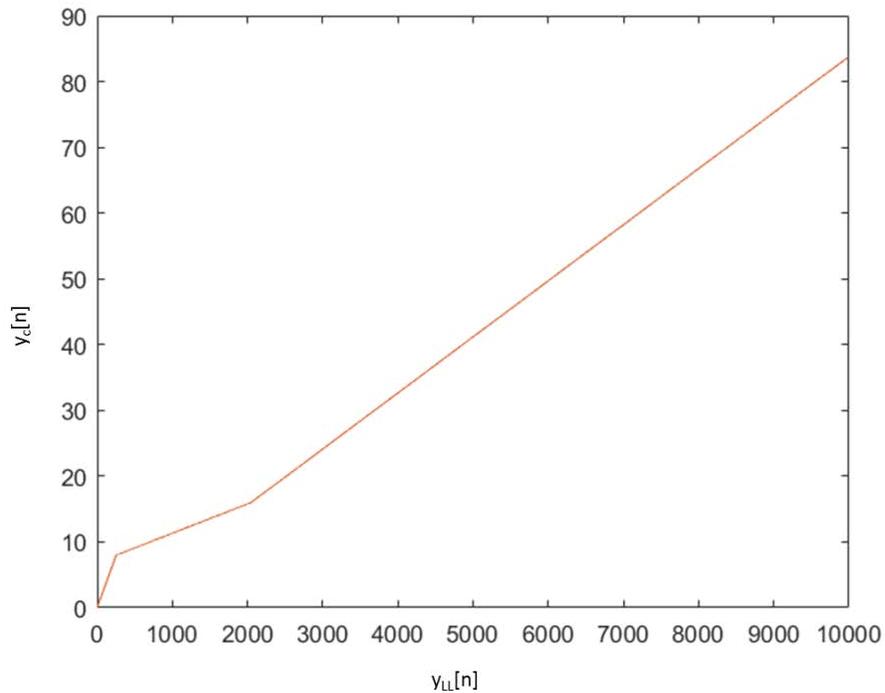


Figure 23: Example plot of the resulting $y_c[n]$ for $y_{LL}[n] = [1:10,000]$ when $\alpha_1 = 256$, $\alpha_2 = 2048$, $\max(|C_{LL}|) = 30196$, $\beta_1 = 8$, $\beta_2 = 16$, and $\beta_3 = 256$

This piecewise linear function is applied only to the initial approximation coefficients at the lowest resolution level of the DWT. It is not recursively applied throughout the reconstruction. It is also important to note that this function preserves the sign of the

coefficient values. In the context of real numbers, a change in sign equates to a change in phase and will have significant impacts when reconstructing the image.

The high-frequency sub-bands are subject to a simple scaling factor, μ , which is computed using the following equation:

$$\mu = \frac{\max(|C_{hf}|)}{2^{(bitdepth-1)}} \quad \text{Eq 35}$$

where $\max(|C_{hf}|)$ is the maximum absolute value coefficient present in the detail coefficients at all resolutions and $bitdepth$ is the desired bit-depth representation of the image. The power of the denominator is $bitdepth-1$ because the positive and negative values each represent half of the range, and $2^{(bitdepth)} \div 2$ is equivalent to $2^{(bitdepth-1)}$. An adaptive scalar based on the maximum absolute value coefficient per resolution level was also explored, but did not produce significant improvements in the resulting images. An example of this is shown in Figure A.2 in Appendix A. In certain examples it over-smoothed the image, blurring details instead of enhancing them.

Benchmark Processing Chain

The initial implementation of the benchmark processing chain performs denoising via the MF algorithm, the same area-based fusion technique from the wavelet processing chain, and a logarithmic tone mapping DRC algorithm. The algorithms were selected based on several factors, including: applicability to the imagery that would be captured by a day- and night-capable helmet-mounted vision system, frequency of use within the image processing community, and the results of SME assessments of visual quality produced by individual algorithms. Computational complexity was a consideration for this processing chain; however, it was secondary to algorithms that result in high visual quality to emphasize the tradeoffs that may be required in designing such systems.

Spatial filters can be effective in reducing additive white Gaussian noise; however, most are smoothing techniques that, while effective at removing noise, tend to

blur edge details. The MF is frequently used in many applications because, in the correct conditions, it will preserve the edge details during denoising. When using a fixed window size, the median filter is effective at removing small to moderate levels of noise [178]. It is also effective in removing speckle noise and salt and pepper noise [179], but exhibits poorer performance for images with high levels of noise. While the algorithm can have significant computation cost if implemented by sorting every entry in the window for every possible window in the image, it can be accelerated using selection algorithms or histogram medians [180].

The standard MF sorts pixels within a predetermined window size in numerical order and replaces the center pixel with the median (middle) value of the sorted list. The implementation of the algorithm for this research filters the image using 25 elements in a 5x5 window, except on the boundaries of the image. The outer edges of the image (the first and last rows and columns) are not filtered at all. The second and second to last rows and columns are filtered using 9 elements in a 3x3 window. These filter sizes are shown in Figure 24.

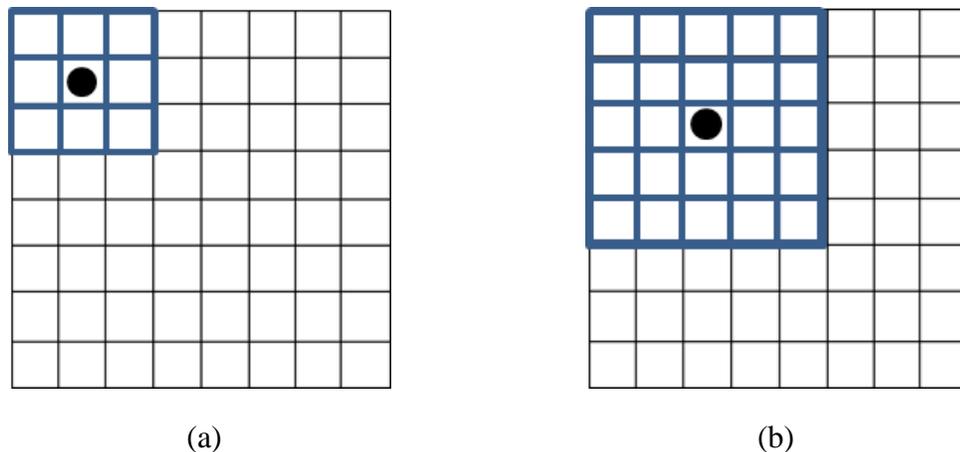


Figure 24: (a) 3x3 pixel and (b) 5x5 pixel filter masks of the median filter algorithm

The fusion algorithm selected for this processing chain is the same area-based fusion technique utilized in the wavelet processing chain. Therefore, it will not be described again here.

Several experiments [130], [181], [182] evaluated and compared DRC algorithms in terms of visual preference, realism, similarity of the compressed image to the HDR image, visibility, and reproduction of detail. Due to the varying results of these analyses, a list of five algorithms representing a diverse list of global, local, biologically based, and filtering approaches were selected for a subjective analysis by SMEs. The algorithms were selected based on factors including computational efficiency, retention and reproduction of detail, strength of compression, availability of software to facilitate a rapid evaluation, and performance in the referenced experiments.

The five algorithms included histogram adjustment [129], adaptive logarithmic mapping [136], gradient-domain compression [183], fast bilateral filtering [184], and a biologically based compression algorithm developed by Teledyne Scientific [185]. The SME assessment performed a pair-wise comparison of images produced by each test algorithm applied to three different scenes (see Figure A.3 - Figure A.5 in Appendix A) containing various characteristics desired by the target application. The algorithm selected for the processing chain [136] consistently performed in the top two across the scenes presented during the SME evaluation. Furthermore, it performed the best at reproducing detail in dark regions [181] and was second for similarity to the HDR scene [130]. The Drago algorithm [136] is a logarithmic tone mapping that uses a bias power function

$$bias_b(t) = t^{\frac{\log(b)}{\log(0.5)}} \quad \text{Eq 36}$$

to perform adaptive adjustment of the logarithmic base depending on each pixel's radiance. The original algorithm design experimentally established a default bias parameter value of $b = 0.85$; however, values ranging from 0.7 to 0.9 also produced

perceptually high-quality images. The tone mapping step computes the resulting display value for each image pixel through the following equation:

$$L_d = \frac{L_{dmax} \cdot 0.01}{\log(L_{wmax} + 1)} \cdot \frac{\log(L_w + 1)}{\log\left(2 + \left(\frac{L_w}{L_{wmax}}\right)^{\frac{\log(b)}{\log(0.5)}} \cdot 8\right)} \quad \text{Eq 37}$$

where L_w is the luminance value for the pixel, L_{wmax} is the maximum luminance value of the scene, and L_{dmax} is the maximum luminance capability of the display. The default value for L_{dmax} is 100 cd/m^2 , a common reference value for CRT displays. L_w and L_{wmax} are pre-scaled by L_{wa} , a “world adaptation luminance” that scales the scene luminance to the output image brightness. L_{wa} is computed using the logarithmic average of the luminance values in the entire scene. If a bias parameter other than the default is used, L_{wa} is further adjusted via the equation

$$L_{wa} = L_{wa} / (1 + b - 0.85)^5 \quad \text{Eq 38}$$

to maintain a constant brightness impression and realism while enhancing contrast.

Subsequent to tone mapping, a transfer function similar to gamma correction is applied to compensate for display non-linearity. Traditional gamma functions exhibit a steep slope near the origin, resulting in drastic mapping of darker pixels that reduces contrast and detail in shadowed areas. The transfer function applied in the Drago algorithm uses a simple linear fit near the origin and frequently used γ values to enhance contrast and details in dark areas. The function is:

$$E' = \begin{cases} slope \cdot L, & L \leq start \\ 1.099L^{\frac{0.9}{\gamma}} - 0.099, & L > start \end{cases} \quad \text{Eq 39}$$

Where $start$ is the abscissa at the point of tangency, and $slope$ is the slope of the line passing the origin and tangent to the curve. For this research, the parameters $start = 0.018$, $slope = 4.5$, and $\gamma = 2$ were selected, equating E' to the international standard ITU-R BT.709 transfer function.

Implementation

Framework Implementation

The processing chains have been designed to support images with varying bit-depth representations. Each input image is first checked for bit-depth, and then, if required, the pixel values are scaled to convert the image to a 16-bit representation. Limiting the implementation to 8-bit images would be representative of the data collected by many of the sensors utilized in the target application, but it was highly desirable to design and implement the frameworks in a manner such that they can easily accommodate a range of imaging modalities and bit-depth representations. The current implementations support only grayscale image data. It would also be simple to extend the frameworks to accommodate color images by applying known color-scale-to-intensity conversions, but the current implementation does not include this functionality at this time.

Many of the algorithms were initially prototyped in MATLAB to support an early analysis of the feasibility of the DWT framework concept. This approach supported rapid prototyping by leveraging the powerful ability of MATLAB to manipulate matrices and display images. Subsequently, the processing chains were re-implemented in the C++ coding language to support the analysis of computational cost and power consumption. The algorithms were implemented, tested, and analyzed using Ubuntu within a Virtual Machine on an Apple MacBook Pro[®] with a 2.8 GigaHertz (GHz) processor and 4 GB of RAM. It is important to note that this means the analysis is based on the processing chains running on the Central Processing Unit (CPU) of the machine, as opposed to a Graphics Processing Unit (GPU) designed to rapidly and efficiently manipulate images.

The processing flow for each of the frameworks is presented in the following figures. The DWT-based framework is presented in Figure 25. Figure 26 depicts the benchmark processing framework.

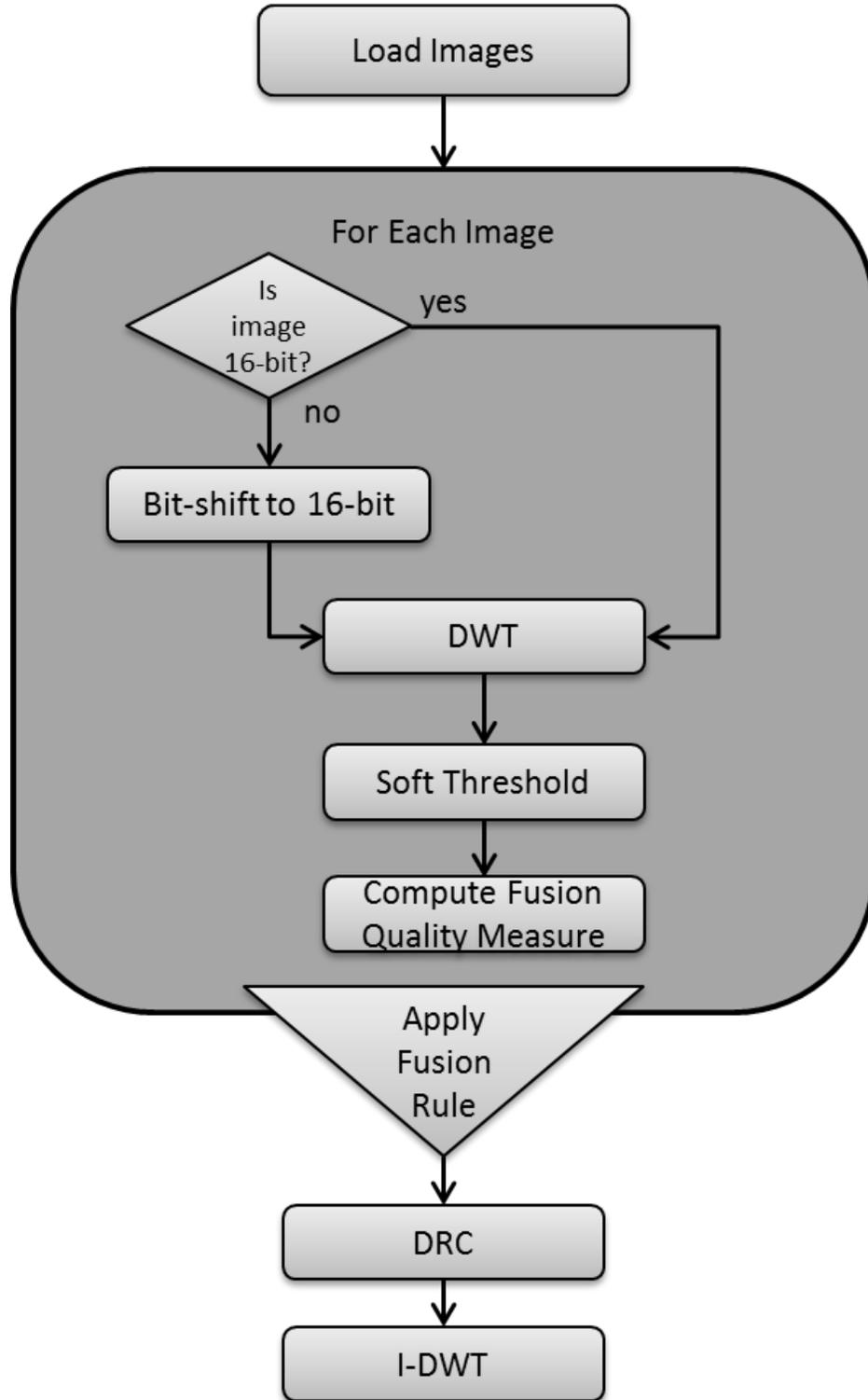


Figure 25: Algorithm flow for DWT-based processing chain

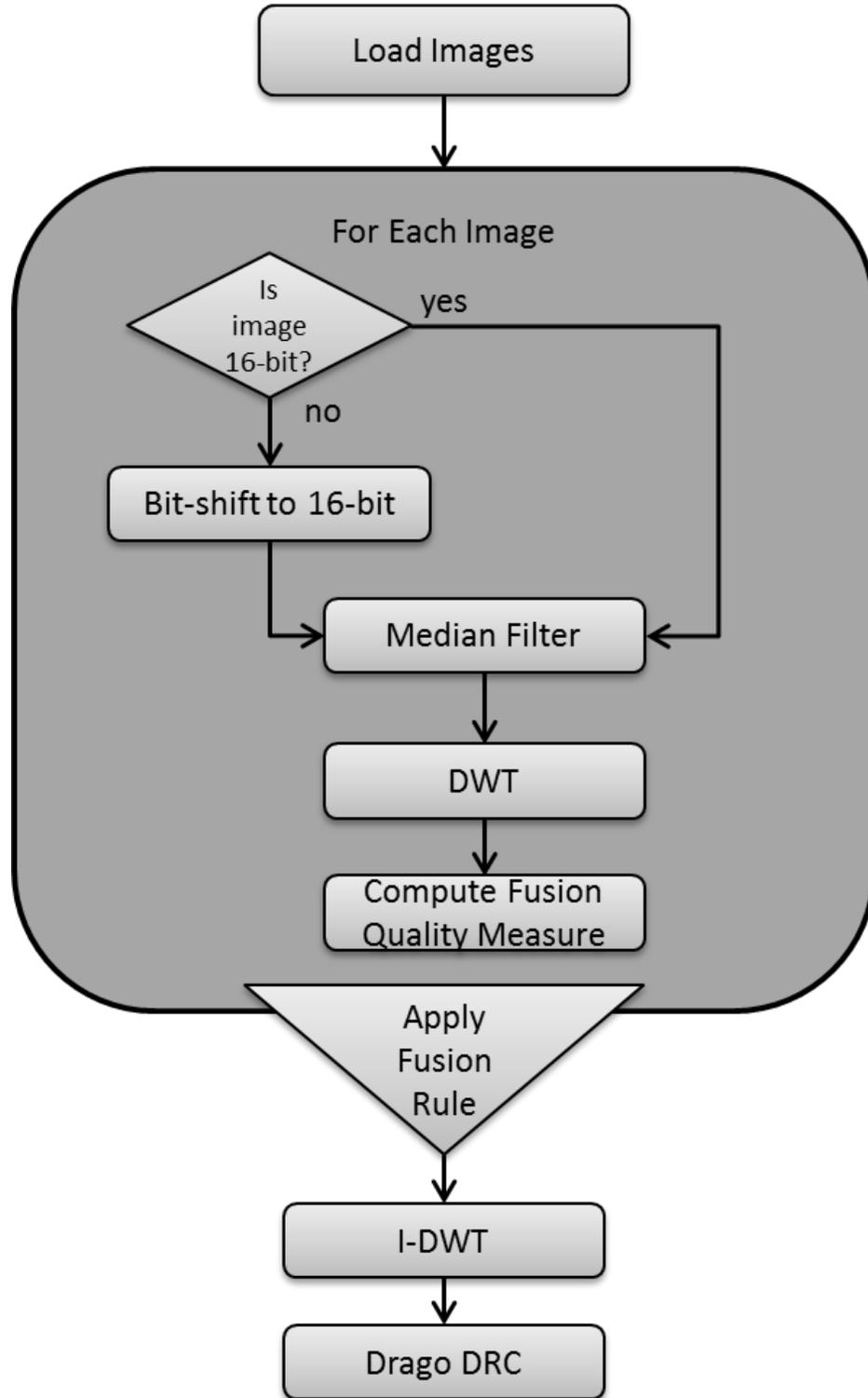


Figure 26: Algorithm flow for benchmark processing chain

Graphical User Interface

A GUI was developed in Qt to streamline the implementation and analysis of the processing chain. Qt is a cross-platform application and GUI development framework that can be used for desktop, embedded, and mobile development [188]. While the framework is written in C++, it can also be used in other programming languages through language bindings. Standard C++ source files are generated by a preprocessor that extends the language with a variety of features to simplify event handling. This capability allows the framework, and the applications or libraries that use it, to be compiled by any standard compliant C++ compiler. The framework supports a multitude of platforms, including Windows, Linux, OS X, Android, iOS, and VxWorks. It is available under GPL 3.0, LGPL 3.0, and LGPL 2.1 free software licenses, with certain modules only available under a GPL license. Qt is also available under commercial licenses that enable development of proprietary applications without restrictions on licensing.

Qt was developed by Haavard Nord and Eirik Chambe-Eng. It became publicly available in May 1995 and grew rapidly from there. The conceptual foundation for the framework was born from a discussion regarding the need for an object-oriented display system. Nord and Chambe-Eng were developing a database application that had a requirement for running with a GUI on Unix, Macintosh, and Windows platforms. Qt 0.90 was uploaded to sunsite.unc.edu on May 20, 1995, with subsequent versions (1.2 and 1.3) in 1997. Qt 2.0 was released in 1999 with a new open source license. In 2000 Qt/Embedded expanded the framework for embedded Linux devices, and Qtopia provided the application framework for mobile devices. Qt 3.0 came on the scene in 2001, providing over 40 new classes that improved Unicode support, new text viewing and editing capabilities, and a regular expression class. Qt 4.0, released in 2005, added significant new functionality including new template containers, Unicode text viewing and editing, and a 2D painting framework, as well as smaller enhancements to previously

available features. Qt 5 was released in 2012, providing substantial advances in speed and ease of use through hardware-accelerated graphics, QML, and JavaScript.

Two versions of a GUI were implemented in Qt 4. The initial GUI implementation, shown in Figure 27, allowed the user to quickly load the source images to be processed and select either the DWT-based framework or the benchmark framework. The source images were displayed, and the resulting image was displayed upon completion of the selected framework. The GUI also presented the overall computational time required to execute the processing chain. While it did not display the processing time for each individual algorithm, this data was stored during execution to support the computational cost analyses.

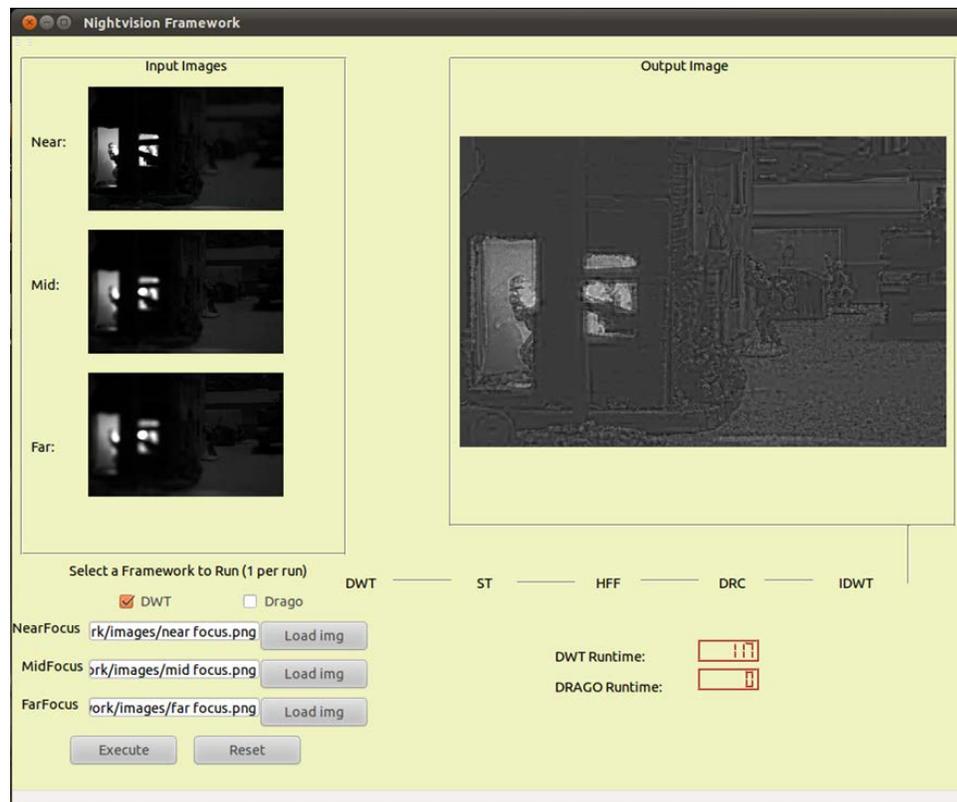


Figure 27: Preliminary Qt 4 GUI

Later in this research effort, the preliminary implementation of the GUI was expanded to support exploration of additional alternatives in the processing chains. The expanded GUI, shown in Figure 28, provides several enhanced capabilities. After preliminary analysis of the image results identified the DRC algorithm as the source of unacceptable artifacts, additional scaling-based DRC approaches were implemented, and the GUI supports exploration of various “breakpoints” in the scaling curves. It also includes an alternative wavelet processing chain based on the DT-CWT and enhancements that streamline loading source imagery.

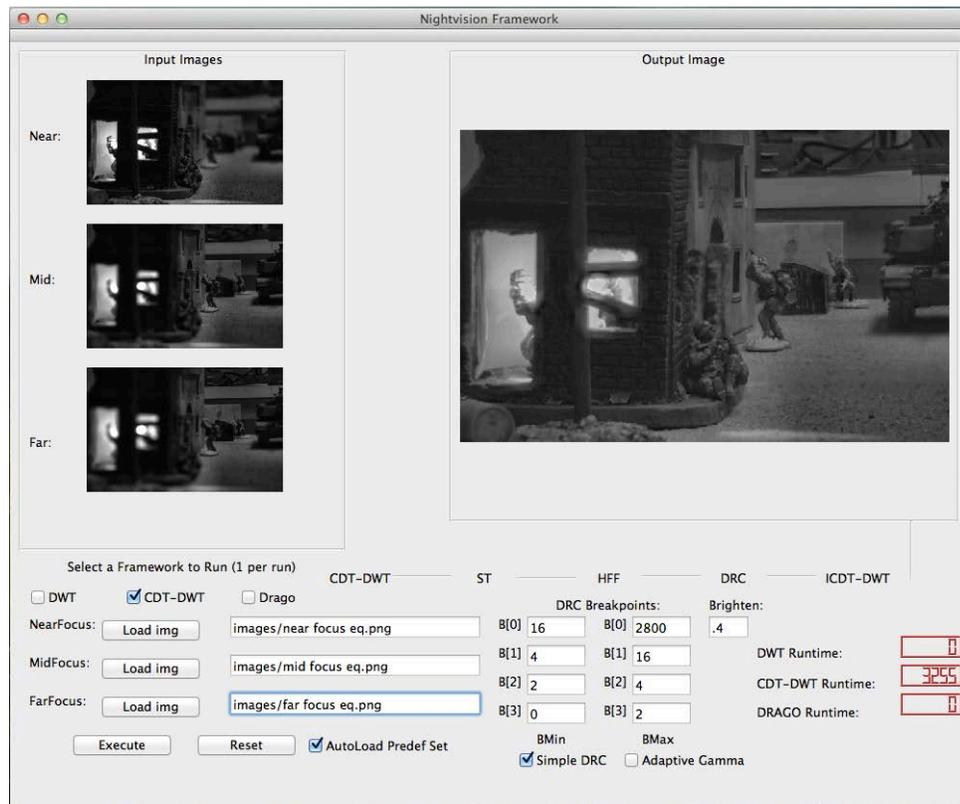


Figure 28: Expanded Qt 4 GUI

CHAPTER 5 RESULTS, ANALYSIS AND DISCUSSION

The importance of this research for the given application lies in the hypothesis that performing most, if not all, of the processing steps in a single domain will provide a significant reduction in computational latency and power consumption. It is well understood within the image processing community that image transforms are computationally expensive, and the wavelet-based processing chain described in the previous chapter minimizes the number of transforms required to enhance the imagery. It eliminates the need to repeatedly transform the imagery in and out of multiple domains, leading to a significant reduction in computational cost and power consumption. Testing this hypothesis requires the quantification of the resulting image quality and characterization of the savings in computational cost and power consumption.

This chapter presents the results of this research effort. It provides the image results for each of the processing chains and the analysis of quality, computational cost, and power consumption. Each of the approaches utilized for quantifying the results are briefly described before introducing and discussing the results of the analyses. Finally, the chapter concludes with a discussion of the implications that the algorithm selection and ordering has with respect to output image quality and the computational cost and power consumption. This discussion presents a set of guidelines to consider when implementing the processing chain for a specific application.

Image Quality Analysis

The increasing application of digital imaging technologies to areas such as medical imaging, surveillance systems, and military vision systems has emphasized the importance of accurate image quality assessment (IQA). This need is well described by a series of questions presented in the literature [186]:

“...no matter what image/video processing problems we are working on, the same issues repeatedly come up – How should we evaluate the images generated from our algorithms/systems?”

How do we know our algorithm/system is creating and improvement between the input and output images, and by how much? How can we know one algorithm/system performs better than another, and by how much? What should be the quality criterion for which the design of our algorithms/systems should be optimized?”

Both subjective and objective IQA are incredibly challenging topics, and selecting the appropriate method must be linked with the specific application.

The human visual system is the final receiver in most image processing applications. This means that subjective analysis is an accurate and reliable solution for IQA, but these methods have several limitations. Subjective assessment is time consuming and expensive, especially when analyzing large datasets. They are inherently an offline process that cannot be applied in real time to adjust algorithm parameters. Finally, results depend greatly on the physical and mental state of the reviewers, as well as how the images are being displayed and in what lighting conditions. These limitations have led to the design of mathematical (objective) methods for quantifying the perceptual quality of images.

Objective IQA has been the subject of much research over the past several decades. Mean Squared Error (MSE) and peak signal-to-noise ratio (PSNR) are two metrics frequently used to assess image quality when designing new algorithms, but these metrics are known to provide misleading results and lack a correlation to human perception. Metrics based on models of human perception and correlation to subjective assessments, including modulation transfer function area (MTFA) and square root integral (SQRI), have limitations in appropriately weighting different types of artifacts that can be introduced when processing digital imagery. Furthermore, there is direct evidence that shows many existing IQA methods demonstrate significant degradation in performance when applied to large datasets containing real-world images [187].

User acceptance is a critical factor for developing a digital vision system for the target application, and the quality of the resulting imagery is a significant factor of user acceptance, as poor quality imagery may interfere with cognitive processes. There are

many problems with defining objective measures for complex image processing applications such as the proposed processing chain, as it is difficult to define an ideal composition image that has been subject to multiple algorithms and manipulations with vastly differing intents. The aspects that determine utility to a user are application- and context-dependent in a way that numerical representations cannot always predict and often do not generalize across all possible ways in which a system can be used. Interpretation by the human is the most critical factor; therefore, the IQA performed for this research is a subjective analysis conducted by a variety of SMEs.

Image Results

This section presents a collection of image results for both the benchmark and wavelet-based processing chains from several of the data sets processed during this research. The primary series of results presented here come from Test Scene 1, the most frequently processed data set, which was used heavily while implementing the processing chains and designing the wavelet-based DRC algorithm. It contains many characteristics that are present in a wide range of the operational conditions an image processing subsystem would encounter in the target application, and image frames were collected in several different lighting conditions. A set of three sample input images from this scene, captured with different focal depths in “mid” lighting, are shown in Figure 29.

The output of the benchmark processing chain is shown in Figure 30, and the output of the wavelet-based processing chain is shown in Figure 31. Both of these figures had a histogram equalization algorithm applied after the processing was complete to enhance the differences in output quality. Figure 32 shows the original output of the wavelet chain without the extra histogram equalization step. Note that the additional histogram equalization processing step exposes gridding artifacts in the images.

Figure 33 and Figure 34 are included in the series of outputs from the wavelet-based processing chain to demonstrate the impacts that the parameters of the piecewise

linear function of the DRC algorithm has on the final results. The first figure shows the results of improperly setting the α parameters for the input range of coefficient values. The second figure shows the results of improperly setting the β parameters that control the output range of the coefficient values. Results from each of the processing chains are shown in Figure 35 through Figure 37 for several additional data sets analyzed while conducting this research.

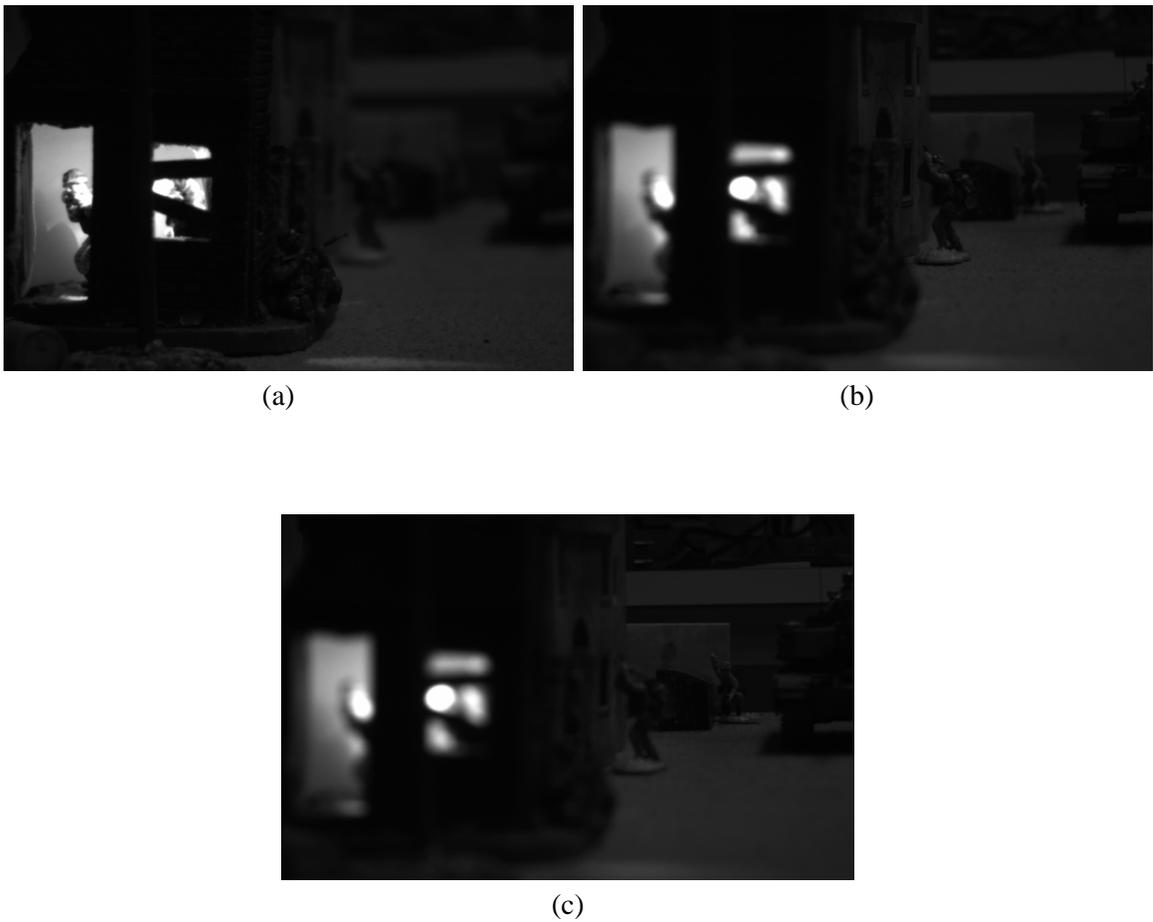


Figure 29: Test Scene 1 – Mid input images with (a) near, (b) mid, and (c) far focus



Figure 30: Output from benchmark framework after applying additional histogram equalization



Figure 31: Sample output of DWT-based framework after applying additional histogram equalization



Figure 32: Sample output of DWT-based framework with no histogram equalization



Figure 33: Sample output of DWT-based framework with modified β values in DRC



Figure 34: Oversaturated output of DWT-based framework from modified α values in DRC



(a)



(b)

Figure 35: Sample Diorama results from (a) benchmark chain and (b) wavelet chain



(a)

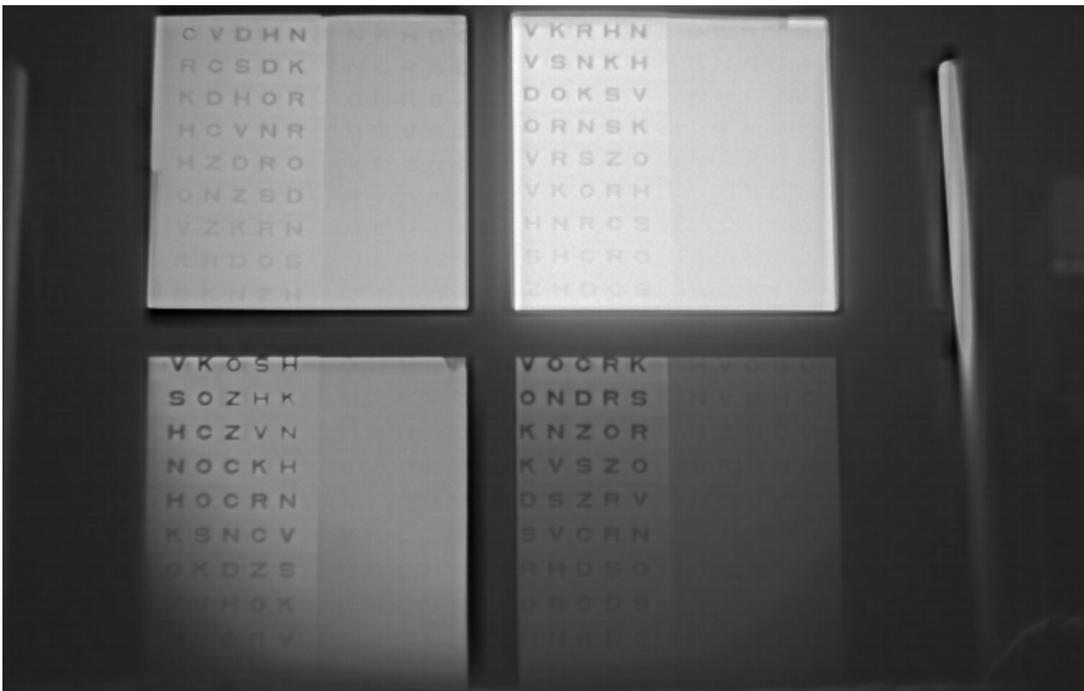


(b)

Figure 36: Sample Dilbert results from (a) benchmark chain and (b) wavelet chain



(a)



(b)

Figure 37: Sample Matrix results from (a) benchmark chain and (b) wavelet chain

Assessment

The subjective assessment of the image results was performed by fourteen participants. Four subjects had operational experience with digital vision systems and/or experience evaluating such systems for the end users. Five subjects were software developers with experience implementing and analyzing image processing algorithms in a wide range of application contexts. Five subjects were system engineers with experience designing and developing helmet-mounted vision systems for a range of operational contexts. The evaluation was conducted in several smaller groups, as it was not possible to coordinate all subjects in a single sitting.

The assessment was conducted by presenting the results to the subjects in a darkened room. Input images were presented to the subjects first, followed by pair-wise comparisons of images produced by the processing chains. Each pairwise comparison comprised a result from the benchmark framework and a result from the wavelet-based framework. All images were displayed on slides with a dark gray background. This study environment was established to provide viewing conditions that were representative of the likely operating environment in the target application. Each slide was presented for up to one minute to allow the subjects to select the preferred image result. Subjects were asked to indicate preference based on the amount of visual information that could be perceived and which image was more visually pleasing. They could also indicate that neither image result was acceptable. The reviewers were asked to provide descriptive feedback for their selection, such as commentary on detail preservation or enhancement and perceived artifacts.

The subjects reviewed eight data sets that were collected in six separate scenes, with Test Scene 1 captured in three different lighting conditions. The data sets included examples for both multi-focal and multi-spectral images. The subjective evaluation included three pairwise comparisons of each of the data sets. Each comparison presented the result of the benchmark processing chain with the results of the wavelet-based

processing chain produced using different DRC α and β values. The percentage of votes received by each framework during the subjective assessment and is presented in Table 3.

Table 3: Votes received by each framework during subjective assessment of results

Scene	Benchmark Chain (%)	Wavelet Chain (%)	Neither (%)
Diorama	45.2	54.8	0
Dilbert	61.9	38.1	0
Matrix	33.3	66.7	0
Test Scene 1 – Dark	2.4	7.1	90.5
Test Scene 1 – Mid	85.7	14.3	0
Test Scene 1 – Light	66.7	28.5	4.8
Track	16.7	11.9	71.4
Lab Scene	57.1	23.8	19.1
Total	46.1	30.7	23.2

There were a wide range of comments received during the evaluation. Select comments include:

- Test Scene 1 – Mid lighting conditions – “This makes the scene look like it was shot in daylight”
- Regarding gridding artifacts that occasionally appeared in images produced by the wavelet framework – “...are a bit distracting”
- Diorama scene from subjects that selected the wavelet framework – “Similar information content, but details are slightly sharper”

- Matrix scene from subjects that selected the wavelet framework – “Letters can be seen more clearly”

Neither framework satisfactorily enhanced the images in Test Scene 1 – Dark lighting conditions, with both producing images that were quite “washed out.” The Track scene was highly corrupted by noise, which neither framework could successfully reduce. The noise was then propagated through the subsequent processing, producing poor results. Surprisingly, there were several comments regarding a preference for “views presented by current NVGs” by several experts familiar with those systems.

Discussion

Upon first review, the results of the subjective assessment seem to be fairly inconclusive, not indicating a strong preference for either framework. However, if the data from Test Scene 1 – Dark lighting conditions and the Track scene (the two scenes where both frameworks performed quite poorly) are excluded, the results present a different interpretation. When these results are excluded, the benchmark framework received 58.3% of the votes, and the wavelet framework received 37.7% of the votes. This is a much more expected result, as the benchmark framework was designed to produce high-quality results and demonstrate the tradeoff that may be required between image quality and processing speed.

The image results clearly show that the DRC algorithm controls the resulting image quality and that the optimal breakpoints are highly scene dependent. This was further demonstrated during the subjective assessment, where the resulting image from the wavelet-based framework that received the most votes in each scene did not exhibit a common set of breakpoint values. α_1 values ranged from 128 to 512, α_2 from 1024 to 4096, β_1 from 8 to 32, β_2 from 32 to 64, and $\beta_3 = 256$ across the different scenes.

One alternative that could be used to produce more consistent results is to apply a substitute tone mapping algorithm outside of the wavelet-based processing chain. This

concept was briefly explored, applying the CLAHE algorithm to the image and using the wavelet framework to reduce the bit depth via a simple scaling of the coefficient values. The result, shown in Figure 38, does not contain any gridding artifacts and even rivals the results of the benchmark framework. This concept should potentially be explored further.



Figure 38: DWT-based framework results with CLAHE applied pre-processing and simple scaling of wavelet coefficients for DRC

It should also be considered that automatic DRC may not necessarily be the “holy grail.” It may, alternatively, make sense to provide the user with a predefined set of adjustments and the ability to select and switch as necessary.

Finally, the comments regarding a preference for view presented by current NVGs, while somewhat surprising, may expose a limitation in the approach used to conduct the review or a perception bias of the subjects. Due to the way they are designed and operate, current night vision devices present imagery that can be described as “green-scale.” The images produced by the frameworks were presented as typical grayscale

images. It is not unreasonable to assume that the familiarity with, and unconscious anticipation of evaluating, green imagery may have influenced the results and comments.

Computational Complexity and Power Consumption

Computational cost is frequently characterized through run-time analysis, a theoretical classification that estimates the computation time of an algorithm as a function of the input size and often presents the result using “Big O” notation. For example, if both the high-pass and low-pass filters are of constant length, the convolution-based filter bank implementation of the DWT grows at a linear rate with respect to the input size, which is denoted $O(N)$. Theoretical notation is important due to the significant impacts inefficient algorithms can have on system performance, especially in applications, such as the target application of this research, where processing and rendering of the results is time-critical. This notation also addresses a relevant limitation in empirical comparisons, where for small input sizes an empirical analysis shows that an algorithm may appear superior, but for sufficiently large input sizes it becomes clear that the growth rate surpasses that of an alternative algorithm. However, for the target application, information regarding the typical sensor resolutions is available, correlating to anticipated input sizes of the data sets. Consequently, meaningful empirical comparisons can also be made.

The computational complexity for both the wavelet-based and benchmark frameworks are briefly characterized in theoretical terms, using Big O notation. The theoretical analysis is essential for providing insights into anticipated run-time for applications beyond the scope of this effort that may use different sensors/imaging devices with varying resolutions. An empirical analysis was also performed, analyzing image sets captured in laboratory and operationally relevant environments. The purpose of this empirical analysis was to further demonstrate the tradeoffs that are often required in designing such digital imaging systems. It is important to note that the empirical

analysis was designed and executed in a manner that properly excluded any loading and GUI-specific processing tasks from the characterization of the computational cost.

While it is also well understood that other components, such as the sensor and display, contribute to the overall power consumption of a digital vision system, only the power consumption associated with processing the imagery is considered here. The computation time required to process the imagery is used to characterize the approximate power consumption for both the wavelet-based and benchmark frameworks. The computations for power consumption are based directly on instructional operations and take into account energy per instruction (in picoJoules/operation (pj/op)), clock speed (in Hertz (Hz)), and frame rate (in Hz). The power analysis first characterizes the energy usage for the platform employed during this research, a general-purpose computing platform, and then is extrapolated to a specific embedded processor that is used for the target application.

Computational Complexity Analysis

A theoretical analysis provides insights into the expected execution time for applications beyond the scope of this effort that may utilize different sensors/imaging devices with resolutions that from those used for the target application. This analysis provides an upper bound on how the processing chain execution time will grow as the size of the input images grow.

The theoretical analysis of the wavelet processing chain provides the following insights:

- The DWT can be implemented for image processing using the 1D DWT in a separable fashion. At each level of decomposition, or reconstruction for the inverse, the transform is applied first to the rows and then to the columns. Thus, for an $n \times m$ image, the computational complexity of the 2D DWT is $O(nm)$. The overall computational complexity contribution for performing the DWT on i input

images will be $O(inm)$. Since only a single inverse is performed at the conclusion of the processing chain, there will be an additional $O(nm)$.

- The soft thresholding operator is a point-processing operator with complexity $O(1)$. Thus, for a 1D signal of length l , the computational complexity is $O(l)$, and for an $n \times m$ image the computational complexity is $O(nm)$. The overall computational complexity contribution for i input images will be $O(inm)$.
- Using proper indexing, the fusion algorithm can be computed in four passes through each image. The quality values can be aggregated down through the resolution levels from highest to lowest in a single pass through the image. Normalizing the values requires a second pass through the image, and propagating the quality values back up through the resolution levels can be performed in a third pass. This is multiplied by the number of input images, i . A final pass is used to find the maximum quality value from the input images and select the coefficient from the image with the highest quality value. Thus, for i input images of resolution $n \times m$, the overall computational complexity is $O(c_1nm)$, where c_1 is a scalar value.
- The dynamic range compression algorithm can be computed in two passes through the image. Using proper indexing, the maximum value of the approximation coefficients and detail coefficients can be computed in one pass through the image. Again using proper indexing, the approximation coefficients and detail coefficients can be modified using the piecewise linear function and scalar value through a second pass through the image. Thus, for an $n \times m$ image the computational complexity is $O(c_2nm)$, where c_2 is a scalar value.

In Big O notation, constant factors and lower terms are dropped. Therefore, the overall computational complexity of the wavelet-based processing chain grows in direct proportion as the dimensions of the image grow and can be stated as $O(nm)$.

The theoretical analysis of the benchmark processing chain provides the following insights:

- The median filter algorithm is known to have high computational cost dominated by the calculation of the median, and more specifically, sorting the pixels in the filter window. The worst-case complexity for a sorting algorithm is $O(w^2)$, where w is the number of pixels in the window that must be sorted. The average case, using more efficient sorting algorithms, is $O(w \log w)$. Thus, for an $n \times m$ image the computational complexity is $O(w \log w \cdot nm)$, and this algorithm is applied to each of the i input images.
- The fusion algorithm is the same algorithm utilized in the processing chain. Therefore the contribution to the computational complexity includes the $O(inm)$ for the DWT, $O(nm)$ for the inverse DWT, and $O(c_1nm)$ for the fusion process described above.
- The Drago implementation scheme is computed in multiple passes through the image. First, the minimum and maximum values are identified. The next pass applies the logarithmic tone mapping and is followed by a third pass that applies the gamma correction. Thus, for an $n \times m$ image the overall computational complexity is $O(c_1nm)$, where c_1 is a scalar value.

The computational cost is dominated by the median filter algorithm. The overall computational complexity of the benchmark processing chain can be stated as $O(w \log w \cdot nm) + O(nm)$. Because the window size in the median filter is not constant, and the number of times each window size is applied varies with the size of the image, that term cannot be simply considered a scalar.

The empirical analysis was conducted by performing 500 executions of both the DWT-based framework and the benchmark framework for each of the scenes that were processed. The software was modified to enable repeated execution of the processing chains without having to reload the source images. The total processing time, as well as

the processing time for each subroutine, was collected after each execution of each framework. An example plot of the execution time data produced from analyzing a single data set with both frameworks is shown in Figure 39.

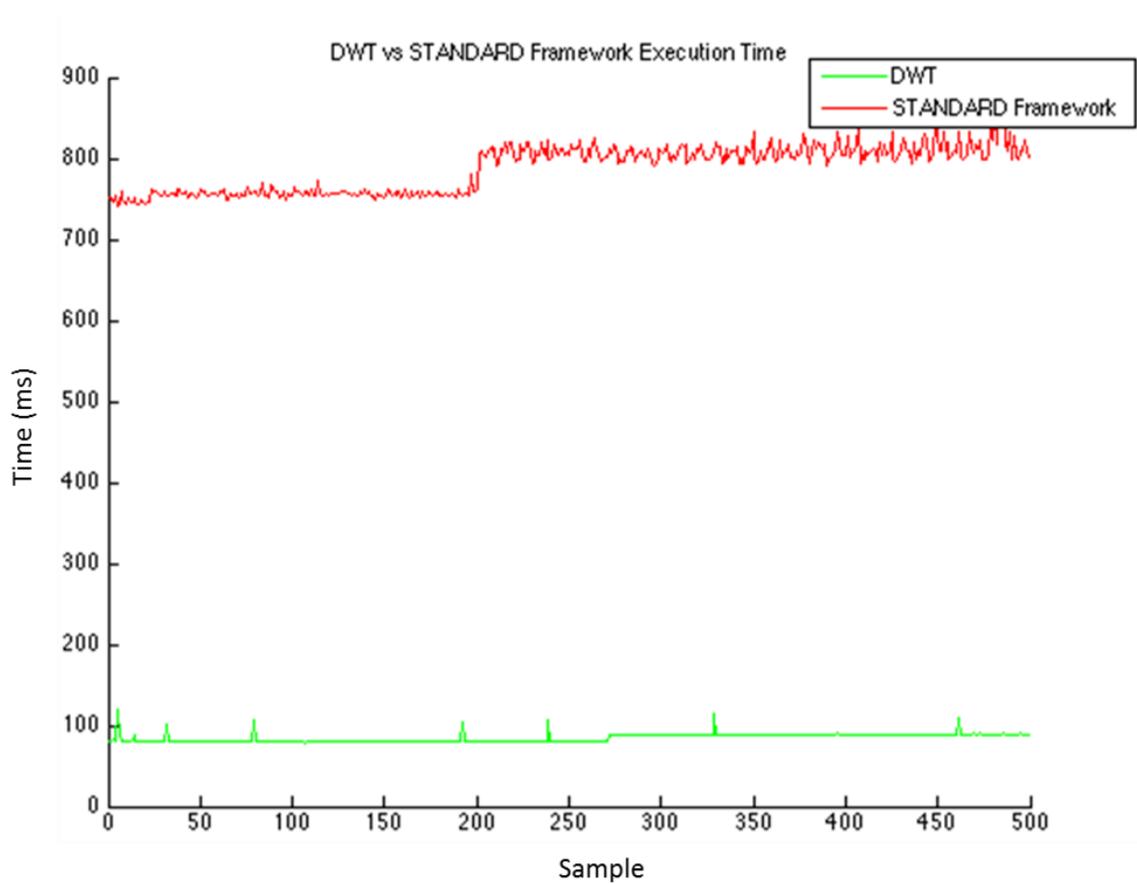


Figure 39: Execution time in milliseconds for DWT-based and benchmark frameworks

For the example plotted above, the average processing time was 83.86 milliseconds (ms) for the wavelet-based processing chain and 786.24 ms for the benchmark processing chain. Table 4 and Table 5 provide the breakdown of these processing times by subroutine, as well as an extrapolation of the processing time to execute the frameworks on the embedded processor.

Table 4: Average processing time of the DWT-based framework

Algorithm	General Purpose Processor Execution Time (ms)	Embedded Processor Execution Time (ms)
DWT	33.71	29.93
Soft Threshold	16.47	14.62
Fusion	19.41	17.23
DRC	1.98	1.76
I-DWT	12.29	10.91
Total	83.86	74.45

Table 5: Average processing time of the benchmark framework

Algorithm	General Purpose Processor Execution Time (ms)	Embedded Processor Execution Time (ms)
Median Filter	490.33	435.35
DWT	33.71	29.93
Fusion	19.41	17.23
I-DWT	12.29	10.91
Drago DRC	230.50	204.65
Total	786.24	698.07

The extrapolation function was generated from a detailed analysis of an implementation of the DWT using the LeGall 5/3 wavelet in both environments, which was applied to images of varying resolutions. The differences in processor capabilities, and the ability

to parallelize with the embedded processor, are factored into the calculation. The results demonstrate a 9.38 times improvement when comparing the two processing chains.

Power Consumption Analysis

As previously stated, it is well understood that hardware components, such as the sensors and display in the target application, contribute substantially to the overall power consumption of an image processing system. The results presented here are derived from the average execution time for the processing frameworks and factor in the instructional operations, energy per instruction (pj/op), clock speed (Hz), and frame rate (Hz). To calculate power consumption for the general-purpose computing platform used in this analysis, the execution time is first converted to instructional operations. Once the number of operations has been calculated, the following overall power consumption is computed, in picoWatts (pW), using the equation

$$power = \frac{frame\ rate \times energy\ per\ instruction \times ops}{clock\ speed} \quad Eq\ 40$$

To extrapolate the power consumption calculation to the embedded processor, the execution time for each algorithm on the general-purpose platform is mapped to an execution time for the embedded processor. As previously mentioned, this mapping was generated from a detailed analysis of an existing implementation of the DWT using the LeGall 5/3 wavelet in both environments, which was applied to images of varying resolutions. The differences in clock speed and ability to highly parallelize with the embedded processor are factored into the calculation. The newly calculated execution time is again converted to instructional operation and used in Equation 40 to compute the power consumption. The specific energy and clock speed information is not provided for the embedded processor at this time, as certain characteristics of the device are considered proprietary. The power consumption of the DWT framework and benchmark framework are provided in Table 6 and Table 7, respectively.

Table 6: Power consumption of the DWT-based framework

Algorithm	Execution Time (ms)	General-Purpose Power Consumption (pW)	Embedded Processor Power Consumption (pW)
DWT	33.71	2,537.44	7.30
Soft Threshold	16.47	1,239.74	3.57
Fusion	19.41	1,461.04	4.20
DRC	1.98	149.04	0.43
I-DWT	12.29	925.10	2.66
Total	83.86	6,312.36	18.16

Table 7: Power consumption of the benchmark framework

Algorithm	Execution Time (ms)	General-Purpose Power Consumption (pW)	Embedded Processor Power Consumption (pW)
Median Filter	490.33	36,908.48	106.14
DWT	33.71	2,537.44	7.30
Fusion	19.41	1,461.04	4.20
I-DWT	12.29	925.10	2.66
Drago DRC	230.50	17,350.36	49.90
Total	786.24	59,182.42	170.20

The tables include the calculated values for both the general-purpose computing platform and the embedded processor. It should be noted that characterizing only the

power consumption associated with processing the images means the results are dependent upon the image resolution. The data presented in the tables is the average execution time for imagery that is representative of the data collected by sensors utilized for a system intended for the target application.

As shown in the tables, the existing implementations of the processing frameworks produced approximate power consumption of 59.182.42 pW for the benchmark chain and 6,312.36 pW for the DWT-based chain on the general-purpose computing platform. The mapping to the embedded processor calculates 170.20 pW for the benchmark chain and 18.16 pW for the DWT-based chain. This demonstrates a 9.37 times reduction in power consumption when comparing the two processing methods. The power consumption calculations also demonstrate that the use of the embedded processor would result in an approximately 347.60 times reduction from the general-purpose processing platform for the DWT-based chain and 347.72 times reduction for the benchmark chain.

Discussion

Processing times that result in latencies perceptible to the human visual system would render the system unusable for the target application in helmet-mounted vision systems. It could also have critical effects in other applications, such as remote surgical procedures. Unfortunately, what constitutes “human perceptible” latency is not a simple thing to define, as it can be highly dependent on the application context and features of the display device, including resolution and FoV. This is noticeably demonstrated by the range of “acceptable” latencies documented in the following examples:

- 100 ms was established as an acceptable threshold for latency in human-computer interaction over four decades ago [189]

- The Federal Aviation Administration (FAA) requirements for Level D certification of flight simulators allow 120 ms latency in visual systems for fixed-wing platforms and 150 ms latency for rotary-wing platforms [190]
- A study on synthetic and enhanced vision systems for rotary-wing platforms deemed 150 ms HMD latency to be unacceptable, 100 ms as marginal, and 50 ms as preferred [191]
- A NASA study on latency in visionic systems concluded that acceptable latencies should be as low as 20 ms [192]
- An evaluation of performance in gaming environments showed that latencies of approximately 40 ms can have negative impacts on tasks such as tracking and shooting accuracy [193]

This range of thresholds makes characterizing the acceptability of the empirical computational complexity analysis quite challenging.

The average execution time of the benchmark framework clearly exceeds all documented thresholds of latencies that are perceivable by the human visual system. The average execution time of the DWT-based framework is not nearly as straightforward. The current results are below the initially established 100 ms threshold but do not meet some of the more stringent metrics identified in more recent literature. It is important to recall, however, that the results of the empirical analysis were obtained from the algorithms executing on a general-purpose CPU, as opposed to a specialized embedded processor or GPU. When considering techniques that can further reduce execution time, such as parallel processing, hardware acceleration, and the ability to process the image as lines are being acquired by the sensors, it is not a significant conceptual leap to conclude that a customized processor and implementation should result in non-perceptible latencies. This is especially true when taking into account that Moore's Law continues to hold true, and the capabilities of processors continue to increase at a rapid pace.

The results of the power consumption analysis are only relevant in limited application contexts. For many applications, such as those in medical fields or when the processing sub-system is able to connect to (or is part of) a platform like an aircraft, there are few limitations that need to be placed on the power consumption. However, for the target application, the system is “untethered,” and the batteries required to power the system are carried by the user. The power consumption demonstrated by the DWT-based framework, when deployed on the target hardware, equates to the system running approximately four to six hours on four standard batteries, including the power required for the sensors and displays.

The results obtained from testing this hypothesis are especially noteworthy considering that the subjective analysis of the resulting images indicated a preference for images produced by the benchmark framework, and that the feedback from the SMEs with operational experience indicated an overall preference for the views provided by their existing analog NVGs. The results reinforce that there may be a significant tradeoff between producing high-quality images and requirements for low-latency/low-power processing to consider when designing and implementing processing frameworks such as the DWT-based chain presented in this work. When considering the cost of existing equipment (such as current NVGs), or that such an image processing sub-system for a given application may not exist, even the less-preferred results may justify the tradeoff.

Algorithm Selection and Ordering

A processing chain of this nature is useful for many applications; however, it would be difficult to design a “one size fits all” solution. Certain applications may require a large number of algorithms to achieve the desired results, while others may only need to utilize a smaller subset. Some applications may produce better results if the algorithms are performed in a different order than that proposed for the intended application.

Therefore, the final goal of this research is to provide an end user the freedom to modify the wavelet processing chain based on the needs of the target application. The critically sampled wavelet utilized in the processing chain is highly susceptible to aliasing, and the modifications each algorithm makes to the coefficients can have significant impacts on the resulting imagery. Thus, tailoring the framework for broader applicability requires a systematic assessment of algorithm selection and ordering in the processing chain. The result of this assessment is a set of guidelines that should be considered when selecting and ordering algorithms in the processing chain for a given application. The guidelines present the tradeoffs in terms of image quality and, when applicable, computational cost and power consumption. It is important to note that not all applications have the same strict requirements for real-time processing and minimizing power consumption.

Fuse Early – The output in terms of computation cost and power consumption will largely be driven by the algorithms that are applied prior to fusion in the processing chain. The intended application and many other potential applications operate on multiple sources of imagery that are combined into a single output, or a reduced number of outputs, at some point in the process. Any algorithm applied prior to fusion in the processing chain will likely be repeated on every input image, and the computational cost and power consumption associated with applying that algorithm to a single image is increased by a factor of the number of input images.

Reduce Noise Early – The image data captured in most, if not all, application contexts for such processing chains will be corrupted by noise. Therefore, noise reduction algorithms are unlikely to ever be tailored out of a processing chain. In the current design, noise reduction is applied individually to each input source prior to performing image fusion. Some noise reduction algorithms can be computationally expensive (e.g., median filtering), and execution time could be significantly reduced by performing noise reduction after the input images have been fused. While this approach

would follow the initial guideline of fusing early, many fusion algorithms can propagate, and in some cases even enhance, the noise in an image. The fusion approach utilized in the processing chains presented in this work has a decreased probability of propagating noise, as the quality measure computation would likely assign a smaller quality value to a coefficient associated with noise than a coefficient associated with a feature. However, this guideline should be observed in processing chains that use alternative fusion methods.

Image Scale Carefully – Algorithms that inherently contain steps that perform image scaling, or are specifically designed for image scaling, should be considered carefully. Image scaling is a key step in many algorithms such as super-resolution and panoramic stitching. It is also a necessary step in applications, such as the target application, where the input images are not inherently the same size and need to be combined (or fused) in some manner. Higher resolution may be desirable to support preservation of detail in subsequent processing steps or to assist operational tasks such as detection, recognition, and identification and there is a significant impact on computational complexity and power consumption. For example, increased resolution means a greater number of pixels associated with the image features, which may lead to better fusion results. However, applying image scaling to each image prior to fusion will come at a significant computational expense. Both the theoretical and empirical analyses discuss the dependency of the execution time of the frameworks on the size of the input image (image resolution), hence the need to carefully consider how and where these types of algorithms are incorporated into the processing chain.

CHAPTER 6 SUMMARY AND FUTURE WORK

Summary

This dissertation presented the design and preliminary implementation of an image processing sub-system that minimizes computational complexity and power consumption by eliminating repeated transformations between processing domains. Specifically, this processing chain utilized the LeGall 5/3 wavelet as the basis for applying multiple algorithms within a single domain. The key factor in designing the processing chain was to establish an understanding of how the wavelet coefficients control characteristics of an image, such as dynamic range, and the sensitivity of the wavelet reconstruction to aliasing. This understanding can be used to extend the processing chain to include additional algorithms.

The wavelet processing chain was compared to a benchmark processing chain comprised of algorithms intended to produce high-quality image results. This comparison characterized the results in terms of image quality, computational costs, and power consumption. Subjective quality assessment concluded that additional work is required to improve the imagery produced by the wavelet-based framework. Analysis of computational cost and power consumption support the hypothesis that the wavelet framework can enable low-latency, low-power applications, demonstrating a nearly a 9.4 times reduction in execution time and power consumption. However, data sets obtained from additional sensor systems of varying resolutions should be processed before stronger claims of validation are made.

Finally, a set of guidelines was provided to support tailoring the processing framework for broader applicability. These guidelines should be considered when selecting and ordering algorithms in the processing chain. They leverage the understanding established during the design phase of how algorithms influence the image

characteristics, and they seek to provide guidance on impacts to image quality, execution time, and power consumption.

Future Work

There is much additional research that can be performed to further extend the principles demonstrated by this work. Extensions to the processing framework can be made through further refinement of the algorithms, integration of additional algorithms, and an analysis of alternative wavelets. Additionally, exploration of hardware solutions to accelerate the algorithms and IQA techniques for these types of processing chains would be significant contributions to the field. Further application-specific research includes the insertion of symbology into the final image and an assessment of artifacts that may be introduced by platform characteristics.

The results clearly show that the breakpoints of the piecewise linear function applied to the approximation coefficients during DRC have a significant impact on the final image quality. Appropriate sets of breakpoints can be identified to produce quality image results, but it is unlikely that values for one environmental context will produce quality image results in all contexts. The optimal set of breakpoints is scene dependent. This property was empirically observed when applying the framework across sets of source imagery, where breakpoints had to be modified when processing different sets to maintain image quality. Thus, a final implementation of such an image processing subsystem could include a dynamic calculation and optimization of the parameters for the piecewise linear function. However, given the complexity of this optimization problem, it was not addressed in this work. As previously stated in the discussion of the image results, an additional alternative is to use manually selectable settings and not perform automatic DRC.

The critically sampled real-valued wavelet analysis filter bank can be implemented incredibly efficiently but produces highly aliased coefficients. The

synthesis filter bank is designed for perfect reconstruction; however, imperfect cancellation occurs when the coefficients are processed in any way, manifesting as ringing or other artifacts around image edges. This holds true for the LeGall 5/3 wavelet used in this implementation and is further exacerbated by the short length of the analysis and synthesis filter banks. This wavelet is also significantly shift variant, and as the scene under view shifts, a dramatically different set of coefficients is computed. These limitations can make incorporating additional algorithms into the processing chain difficult, as further manipulation of the wavelet coefficients can easily introduce artifacts into the output images.

An alternative approach that may address these effects would be developing the framework utilizing an alternative wavelet, either real-valued or complex. For example, the DT-CWT supports near state-of-the-art algorithms for most of the candidate processing tasks, and a brief investigation into utilizing it for the framework is summarized in Appendix B. The DT-CWT framework generated visually appealing image results; however, the current approaches for implementing the DT-CWT drive computational complexity to levels that significantly exceed those of even the benchmark processing chain. In order to make it a reality on practical fixed-point image processing hardware, a transform with wavelet and scaling filters that take only integer values would need to be designed. Additionally, the redundant nature of the transforms will also drive increased memory requirements that would be difficult to support for the intended application to helmet-mounted vision systems.

The focus of this work was on developing a low-latency, low-power software solution. As discussed in the results section, the computational cost and power consumption could be further reduced by coupling the software with a custom hardware solution, such as a specialized processor. This concept is becoming more popular, which can be clearly seen in the emerging class of microprocessors, called Vision Processing Units (VPUs), which are specially designed to accelerate machine vision algorithms.

Examples include the Movidius Myriad 2 family of VPUs designed for machine vision, machine learning, and augmented/virtual reality applications [194] and the NXP Semiconductors S32V vision processor for automotive vision systems [195].

Developing IQA methods has been an active research topic for several decades, but further research is required. The limitations in objective quality assessment techniques were discussed previously as part of the research design. There is no single existing IQA technique that can properly assess the results of processing chains such as those presented in this work. The design of such a metric would be a significant contribution to the field of image processing, especially as applications of image processing systems continue to expand.

Most fielded helmet-mounted vision systems include 2D or 3D conformal line drawn symbology that provides enhanced situational awareness. Injecting the symbology into the final image results could be treated as a fusion problem; however, this approach would likely degrade the line quality of the symbology. Identifying an appropriate method and position to insert symbology into the processing chain would be a valuable application-specific extension of this work. The most likely approach to provide symbology, however, would be through a merge plane as a simple overlay that could be switched on and off by the user.

In certain operational platforms, specific characteristics of the platform itself may introduce artifacts into the imagery. For example, multi-function displays in many airborne platforms are line scanned displays that, when imaged by a digital sensor, are likely to create unacceptable artifacts in the resulting imagery. Identifying these platform-specific issues and developing mitigation strategies within the processing chain itself will be critical to fielding digital helmet-mounted vision systems.

REFERENCES

- [1] Melzer, J.E., and Moffitt, K. (1997). "Head-Mounted Displays: Designing for the User." New York: McGraw-Hill.
- [2] Rash, C.E. (editor) (2000). "Helmet-mounted displays: Design issues for the rotary-wing environment." Bellingham, WA: SPIE Press.
- [3] Manning, S.D., and Rash, C.E. (2007). "A moveable view." *AeroSafetyWorld*. Flight Safety Foundation. 2, 41-44.
- [4] Shontz, W.D., and Trumm, G.A. (1969). "Perceptual processes and current helmet-mounted display concepts." Minneapolis, Minnesota: Honeywell Inc. Technical Note TN-1.
- [5] Walker, D.J., Verona, R.W. and Brindle, J.H. (1980) "A newly developed helmet-mounted display system for attack helicopters." *Proceedings of the Society for Information Displays*. Society for Information Displays. San Diego, CA.
- [6] Rash, C.E., and Martin, J.S. (1988). "The Impact of the U.S. Army's AH-64 Helmet Mounted Display on Future Aviation Helmet Design." Fort Rucker, Alabama: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 88-13.
- [7] Thales. (2010). TopOwl[®]. Retrieved April 13, 2014 from <https://www.thalesgroup.com/sites/default/files/asset/document/TopOwl%20Brochure%20Nov%202010.pdf>.
- [8] Bayer, M.M., Rash, C.E., and Brindle, J.H. (2009). "Introduction to helmet mounted displays" in Rash, C.E., Russo, M.B., Letowski, T.R., and Schmeisser, E.T. (eds.), *Helmet-mounted displays: Sensation, perception, and cognitive issues*. U.S. Army Aeromedical Research Laboratory, 47-108.
- [9] Harding, T.H., Beasley, H.H., Martin, J.S., Rash, C.E., McLean, W.E., Mora, J.C., and McEntire, B.J. (1998). "Optical and biodynamic evaluation of the Helmet Integrated Display Sight System (HIDSS) for the RAH-66 Comanche development and validation program phase." Fort Rucker, Alabama: U. S. Army Aeromedical Research Laboratory. USAARL Report No. 98-22.
- [10] Efke, A. and Jenkins, D. (1980). "Development of an Aviator's Night Vision Imaging System (ANVIS)." Presented at *SPIE Int. Tech. Symposium Exhibit*. San Diego, CA.
- [11] Wood, R.B. (1992). "Holographic head-up displays." *Electro-optical displays*, In: Karim, M. (ed.). New York: Marcel Dekker. 337-415.
- [12] Cameron, A.A., and Steward, D.G. (1994). "The Viper HMD – from design concept to flight test." *Helmet- and Head-Mounted Displays and Symbolology Design Requirements, Proceedings of SPIE*, 2218, 137-148.
- [13] Vos, G.D., and Brandt, G. (1990). "Use of holographic optical elements in HMDs." *Helmet Mounted Displays II, Proceedings of SPIE*, 1290, 70-80.

- [14] BAE Systems. (2007). "The Q-Sight™ family of helmet display products." Retrieved from www.baesystems.com/BAEProd/groups/public/documents/bae_publication/bae_pdf_eis_q-sight.pdf
- [15] SBG Lab. (2014). "Head Mounted Display DigiLens® Enables "Heads-Up" Holographic Imagery." Retrieved April 12th 2014 from http://www.digilens.com/Head_Mounted_Display.html.
- [16] Johnston, R.S., and Willey, S.R. (1995). "Development of a commercial retinal scanning display." *Helmet- and Head-Mounted Displays and Symbology Design Requirements II, Proceedings of SPIE*, 2465, 2-13.
- [17] Marshall, G.F. (1989). "Back from the past, helmet-mounted display." *Proceedings of SPIE*, 1116, 2-11.
- [18] Braybrook, R. (1998). "Looks can kill." *Armed Forces Journal International*, 4, 44.
- [19] Klass, P.J. (1972). "Navy pilots to use helmet sight." *Aviation Week and Space Technology*, January 31, 37-40.
- [20] Rash, C.E. (2008). "A 25-year retrospective review of visual complaints and illusions associated with a monocular helmet-mounted display." *Health and Safety Aspects of Visual Displays, Displays*, 29 (2), 70-80.
- [21] Vision Systems International. (n.d.). "Joint Helmet Mounted Cueing System." Retrieved November 10, 2013 from <http://vsi-hmcs.com/jhmcs.htm>.
- [22] Beal, C., and Sweetman, B. (1994). "Helmet-mounted displays: Are we jumping the gun?" *International Defense Review*, September, 69-75.
- [23] Merryman, R.K. (1994). "Vista Sabre II: Integration of helmet-mounted tracker/display and high boresight missile seeker into F-15 aircraft." *Proceedings of SPIE*, 2218, 173-184.
- [24] Lucas, T. (1994). "Advances in helmet-mounted displays." *Avionics*, June, 20-27.
- [25] Elbit Systems. (2004). "MiDASH Modular Integrated Display and Sight System for helicopter pilots."
- [26] Vision Systems International. (n.d.). "F-35 Gen II HMDS." Retrieved November 10, 2013 from <http://vsi-hmcs.com/f35.htm>.
- [27] Brickner, M.S. (1989). "Helicopter Flights with Night Vision Goggle – Human Factors Aspects." NASA Technical Memorandum 101039.
- [28] McLean, W., Rash, C., McEntire, J., Braithwaite, M. and Mora, J. (1997). "A Performance History of AN/PVS-5 and ANVIS Image Intensification Systems in US Army Aviation." *Proceedings of SPIE*, 3058, 265-298.
- [29] – (1946). "Bull's-Eyes in the Night". *Popular Science*. July, 73.
- [30] – (1958). *Image Intensifier Symposium (Proceedings)*, U.S. Army Engineer Research and Development Laboratories, October.
- [31] Lampton, M. (1981). "The Microchannel Image Intensifier." *Scientific American*. 245(2), 62-71.

- [32] Aeronautical Design Standard. (1981). "Aircrew Station Lighting for Compatibility with Night Vision Goggle Use." ASD-23, U.S. Army Aviation Research and Development Command.
- [33] Military Specification. (1986). "Lighting, Aircraft, Interior, Aviator's Night Vision Imaging System (ANVIS) Compatible." MIL-L-85762.
- [34] Military Specification. (1988). "Lighting, Aircraft, Interior, Night Vision Imaging System (NVIS) Compatible." MIL-L-85762A.
- [35] Reetz III, F. (1987). "Rationale Behind the Requirements Contained in Military Specifications MIL-L-85762 and MIL-L-85762A." Report NADC-87060-20, Naval Air Development Center.
- [36] House of Representatives. (1989) "Night Vision Goggles." Hearing before the Investigations Subcommittee of the Committee on Armed Services, held March 21. U.S. Government Printing Office, Washington, D.C.
- [37] Office of the Director. (1989). "Review of Testing Performed on AN/PVS-5 and AN/AVS-6 Aviation Night Vision Goggles." Operation Test and Evaluation.
- [38] – (1989). "Review of the use of NVG in Flight Training." Report for the Deputy Secretary of Defense.
- [39] Green, D.L. (1994a). "Rotorcraft Night Vision Goggle Evaluation." Report DOT/FAA/RD-19/11.
- [40] Green, D.L. (1994b). "Civil Use of Night Vision Devices – Evaluation Pilot's Guide Part I." Report FAA/RD-94/18.
- [41] Green, D.L. (1994c). "Civil Use of Night Vision Devices – Evaluation Pilot's Guide Part II." Report FAA/RD-94/19.
- [42] Green, D.L. (1994d). "Assessment of Night Vision Goggle Workload – Flight Test Engineer's Guide." Report FAA/RD-94/20.
- [43] – (2008). "War Creates Shortage in Night Vision Goggles." *Associated Press*. Retrieved November 10, 2013 from <http://www.msnbc.msn.com/id/24971226/>.
- [44] Xu, F., Liu, X., and Fujimura, K. (2005). "Pedestrian Detection and Tracking with Night Vision." *Intelligent Transportation Systems, IEEE Transactions*, 6, 63-71.
- [45] Digregorio, B. (2006). "Safer Driving in the Dead of Night [Infrared Vision Systems]." *Spectrum, IEEE*, 43, 20-21.
- [46] Hoffman, M. (2014). "Scientists Develop Night Vision Contact Lens." Retrieved April 26, 2014 from <http://defensetech.org/2014/03/28/scientists-develop-night-vision-contact-lens/>
- [47] Smith, G., and Atchison, D.A. (1997). *The eye and visual optical instruments*. New York: Cambridge University Press.
- [48] Sutherland, I.V. (1965). "The Ultimate Display." *Proceeding of the IFIP Congress*, 2, 506-508.
- [49] Keller, K., and Colucci, D. (1998). "Perception in HMDs: What is it in head-mounted displays (HMDs) that really make them all so terrible?" *Proceedings of SPIE*, 3362, 46-53.

- [50] Hopper, D.G. (2000). "Reality and Surreality of 3-D Displays: Holodeck and beyond." Invited Keynote Paper in the *Electronic Information Display Conference*, Society for Information Display ExCel, England U.K., November 21-23, 2000.
- [51] Haar, R. (2005). "Virtual reality in the military: Present and future." Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, the Netherlands.
- [52] Jennings, S. (1997). "An in-flight evaluation of stereoscopic imagery in helicopter cockpits." *SPIE Aerosense, Proceedings of SPIE*, 3057, 108/117.
- [53] Arthur, J.J., Bailey, R.E., Williams, S.R., Prinzel, L.J., Shelton, K.J., Jones, D.R., and Houston, V. (2015). "A review of head-worn display research at NASA Langley Research Center." *Proceedings of SPIE*, 9470.
- [54] Sweetman, B. (2015). "Thales Unveils Head-Worn Display for Business Jets." Retrieved March 12, 2016 from <http://aviationweek.com/business-aviation/thales-unveils-head-worn-display-business-jets>
- [55] Duchowski, A. (2000). "Binocular Eye Tracking in virtual reality for inspection training." *Proceedings of the Eye Tracking Research and Applications Symposium*, Palm Beach Gardens, FL, November 6-8, 89-96.
- [56] Bajura, M. (1992). "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient." *Computer Graphics, Proceedings of SIGGRAPH*, 26(2), 203-210.
- [57] Ackerman, J. (2002). "Application of augmented reality to laparoscopic surgery." Doctorate dissertation submitted to the faculty of The University of North Carolina at Chapel Hill.
- [58] Chung, J. (1989). "Exploring virtual worlds with HMDs." TR89-009.
- [59] Rolland, J. (2002). "Head-mounted projective displays for creating distributed collaborative environments." *Proceedings of SPIE*, 4711, 395-399.
- [60] Davis, L. (2002). "Application of Augmented Reality to Visualizing Anatomical Airways." *Proceedings of SPIE*, 4711, 400-405.
- [61] Reinhart, W. (1991). "Depth cueing for visual search and positioning." *Proceedings of SPIE*, 1457, 221-232.
- [62] Primeau, G. (2000). "Wide field-of-view SVGA sequential color HMD for use in anthropometric telepresence application." *Proceedings of SPIE*, 4711, 11-20.
- [63] Hanlon, M. (2006). "Trimersion HMD takes first person shooter games to a new level of reality." Retrieved December 20, 2013 from <http://www.gizmag.com/go/6410/>.
- [64] Oculus VR®. (2014) "Oculus Rift: Next-Gen Virtual Reality." Retrieved March 15, 2014 from <http://www.oculusvr.com/rift/>.
- [65] Google. (2012). "Glass: What it Does" Retrieved March 15, 2014 from <http://www.google.com/glass/start/what-it-does/>.
- [66] Goldman, D. (2012). "Google unveils 'Project Glass' virtual-reality glasses." *Money* (CNN) Retrieved March 15, 2014 from http://money.cnn.com/2012/04/04/technology/google-project-glass/?source=cnn_bin.

- [67] Haar, A. (1910). "Zur Theorie der orthogonalen Funktionensysteme." *Mathematische Annalen*, 69, 331-371.
- [68] Mallat, S. (1989). "A theory for multiresolution signal decomposition: The wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 (7), 674-693.
- [69] Daubechies, I. (1988). "Orthonormal bases of compactly supported wavelets." *Communications on Pure and Applied Mathematics*, 41 (7), 909-996.
- [70] Acharya, T., and Ray, A.K. (2005). "Image Processing: Principles and Applications." Hoboken, NJ: John Wiley & Sons, 79-104.
- [71] Daubechies, I. (1992). "Ten Lectures on Wavelets." New York: *Society for Industrial and Applied Mathematics*.
- [72] Selesnick, I.W., Baraniuk, R.G., and Kingsbury, N. (2005). "The dual-tree complex wavelet transform - A coherent framework for multiscale signal and image processing." *IEEE Signal Processing Magazine*, 22 (6), 123-151.
- [73] Kingsbury, N. (1999). "Image processing with complex wavelets." *Philosophical Transactions of the Royal Society of London*, 357, 2543-2560.
- [74] Simoncelli, E.P., Freeman, W.T., Adelson, E.H., and Heeger, D.J. (1992). "Shiftable Multi-Scale Transforms." *IEEE Transactions on Information Theory*, 38 (2), 587-607.
- [75] Zitova, B., and Flusser, J. (2003). "Image registration methods: A survey." *Image and Vision Computing*, 21 (11), 977-1000.
- [76] Brown, L.G. (1992). "A survey of image registration techniques." *ACM Computing Surveys*, 24, 326-376.
- [77] Maintz, J.A., and Viergever, M.A. (1998). "A survey of medical image registration." *Medical Image Analysis*, 2 (1), 1-36.
- [78] Szeliski, R. (2006). "Image alignment and stitching: A tutorial." *Foundations and Trends in Computer Graphics and Vision*, 2 (1), 1-104.
- [79] Deshmukh, M., and Bhosle, U. (2011). "A survey of image registration" *International Journal of Image Processing (IJIP)*, 5 (3), 245-269.
- [80] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). "Deformable Medical Image Registration: A Survey." *IEEE Transactions on Medical Imaging*, 32 (7), 1153-1190.
- [81] Mani, V.R.S, and Arivazhagan, S. (2013). "Survey of Medical Image Registration." *Journal of Biomedical Engineering and Technology*, 1 (2), 8-25.
- [82] Prabala, A. (2000). "CCD vs. CMOS Imagers for Sci/Industrial Cameras." *Advanced Imaging*, January 2000, pp 16.
- [83] Jain, A., Kerhalkar, S.G., and Ahmed, M. (2014). "Review on Denoising for the AWGN signal introduced in a stationary image." *International Journal of Engineering Science Invention*, 3 (4), 1-10.
- [84] Saxena, C., and Kourav, D. (2014). "Noises and Image Denoising Techniques: A Brief Survey." *International Journal of Emerging Technology and Advanced Engineering*, 4 (3), 878-885.

- [85] Tiwari, S., and Naiyar, N. (2014). “De-Noising Techniques: A Comparative Approach.” *International Journal of Scientific Engineering and Research (IJSER)*, 2 (1), 44-47.
- [86] Ram, B.P., and Choudhary, S. (2014). “Survey Paper on Different Approaches for Noise Level Estimation and Denoising of an Image.” *International Journal of Science and Research (IJSR)*, 3 (4), 618-622.
- [87] Prasad, G., Mishra, A.K., and Singh, A.K. (2014). “In-depth Analysis of Wavelet Transform based Denoising Scheme for Smooth and Textured Images Corrupted with Gaussian Noise.” *International Journal of Applied Information Systems (IJ AIS)*, 6 (7), 27-32.
- [88] Yasar, H., Ceylan, M., and Ozturk, A.E. (2014). “Comparison of Real and Complex-Valued Versions of Wavelet Transform, Curvelet Transform and Ridgelet Transform for Medical Image Denoising.” *International Journal of Electronics; Mechanical and Mechatronics Engineering*, 3 (1), 427-436.
- [89] Kaveri, A.P., and Amrutkar, K.J. (2014). “Median filtering Frameworks and their Application to Image Enhancement.” *International Journal of Application or Innovation in Engineering & Management (IJA IEM)*, 3 (3), 509-512.
- [90] Jain, A., Mishra, S.K., and Richariya, V. (2014). “A Review: Impulse Noise Reduction.” *International Journal of Engineering & Management Innovations (IJEMI)*, 2 (1).
- [91] Buades, A., Coll, B., and Morel, J.M. (2005). “A Review of Image Denoising Algorithms, With a New One.” *Society for Industrial and Applied Mathematics Journal on Multiscale Modeling and Simulation*, 4 (2), 490-530.
- [92] Perona, P., and Malik, J. (1990). “Scale space and edge detection using anisotropic diffusion.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 629-639.
- [93] Yu, Y., and Acton, S. (2002). “Speckle reducing anisotropic diffusion.” *IEEE Transactions on Image Processing*, 11, 1260-1270.
- [94] Donoho, D.L. (1995). “De-noising by soft thresholding.” *IEEE Transactions on Information Theory*, 41 (3), 613-627.
- [95] Donoho, D.L., and Johnstone, I.M. (1995). “Adapting to unknown smoothness via wavelet shrinkage.” *Journal of the Statistical Association*, 90 (432), 1200-1224.
- [96] Chang, S.G., Yu, B., and Vetterli, M. “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing*, 9, 1532-1546.
- [97] Dan, L., Yan, W., and Ting, F. (2011). “Wavelet Image Denoising Algorithm Based on Local Adaptive Wiener Filtering.” *2011 International Conference on Mechatronic Science, Electric Engineering and Computer*, 2305-2307.
- [98] Sheng, M., Pang, Y., Wan, L., and Huang, H. (2014). “Underwater Images Enhancement Using Multi-Wavelet Transform and Median Filter.” *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12 (3), 2306-2313.

- [99] Jiang, D., Zhuang, D., Huang, Y., and Fu, J. "Survey of Multispectral Image Fusion Techniques in Remote Sensing Applications." Yufeng Zheng, Alcorn State University, USA (Ed), ISBN: 978-953-307-182-4, InTech, Available from: <http://www.intechopen.com/articles/show/title/survey-of-multispectral-image-fusion-techniques-in-remote-sensing-applications>.
- [100] Martinsen, G.L., Hosket, J.S., and Pinkus, A.R. (2008). "Correlating Military Operators Visual Demands With Multi-spectral Image Fusion." In *Proceedings of SPIE*, 6968, 69681S-1-69681S-7.
- [101] James, A.P, and Dasarathy, B.V. (2014). "Medical Image Fusion: A survey of the state of the art." *Information Fusion*, 19, 4-19.
- [102] Zhou, Y., Mayyas, A., and Omar, M.A. (2011). "Principal Component Analysis Based Image Fusion Routine With Application To Stamping Split Detection." *Research in Nondestructive Evaluation*, 78-79.
- [103] Yang, Y., Park, D.S., Huang, S., and Rao, N. (2010). "Medical Image Fusion via an Effective Wavelet-Based Approach." *EURASIP Journal on Advances in Signal Processing*, 2010, 1-12.
- [104] Wan, T., Zhu, C., and Qin, Z. (2013). "Multifocus image fusion based on robust principal component analysis." *Pattern Recognition Letters*, 34 (9), 1001-1008.
- [105] Burt, P.J., and Adelson, E.H. (1983). "The Laplacian Pyramid as a Compact Image Code." *IEEE Transactions on Communications*, 31, 532-540.
- [106] Anderson, H. (1987). "A Filter-subtract-decimate Hierarchical Pyramid Signal Analyzing and Synthesizing Technique." U.S. Patent 718 104.
- [107] Toet, A. (1996). "Image Fusion by a Ratio of Low-pass Pyramid." *Pattern Recognition Letters*, 9, 245-253.
- [108] Toet, A., van Ruyven, J.J., and Valetton, J.M. (1989). "Merging Thermal and Visual Images by a Contrast Pyramid." *Optical Engineering*, 28 (7), 789-792.
- [109] Lan, Y., Ren, H., Zhang, Y., and Hung, C. (2013). "Multi-band Vector Wavelet Transformation based Multi-Focus Image Fusion Algorithm." *Journal of Software*, 8 (1), 208-217.
- [110] Kannan, K., Perumal, S.A., and Arulmozhi, K. (2010). "The Review of Feature Level Fusion of Multi-Focused Images Using Wavelets." *Recent Patents on Signal Processing*, 2, 28-38.
- [111] Zhang, Q., Ma, Z., and Wang, L. (2013). "Multimodality image fusion by using both phase and magnitude information." *Pattern Recognition Letters*, 34 (2), 185-193.
- [112] Surthy, S., Parameshwaran, L., and Sasi, A.P. (2013). "Image Fusion Technique using DT-CWT." *2013 International Multi-Conference on Automation, Computing, Communication, Control, and Compressed Sensing*, 160-164.
- [113] Schnelle, S.R., and Chan, A.L. (2011). "Fusing Infrared and Visible Imageries for Improved Tracking of Moving Targets." Adelphi, Maryland: U.S. Army Research Laboratory. ARL-TR-5552.

- [114] Das, S., and Kundu, M.K. (2011). "Ripplet Based Multimodality Medical Image Fusion Using Pulse-Coupled Neural Network and Modified Spatial Frequency." *Proceedings of the International Conference on Recent Trends in Information Systems (ReTIS-11)*, 229-234.
- [115] Alparone, L., Baronti, S., Garzelli, A., and Nencini, F. (2005). "The curvelet transforms for fusion of very-high resolution multi-spectral and panchromatic images." *Proceedings of the 25th EARSeL Symposium*, Porto, Portugal, Millpress Science Publishers, The Netherlands.
- [116] Song, M., Chen, X., and Guo, P. (2009). "A fusion method for multispectral and panchromatic images based on HIS and Contourlet transformation." *10th Workshop on Image Analysis for Multimedia Interactive Services*, 77-80.
- [117] Al-Helali, A.H.M., Ali, H.A., Al-Dulaimi, B., Alzubaydi, D., and Mahmmoud, W.A. (2009). "Slantlet Transform for Multispectral Image Fusion." *Journal of Computer Science*, 5 (4), 263-269.
- [118] Xu, N., Tan, K., Arora, H., and Ahuja, N. (2004). "Generating omnifocus images using graph cuts and a new focus measure." *Proceedings of the International Conference on Pattern Recognition*, 697-700.
- [119] Agarwala, A., et. al. (2004). "Interactive digital photomontage." *ACM Transactions on Graphics*, 23 (3), 294-302.
- [120] Duan, J., Meng, G., Xiang, S., and Pan, C. (2013). "Multifocus Image Fusion via Region Reconstruction." *2013 Second IAPR Asian Conference on Pattern Recognition*, 396-400.
- [121] Zhao, H., Shang, Z. Tang, Y.Y., and Fang, B. (2013). "Multi-focus image fusion based on the neighbor distance." *Pattern Recognition*, 46 (3), 1002-1011.
- [122] Liang, J., He, Y., Lin, D., and Zeng, X. (2012). "Image fusion using higher order singular value decomposition." *IEEE Transactions on Image Processing*, 21 (5), 2898-2909.
- [123] Mostafa, M.G., Farag, A.A., and Essock, E. (2000). "Multimodality image registration and fusion using neural network." *Proceedings of the Third International Conference on Information Fusion*, 2, Wed3/3-Wed3/9.
- [124] Li, W., and Zhu, X. (2005). "A new algorithm of multi-modality medical image fusion based on pulse-coupled neural networks." In *Advances in Natural Computation*, Springer, 995-1001.
- [125] Wang, Z., and Ma, Y. (2008). "Medical image fusion using *m*-PCNN." *Information Fusion*, 9 (2), 176-185.
- [126] Teng, J., Wang, S., Zhang, J., and Wang, X. (2010). "Neuro-fuzzy logic based fusion algorithm of medical images." *3rd International Congress on Image and Signal Processing (CISP)*, 4, 1552-1556.
- [127] Zaveri, T., Makwana, I., and Zaveri, M. (2010). "A fuzzy based hybrid multispectral image fusion method using DWT." *10th International Conference on Hybrid Intelligent Systems (HIS)*, 13-18.

- [128] Wang, N., Ma, Y., and Wang, W. (2014). "DWT-based Multisource Image Fusion Using Spatial Frequency and Simplified Pulse Coupled Neural Network." *Journal of Multimedia*, 9 (1), 159-165.
- [129] Larson, G.W., Rushmeier, H., and Piatko, C. (1997). "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes." *IEEE Transactions on Visualization and Computer Graphics*, 3 (4), 291-306.
- [130] Ashikhmin, M., and Goyal, J. (2006). "A Reality Check for Tone-Mapping Operators" *ACM Transactions on Applied Perception (TAP)*, 3 (4), 399-411.
- [131] Tumblin, J. (1999). "Three methods of detail-preserving contrast reduction for displayed images." PhD thesis, College of Computing Georgia Institute of Technology.
- [132] DiCarlo, J.M., and Wandell, B.A. (2001). "Rendering high dynamic range images." In *Proceedings of the SPIE: Image Sensors*, 2965, 392-401.
- [133] Devlin, K., Chalmers, A., Wilkie, A., and Purgathofer, W. (2002). "Star: Tone reproduction and physically based spectral rendering." In *State of the Art Reports, Eurographics*, 101-123.
- [134] Battiato, S., Castorina, A., and Mancuso, M. (2003). "High dynamic range imaging for digital still camera: an overview." *Journal of Electronic Imaging*, 12, 459-469.
- [135] Vishwakarma, A.K., and Mishra, A. (2012). "Color Image Enhancement Techniques: A Critical Review." *Indian Journal of Computer Science and Engineering (IJCSE)*, 3 (1), 39-45.
- [136] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003). "Adaptive logarithmic mapping for displaying high contrast scenes." In *Computer Graphics Forum*, 22 (3), 419-426. Blackwell Publishing, Inc.
- [137] Lal, S., and Chandra, M. (2014). "Efficient Algorithm for Contrast Enhancement of Natural Images." *The International Arab Journal of Information Technology*, 11 (1), 95-102.
- [138] Li, Z., Zhen, J., Zhu, Z., Yao, W., Wu, S., and Rahardja, S. (2013). "Content adaptive bilateral filtering." *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 1-6.
- [139] Land, E.H., and McCann, J.J. (1971). "Lightness and retinex theory." *Journal of the Optical Society of America*, 61, 1-11.
- [140] Zosso, D., Tran, G., and Osher, S. (2013). "A unifying retinex model based on non-local differential operators." *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics.
- [141] Robinson, P., Roodt, Y., and Nel, A. (2012). "Adaptive Multi-Scale Retinex algorithm for contrast enhancement of real world scenes." In *Proceedings of 23rd Annual Symposium of the Pattern Recognition Association of South Africa*, Pretoria, South Africa.
- [142] Vonikakis, V., Andreadis, I., and Gasteratos, A. (2007). "Fast Dynamic Range Compression for Grayscale Images." *International Workshop on Advanced Image Technology 2007 (IWAIT 2007)*, Bangkok, Thailand.

- [143] Kwon, H.-J., Lee, S.-H., Lee, G.-Y., and Sohng, K.-I. (2014). "Luminance adaptation transform based on brightness functions for LDR image reproduction." *Digital Signal Processing*, In Press. Available online 1 April 2014.
- [144] Naik, N., and Mishra, A. (2015). "Low Contrast Image Enhancement using Wavelet Transform based Algorithms: A Literature Review." *International Journal of Engineering and Technical Research (IJETR)*, 3 (6), 123-128.
- [145] Li, Y., Sharan, L., and Adelson, E.H. (2005). "Compressing and companding high dynamic range images with subband architectures." *ACM Transactions on Graphics*, 24 (3), 836-844.
- [146] Hanika, J., Dammertz, H., and Lensch, H.P.A (2011). "Edge-Optimized À-Trous Wavelets for Local Contrast Enhancement with Robust Denoising." *Computer Graphics Forum*, 30 (7), 1879-1886.
- [147] Bhardwaj, A., and Singh, M.K. (2012). "A Novel approach of medical image enhancement based on Wavelet transform." *International Journal of Engineering Research and Applications (IJERA)*, 2 (3), 2356-2360.
- [148] Indira, K.P., and Hemamalini, R.R. (2012). "A Method for Contrast Correction and Enhancement For Medical Images using Wavelet Fusion." *2012 International Conference on Computing and Control Engineering (ICCCCE 2012)*.
- [149] Lee, J.W., Park, R.-H., and Chang, S.K. (2012). "Noise Reduction and Adaptive Contrast Enhancement for Local Tone Mapping." *IEEE Transactions on Consumer Electronics*, 58 (2), 578-586.
- [150] Loza, A., Bull, D.R., Hill, P.R., and Achim, A.M (2013). "Automatic Contrast Enhancement of Low-Light Images Based on Local Statistics of Wavelet Coefficients." *Digital Signal Processing*, 23 (6), 1856-1866.
- [151] Starck, J.L., Murtagh, F., Candes, E.J., and Donoho, D.L. (2003). "Gray and color image contrast enhancement by the curvelet transform." *IEEE Transactions on Image Processing*, 12, 706-717.
- [152] Zakaria, M.F., Ibrahim, H., and Suandi, S.A. (2010). "A Review: Image Compensation Techniques." *2nd International Conference on Computer Engineering and Technology*, 7, 404-408.
- [153] Liu, T.T., Jiang, Y.D., Ding, W.Y., Meng, X.S., and Wang, X. (2013). "A New High-Dynamic-Range compression and detail Enhancement Algorithm for Infrared Images." *Applied Mechanics and Materials*, 427-429, 1813-1816.
- [154] Chouhan, R., Kumar, C.P., Kumar, R., and Jha, R.K. (2012). "Contrast Enhancement of Dark Images using Stochastic Resonance in Wavelet Domain." *International Journal of Machine Learning and Computing*, 2 (5), 711-715.
- [155] Rajput, Y., Rajput, V.S., Thakur, A., and Vyas, G. (2012). "Advanced Image Enhancement Based on Wavelet & Histogram Equalization for Medical Images." *IOSR Journal of Electronics and Communication Engineering (IOSRJECE)*, 2 (6), 12-16.

- [156] Arun, M., Karthikeyan, R., and Ramya, A. (2013). "Multi Resolution And Contrast Enhancement Using Wavelet Transforms and Singular Value Decomposition." *International Journal of Advanced Trends in Computer Science and Engineering*, 2 (2), 209-215.
- [157] Unaldi, N., Temel, S., and Demirci, S. (2013). "Undecimated Wavelet Transform Based Contrast Enhancement." *World Academy of Science, Engineering and Technology, International Science Index 81, International Journal of Computer, Information Science and Engineering*, 7 (9), 38-41.
- [158] Tai, S.-C., Chen, Z.-S., and Lin, S.-C. (2013). "Adaptive tone reproduction operator based on an ideal perceptive luminance of human visual system." Available online from: <http://pdcat13.csie.ntust.edu.tw/download/papers/P10006.pdf>.
- [159] Ahmed, N., Natarajan, T., and Rao, K.R. (1974). "Discrete Cosine Transform." *IEEE Transactions on Computers*, C-23 (1), 90-93.
- [160] Huffman, D.A. (1962). "A method for the construction of minimum redundancy codes." In *Proceedings IRE*, 40, 1098-1101.
- [161] Pennebaker, W.B, Mitchell, J.L., et. al. (1988). "Arithmetic coding articles." *IBM Journal of Research and Development*, 32 (6), 717-774.
- [162] Wallace, G.K. (1992). "The JPEG Still Picture Compression Standard." *IEEE Transactions on Consumer Electronics*, 38 (1), 18-34.
- [163] Christopoulos, C., Skodras, A., and Ebrahimi, T. (2000). "The JPEG2000 Still Image Coding: An Overview." *IEEE Transactions on Consumer Electronics*, 46 (4), 1103-1127.
- [164] Cohen, A., Daubechies, I., and Feauveau, J.C. (1992). "Biorthogonal bases of compactly supported wavelets." *Communications on Pure and Applied Mathematics*, 45 (5), 458-560.
- [165] LeGall, D., and Tabatabai, A. (1988). "Subband Coding of Digital Images Using Symmetric Short Kernel Filters and Arithmetic Coding Techniques." *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 761-765.
- [166] Marcellin, M.W., Gormish, M.J., Bilgin, A., and Boliek, M.P. (2000). "An overview of JPEG-2000." In *Proceedings of 2000 Data Compression Conference*, 523-544.
- [167] Taubman, D. (2000). "High Performance Scalable Image Compression with EBCOT." *IEEE Transactions on Image Processing*, 9 (7), 1158-1170.
- [168] Santa-Cruz, D., Ebrahimi, T., Askelof, J., Larsson, M., and Christopoulos, C. (2000). "JPEG 2000 Still Image Coding Versus other Standards." ISO/IEC JTCl/SC29/WG1 (ITU-T SG8).
- [169] Guo, L., Au, O. C., Ma, M., and Liang, Z. (2007). "Temporal video denoising based on multihypothesis motion compensation." *IEEE Transactions on Circuits and Systems for Video Technology*, 17 (10), 1423-1429.

- [170] Rajagopalan, R., and Orchard, M. T. (2002). "Synthesizing processed video by filtering temporal relationships." *IEEE Transactions on Image Processing*, 11 (1), 26-36.
- [171] Sweldens, W. (1995). "The Lifting Scheme: A new philosophy in biorthogonal wavelet constructions." In *Proceedings of SPIE*, 2569, 68-79.
- [172] Daubechies, I., and Sweldens, W. (1998). "Factoring wavelet transforms into lifting steps." *Journal of Fourier Analysis and Applications*, 4 (3), 247-269.
- [173] Fairhurst, A.M. (1998). "Method of predicting the probability of human observers recognizing targets in simulated thermal images." *Optical Engineering*, 37 (3), 744-751.
- [174] Healey, G.E., and Kondepudy, R. (1994). "Radiometric CCD camera calibration and noise estimation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (3), 267-276.
- [175] Boie, R.A, and Cox, I.J. (1992). "An analysis of camera noise." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (6), 671-674.
- [176] Donoho, D.L., and Johnstone, I.M. (1994). "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, 81, 425-455.
- [177] Li, H., Manjunath, B.S., and Mitra, S.K. (1995). "Multisensor image fusion using the wavelet transform." *Graphical Models and Image Processing*, 57 (3), 235-245.
- [178] Arias-Castro, E., and Donoho, D.L "Does median filtering truly preserve edges better than linear filtering?" *Annals of Statistics*, 37 (3), 1172–2009.
- [179] Arce, G.R. (2004). "Nonlinear Signal Processing: A Statistical Approach." John Wiley & Sons, Inc.: Hoboken, New Jersey, USA.
- [180] Huang, T., Yang, G., and Tang, G. (1979). "A fast two-dimensional median filtering algorithm." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27 (1), 13-18.
- [181] Yoshida, A., Blanz, V., Myszkowski, K, and Seidel, H.P. (2005). "Perceptual Evaluation of Tone Mapping Operators with Real-World Scenes," *Proceedings of the SPIE*, 5666, 192-203.
- [182] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H. (2005). "Evaluation of Tone Mapping Operators using a High Dynamic Range Display," *ACM Transactions on Graphics (TOG)*, 24 (3), 640-648.
- [183] Fattal, R., Lischinski, D., and Werman, M. (2002). "Gradient Domain High Dynamic Range Compression," *ACM Transactions on Graphics (TOG)*, 21 (3), 249-256.
- [184] Durand, F., and Dorsey, J. (2002) "Fast Bilateral Filtering for the Display of High-Dynamic-Range Images," *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press, 257-266.
- [185] Aguilar, M., Grater, J., Tarquini, R., and Johnson, J. (2006). "Enhancement, fusion, and visualization of third generation FPA imagery." *Meeting of the Military Sensing Symposia (MSS), Specialty group on Passive Sensors*. Orlando, FL.

- [186] Wang, Z. (2016). "Objective image quality assessment: Facing the real-world challenges." *Proceedings of IS&T Electronic Imaging, Image Quality and System Performance XIII*, 1-6.
- [187] Ghadiyaram, D., and Bovik, A.C. (2016). "Massive Online Crowdsourced Study of Subjective and Objective Picture Quality." *IEEE Transactions on Image Processing*, 25 (1), 372-287.
- [188] Blanchette, J., and Summerfield, M. (2008). *C++ GUI Programming with Qt 4, Second Edition*. Upper Saddle River, NJ: Prentice Hall Press.
- [189] Miller, R.B. (1968). "Response time in man-computer conversational transactions." *Proceedings of AFIPS Fall Joint Computer Conference*. 33, 267-277.
- [190] U.S. National Archives and Records Administration. (2012) *Code of Federal Regulations*. Title 14, Flight Simulation Training Device Initial and Continuing Qualification and Use. (Section 60)
- [191] Link, N.K, Kruk, R.V., McKay, D., Jennings, S., and Craig, G. (2002). "Hybrid Enhanced and Synthetic Vision System Architecture for Rotorcraft Operations." *Proceedings of SPIE: Enhanced and Synthetic Vision*. 4713, 190-201.
- [192] Bailey, R.E, Arthur, J.J., Williams, S.P., and Kramer, L.J. (2005). "Latency in visionic systems: test methods and requirements." *NASA Technical Report RTO-MP-HFM-125*.
- [193] Ivkovic, Z., Stavness, I., Gutwin, C., and Sutcliff, S. (2015). "Quantifying and mitigating the negative effects of local latencies on aiming in 3D shooter games." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 135-144.
- [194] Movidius, Inc. (2016). "Myriad 2 MA2x5x Vision Processor." Online at https://uploads.movidius.com/1463156689-2016-04-29_VPU_ProductBrief.pdf.
- [195] NXP Semiconductors. (2016). "Automotive Vision Systems." Online at <http://www.nxp.com/applications/solutions-for-the-iot-and-adas/automotive/adas-autonomous-driving/automotive-vision-systems:VISION-PROCESSING-SYSTEMS#overviewExpand>.

APPENDIX A SUPPLEMENTARY IMAGE DATA

Multi-Focus Fusion Results

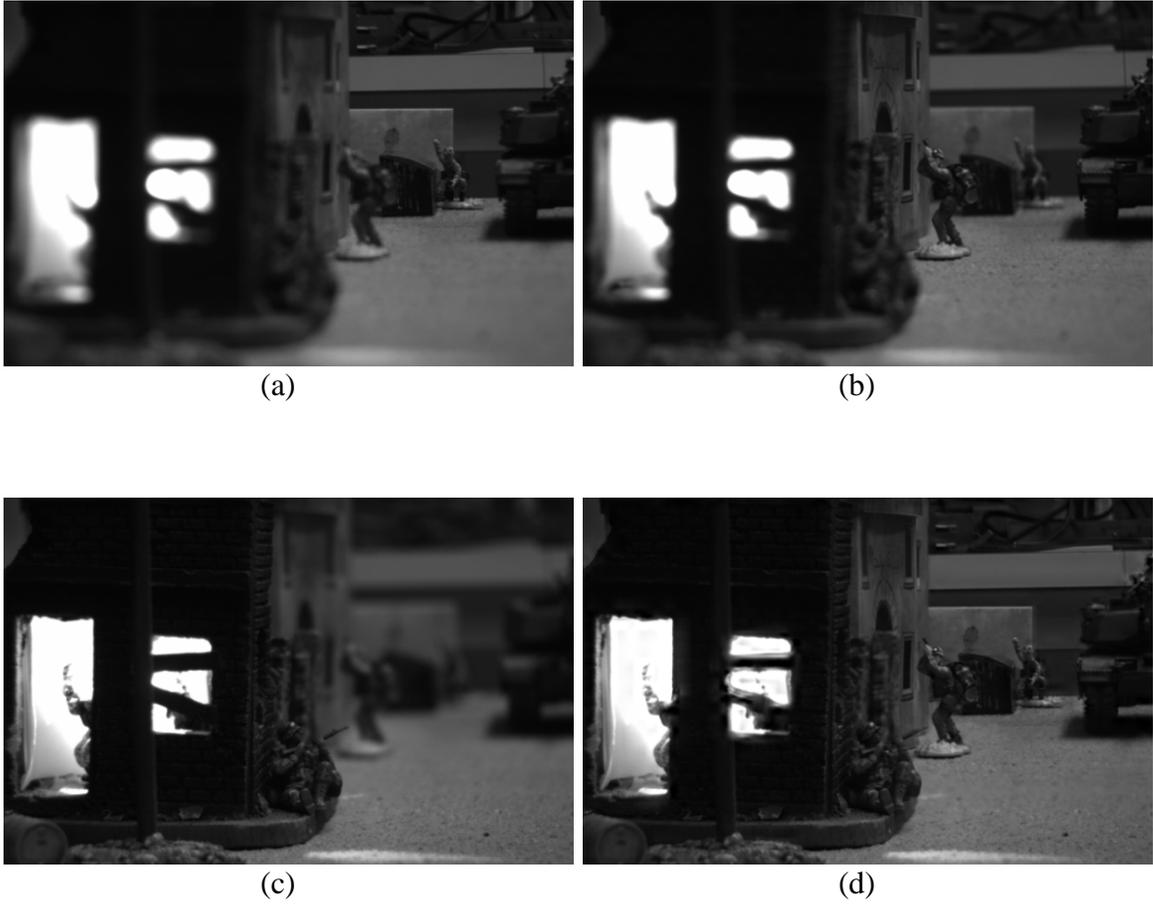


Figure A.1: Example (a) far focus, (b) mid focus, (c) and near focus inputs with (d) fused results

Comparison of Adaptive Scaling of High-Frequency Sub-bands



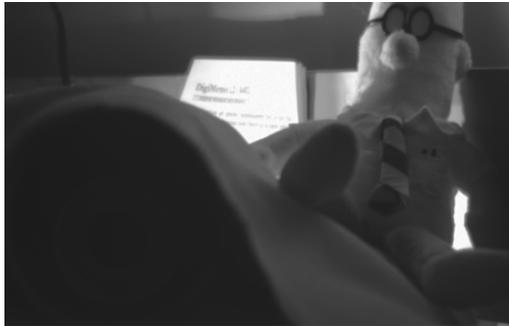
(a)



(b)

Figure A.2: DRC with (a) a single scalar and (b) adaptive scaling of detail coefficients

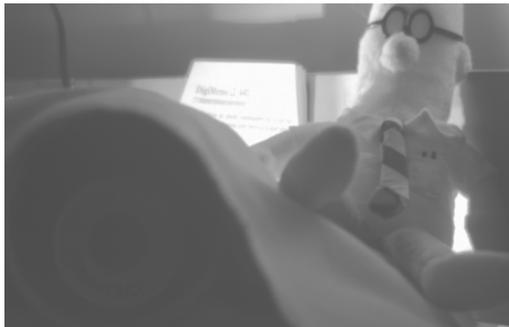
Images Used in DRC SME Evaluation



(a)



(b)



(c)



(d)



(e)



(f)

Figure A.3: Dilbert (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results

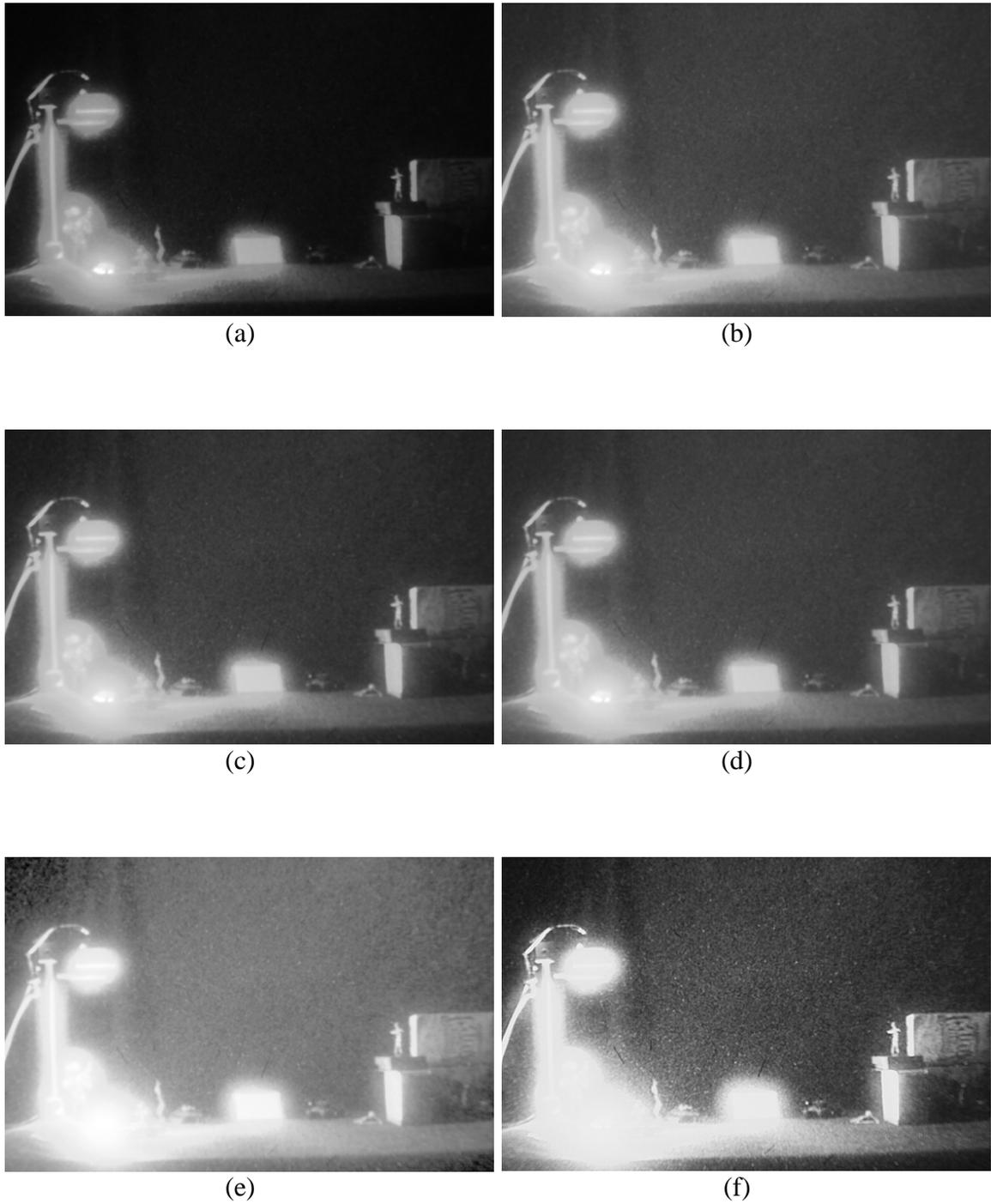
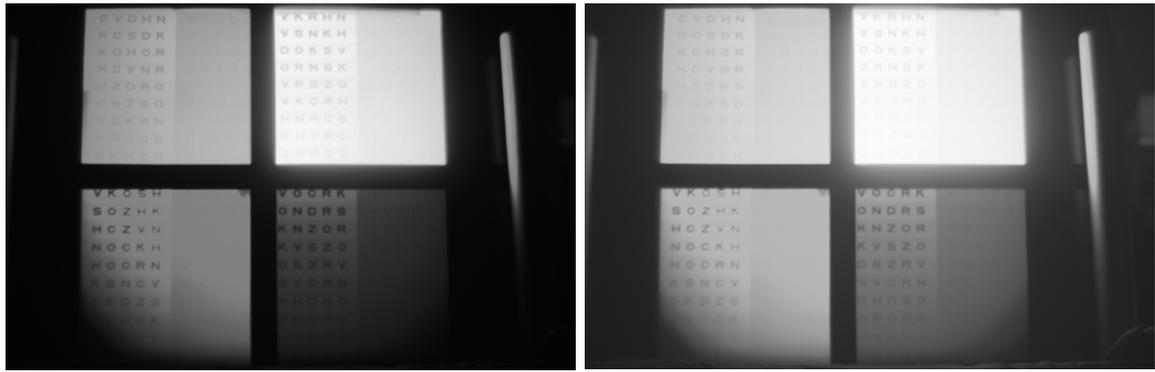
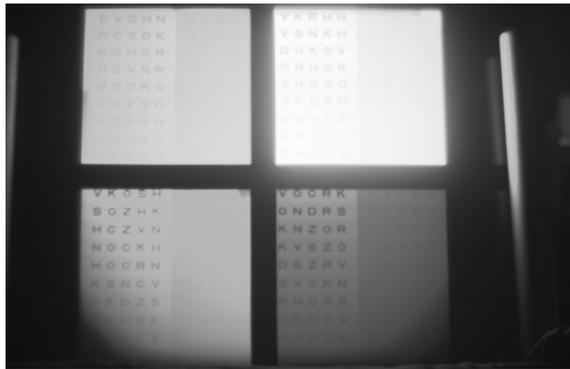


Figure A.4: Diorama (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results

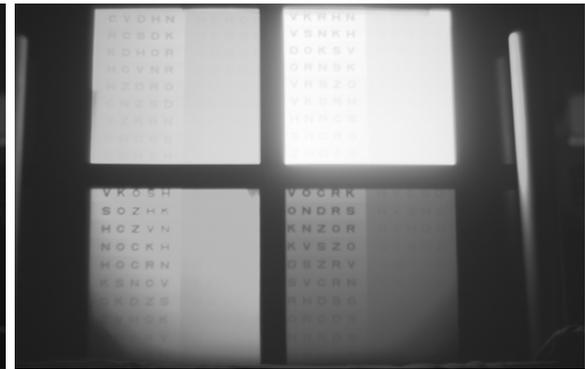


(a)

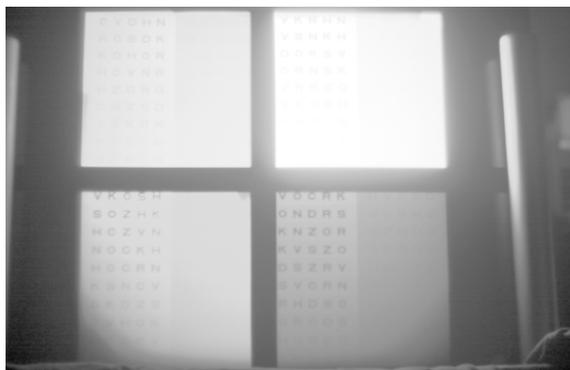
(b)



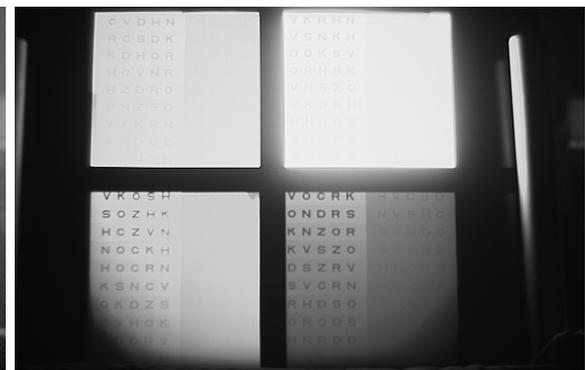
(c)



(d)



(e)



(f)

Figure A.5: Matrix (a) HDR Source Image, (b) Bilateral, (c) Drago, (d) Gradient, (e) Histogram Equalization, and (f) Biologically Inspired DRC Results

APPENDIX B DT-CWT WAVELET FRAMEWORK

A brief investigation was conducted into utilizing the DT-CWT in place of the LeGall 5/3 wavelet. This approach has six angularly selective sub-bands, shown in Figure B.1, and a spectrum structure that reduces the frequency aliasing and other artifacts introduced by decimation, which were observed in early implementations of the original processing chain.

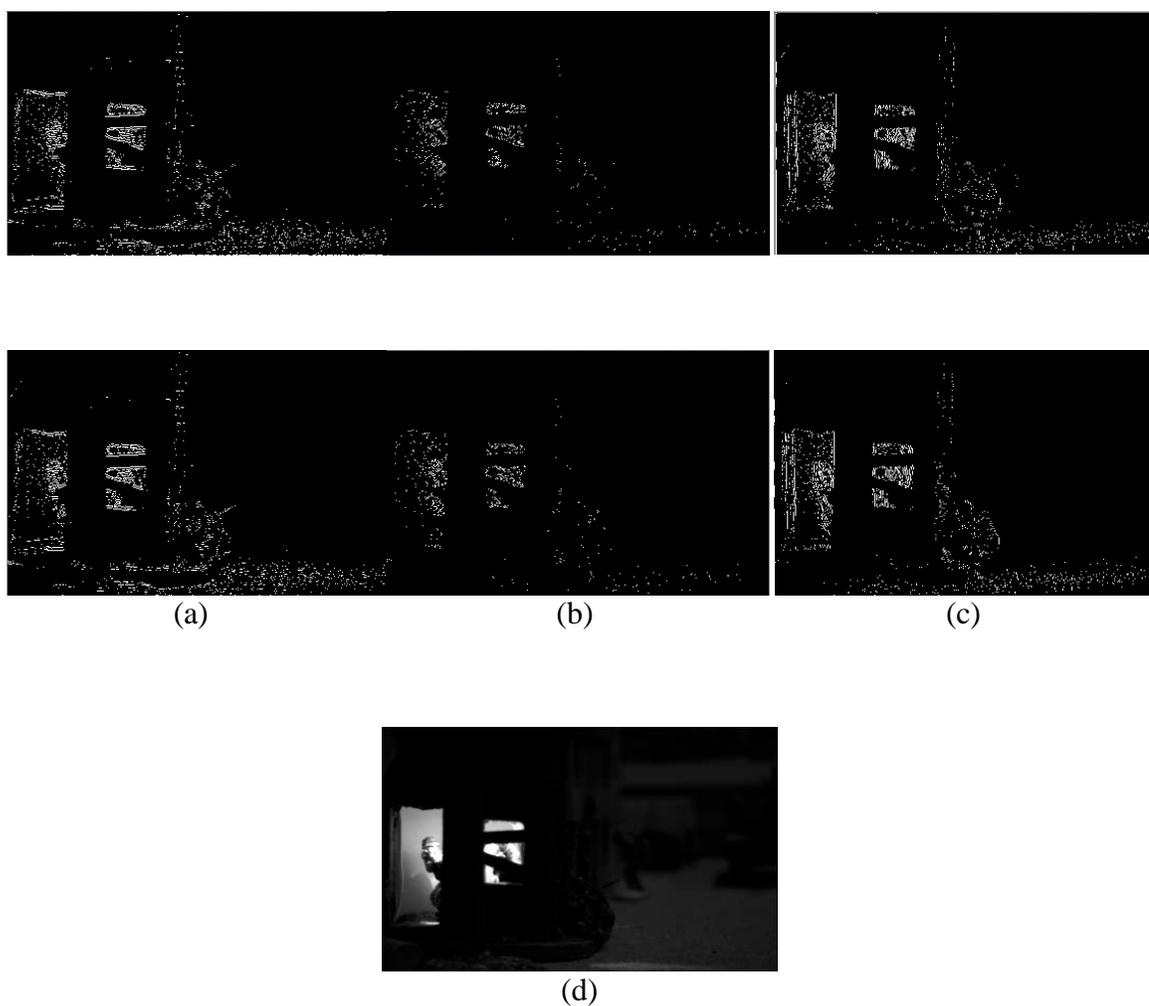


Figure B.1: DT-CWT (a) HL Orientations, (b) HH Orientations, (c) LH Orientations, (d) LL approximation coefficients

Similar to the DWT framework, the results produced by the DT-CWT framework are highly influence by the DRC algorithm. This can clearly be seen from the sample image results provided in Figure B.2.



Figure B.2: Sample image results after applying the DT-CWT processing chain

Subjective assessment of the image results from this processing chain elicited a range of feedback, including highly positive feedback on the results in Figure B.2b. Despite this positive feedback, there is significant additional work that would be required to reduce the execution time to acceptable levels. The average execution time observed during this preliminary assessment was approximately 3.5 seconds, nearly 5 times the execution time of the benchmark framework and 42 times the execution time of the DWT framework based on the LeGall wavelet.