

A Content Enhancement Framework for Multi-Projector Systems

by

Andrew Hryniowski

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2018

© Andrew Hryniowski 2018

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The content from one published conference paper and two future publications are included in this thesis with contributions from others. The contributions for each work are listed below. For each of these work I performed the majority of the design conceptualization, background review, experimentation, writing, and editing.

The published work "Multi-Projector Resolution Enhancement Through Biased Interpolation", in Conference of Computer and Robotic Vision (CRV), 2018, contains much of the content related to the Biased Interpolation model presented in Chapter 6.1. The paper's authors, excluding myself, include Ibrahim Ben Daya, Dr. Ahmed Gawish, Mark Lamm, Dr. Alexander Wong, and Dr. Paul Fieguth. Ibrahim Ben Daya supported with some of the experimental results, and helped with the writing. All the authors supported the problem conceptualization, provided guidance regarding the publication's structure, and helped with proofing the paper.

The first group of work planned for publication relates to the following content in this document: Sub-Pixel Integration in Chapter 4, measuring multi-projector configuration properties in Chapter 5, and the proposed Optimal model in Chapter 6.3. The paper's authors, excluding myself, will include Dr. Ahmed Gawish, Mark Lamm, Dr. Alexander Wong, and Dr. Paul Fieguth. All the authors supported the problem conceptualization. Dr. Alexander Wong and Dr. Paul Fieguth provided useful advice on the structure of the report. The intent is to publish this work in IEEE Transactions on Computational Imaging.

The second group of work planned for publication relates to the Kernel models presented in Chapter 6.2. The paper's authors, excluding myself, will include Dr. Zohreh Azimifar, Mark Lamm, Dr. Alexander Wong, and Dr. Paul Fieguth. The intent is to publish this work at the Conference on Vision and Imaging Systems.

Abstract

Projectors are a convenient technology for displaying content on large, abnormal, or temporary surfaces where mounting other forms of light emitting devices is too impractical or too expensive. Common uses of projectors include movie cinemas, concert halls, 3D model colourization, planetariums, etc. Many of these applications require multiple projectors to either cover the entire display surface, like planetariums, or to achieve the require brightness, like outdoor projection.

Aligning the content between projectors is typically required to ensure that overlapping regions between projectors display the same content. Naive approaches of aligning content treat the relationship between the content and a projector independently of all other projectors in the configuration. Aligning content can limit the quality of the superimposed image as high frequency signals are often degraded during the alignment process. Previous works have shown it is possible to improve the perceptual quality of the aligned content by giving each content-to-projector transformation prior knowledge of all projectors in the configuration. However, these works either make theoretical assumptions, require special hardware, severely limit the types of applications their systems work on, or only use qualitative analysis to evaluate their system’s performance.

In this work, a framework capable of simulating a multi-projector configuration for any number of projectors on a flat surface is proposed. A method of comparing the ideal content with the projected content is developed using the proposed simulation in conjunction with an existing image comparison technique. Different system setups are tested for a two projector configuration. The quality of each configuration is measured using the developed comparison metric across a dataset of natural images. Finally, the proposed framework is used to train three different models, in an end-to-end fashion, that are capable of improving the perceptual quality of the superimposed image.

The first two models are parametric and content independent, while the third model is non-parametric and content dependent. The first model directly integrates with existing interpolation methods used during the content-to-projector alignment. The second model applies a post transformation filtering operation using a set of learned linear convolutional kernels. The third model directly optimizes the projected images to improve the perceptual quality of the superimposed image.

Acknowledgements

I would like to thank my two co-supervisors Dr. Paul Fieguth and Dr. Alexander Wong for providing me with a nourishing environment in which I have been able to continue my research oriented education. Thank you both for stoically enduring my development as a researcher, and for providing me with two examples of dedicated researches. I aspire to one day have the thoughtfulness and thoroughness of you both.

I would like to thank both Dr. David Clausi and Dr. Bryan Tripp for serving as readers of my Master's thesis.

I would also like to thank all the professors and students in the Video and Image Processing Lab. This group of people have an unending inquisitiveness that I relish being a part of; I look forward to interacting with all of you everyday.

In addition, I would like to thank Mark Lamm and Christie Digital for providing me an excellent internship opportunity through which I have been able to experience industry oriented research.

Most importantly, I would like to thank my loving parents and friends who have always given me the emotional support I have needed. My life would be for naught without all of you.

Table of Contents

List of Figures	viii
1 Introduction	1
1.1 Motivation	1
1.2 Previous Content Enhancement Approaches	2
1.3 Thesis Contributions and Outline	4
2 Background	6
2.1 Multi-Projector Configurations	6
2.2 Content Alignment	8
2.2.1 Projector-to-Content Calibration	8
2.2.2 Space Mapping Models	10
2.2.3 Interpolation	11
2.3 Image Comparison	13
2.4 Other Aspects of Multi-Projector Configurations	15
3 Quantitative Problem Formulation	18
3.1 Problem Conceptualization	18
3.2 Moiré Interference Patterns	19
3.3 Multi-Projector Configuration Constraints	21
3.4 Non-Warping Assumptions	24

4	Sub-Pixel Integration	26
4.1	Sub-Pixel Calculation	27
4.2	Loss Function Adjustment	29
5	Quantitative Properties of Multi-Projector Configurations	32
5.1	SPI-SSIM Pixel-Length-Ratio vs. Rotation	32
5.2	Inter-Rotation vs. Intra-Rotation	34
5.3	Pixel-Length-Ratio Exploration	37
6	Content Improvement Models	42
6.1	Biased Interpolation	44
6.2	Kernel Models	45
6.2.1	Kernel-Space Model	45
6.2.2	Kernel-Limited Model	46
6.3	Optimal Model	47
6.4	Results	47
6.4.1	Kernel-Space Spanning Set	47
6.4.2	Kernel-Space vs. Kernel-Limited	49
6.4.3	Model Comparison	50
7	Conclusions	57
7.1	Summary of Thesis and Contributions	57
7.2	Impact	58
7.3	Future Work and Preliminary Results	59
7.3.1	Improved Comparison Metric	59
7.3.2	Complex Surfaces	60
7.3.3	Testing More than Two Projectors	60
7.3.4	Parametric Content Dependent Models	60
	References	62

List of Figures

1.1	Biased Interpolation Example Results	3
2.1	Projector Configuration Comparison	7
2.2	Examples of Projector Calibration Methods	10
2.3	MSE vs. SSIM	15
2.4	Multi-Projector System Overview	16
3.1	Single Projector Content Correction Example	20
3.2	Pixel-Grid Density vs. Pixel-Grid Resolution	21
3.3	Moiré Pattern Comparison	22
4.1	Sup-Pixel Integration Example	28
4.2	Sub-Pixel Integration Applied to SSIM	29
5.1	SPI-SSIM, Pixel-Length-Ratio vs. Rotation	33
5.2	Comparing Individual Projected Images	35
5.3	Inter-Rotation vs. Intra-Rotation	36
5.4	Optimal Configuration Comparison	38
5.5	SPI-SSIM vs. Pixel-Length-Ratio	39
5.6	Pixel-Length-Ratio Image Examples	41
6.1	Learning Framework Overview	43
6.2	Kernel-Space Spanning Set Comparison	48

6.3	Kernel-Limited Model, Performance vs. Number of Kernels	49
6.4	Content Independent Model Comparison	51
6.5	Text - Optimal Two Projector Configuration Example	53
6.6	Text - 22.5 Degrees Two Projector Configuration Example	54
6.7	Building - Optimal Two Projector Configuration Example	55
6.8	Building - 22.5 Degree Two Projector Configuration Example	56

Chapter 1

Introduction

Projectors are optical devices that reconstruct a specified pattern using millions of pixels on a given surface. The patterns that these devices can display are limited only by the arrangement of pixels, the number of pixels, the refresh rate of the projector, and the owners imagination. Projectors are used for a wide variety of tasks, including, cinema projectors, flight simulators, 3D model colourization, side-of-building projection, etc. Many of these applications require more than one projector to cover the entire display surface since a single projector may not have sufficient pixel density, or the display surface is an irregular shape and covering the entire surface is impossible with a single projector (e.g., a round display surface). Other applications require multiple projectors to increase the redundancy of the entire system in case of projector failure, increase the total brightness of the content being projected, or to increase the perceptual quality of the content beyond the capabilities of a single projector.

1.1 Motivation

Aligning the same content across multiple overlapping projectors is a simple idea in principle. In practice, however, aligning content between projectors is a complex task as the differences between projectors are not easily overcome. For example, the position of each projector relative to the display surface has a large role in determining the achievable content quality of the final superimposed image. Other inter-projector properties that affect the quality of the alignment include projector colour responses, the focus of each projector at a given location of the display surface, and the quality of pixel correspondences used to generate a transformation between spaces.

One specific area that has shown to improve the perceptual quality of the superimposed image is to have the images from each projector *co-operate* with one another. That is, filter each sub-image so that when stacked they produce a superimposed image with an improved perceptual quality. It is common for the content's frame-of-reference to be different than each projector's frame-of-reference to ensure that the content is correctly aligned with the display screen. This requires that each projector sample the correct portion of the content image such that the content is aligned between projectors for any given overlapping segment. The transformation between spaces relies on interpolation to approximate the content when a sampling point lies between content pixels. The interpolation process results in a degradation of high frequency patterns. Normal space transformations and sampling procedures only consider the spaces being operated on; the functions are agnostic to the state of other existing space. This is disadvantageous for multi-projector configurations as the overlapping pixel grids may complement or interfere with each others content. A better superimposed image may be achieved by providing the each content-to-projector space transformation with prior knowledge of all other existing spaces.

A key component to improving the perceptual quality of the realized image is the effective resolution at any given area. A greater resolution allows higher frequency content signals to be projected. The density of pixels within an area increase with the number of projectors. This effectively means that two projectors acting together, each with the same given resolution, have the potential to realize a greater superimposed content resolution than either projector could separately (on average). Figure 1.1 shows the results from one of the models presented in this work.

1.2 Previous Content Enhancement Approaches

Image stacking can be achieved through a single projector or with multiple projectors. In the single projector case the concept of wobulation [1] is used. An opto-mechanical image shifter moves the entire pixel-grid by a fraction of a pixel in both axes. The single projector operates at double the frequency of a normal projector and alternates between the shifted and un-shifted images. The effectiveness of this system is due to the precise global fractional pixel shifts achieved. Wobulation is cumbersome and expensive to implement in practice [2]. Scaling this technology to more than two shifted images drastically increases the cost of the projector as the required operating frequency of the projector scales linearly with the number of shifted images.

Unlike wobulation, the images from each projector in a multi-projector configuration have distinct points of projection. This requires the sub-images to be transformed such that



Figure 1.1: An example of the achieved resolution enhancement using the Biased Interpolation system. There are four sub-figures, each showing the same content under different conditions. Each sub-figure shows the alphabet in arial text at different font sizes, and a zoomed-in image of three letters *stu*; marked with a red boarder. Top-left is the original content, top-right is a simulation of a single projector, the bottom-left is a simulation of two projectors, and the bottom-right is a simulation of two projectors with the proposed filtering applied. Notice the quality of the *t* before and after a second projector is added. With one projector the *t* appears to be cut in half with one half shifted by a pixel. When the second projector is added the *t* appears like a single undivided entity. The *t* then becomes less blurry once the filtering model is applied.

the content from each image is aligned. Precise calibration is required as any misalignment will degrade the quality of the superimposed approximate high-resolution image [3]. Having different points of projection creates a non-uniformity in pixel shift patterns throughout the overlapping region [4]. Content enhancement is only attainable when fractional pixel shifts between projectors are present within a local region of the overlapping projector field [4]. Regions with no fractional pixel shifts are only capable of producing a direct average of the content being projected from each projector.

Work in multi-projector perceptual quality enhancement has focused on developing post alignment filter banks to sharpen images in either the spatial or frequency domain [2,3,4,5,6,7]. These works use a combination of simulated results, captured images, and qualitative comparison to measure their algorithm’s effectiveness. Directly comparing techniques outside a simulated domain is difficult as projector properties differ, developed systems solve a varying degree of sub-problems within multi-projector content enhancement, or the developed system is only shown to work with an unrealistic constant global pixel overlapping pattern (outside of wobulation type configurations).

Metrics used in multi-projector resolution enhancement systems include Structured Similarity (SSIM) [8], its variants, and Mean Squared Error (MSE). Both have been used as comparisons metrics but SSIM has never been used to directly train such a system [2,3,4,5,6,7].

1.3 Thesis Contributions and Outline

The purpose of this thesis is to develop methods to measure and improve upon the perceptual quality of the superimposed images produced by a multi-projector configuration. The three main contributions of this thesis are

1. Developing a framework to quantitatively measure the perceptual quality of a superimposed image with respect to the ideal content, Chapter 4.
2. Evaluating the degradation of perceptual quality across a variety of two projector configurations, Chapter 5.
3. Using the proposed framework to learn three different models that are capable of improving the perceptual quality of superimposed images, Chapter 6.

The rest of the this work is structured as follows, Chapter 2 reviews the knowledge required to understand the contributions of this work. Different types of multi-projector

configurations are distinguished, standard content alignment procedures are reviewed, image comparison techniques are stated, and different parts of a multi-projector pipeline are discussed. Chapter 3 provides an in depth look at what causes perceptual image quality degradation in multi-projector configurations and how improving the perceptual quality of the projected superimposed image is possible. In addition, assumptions made in the problem formulation are listed. Chapter 4 proposes a method of quantitatively measuring the perceptual image quality of the superimposed projected images. This method first simulates the superimposed image by reducing pixels from each projector into small components called sub-pixels. A standard image comparison technique is then adjusted to accommodate the sub-pixel representation of images. Chapter 5 analyses the performance of naive multi-projector configurations using the proposed comparison metric. Chapter 6 proposes three different models to improve the perceptual quality of superimposed images. Two of the models are parametric and content independent, and the third model is non-parametric and content dependent. A variety of tests are performed on each model individually, and all three models are compared. Finally, Chapter 7 summarizes the thesis, restates conclusions, and provides possible future avenues of research.

Chapter 2

Background

This chapter reviews relevant information required to understand the proposed content in this work. First, various types of multi-projector configurations are distinguished in Section 2.1. Next, Section 2.2 formalizes aspects of inter-projector content alignment, this includes calibrating every projector to a common frame-of-reference, building a model for each calibration, and moving content between spaces. Then, Section 2.3 defines metrics for comparing images. Finally, Section 2.4 reviews a more complete picture of multi-projector systems.

2.1 Multi-Projector Configurations

A multi-projector configuration is a system consisting of more than one aligned projectors where each unit is displaying a portion of, or all of, the desired content. A key feature of multi-projector configurations is the amount of overlap between the projectors. Overlap within a configuration ranges from no overlap between the projectors to complete overlap between projectors. The type of overlap in the systems is generally determined by the application, and more specifically, the characteristics of the display surface and the desired content.

Systems with no overlap, Figure 2.1a, between projectors are used for applications where the content between projectors is non-contiguous, or when significant time can be spent on physically aligning each projector with all neighbouring projectors. The latter scenario is extremely difficult to achieve in practice as minor differences in projector placement or lens characteristics will result in a misalignment. This scenario essentially limits physical alignment to flat display surfaces or where custom projector frames are built.

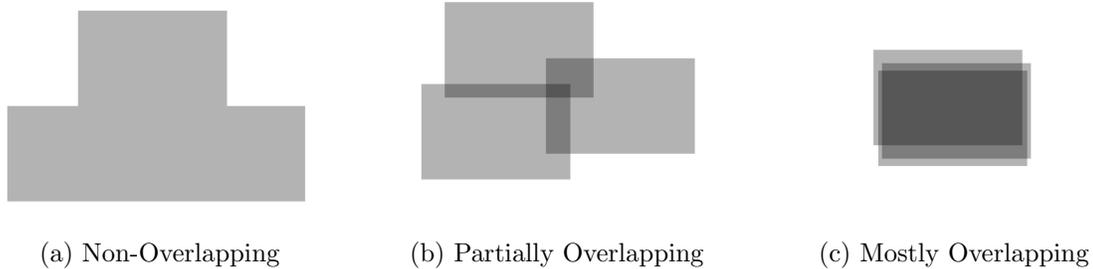


Figure 2.1: This figure compares the types of overlapping projector configurations. Each image has three projectors represented by grey rectangles. Notice that overlapping regions appear darker; the more overlapping regions the darker the region is. In reality this relationship is flipped and the regions get brighter due to back light bleed from each projector.

Partially overlapping configurations, Figure 2.1b, have limited overlap between neighbouring projector pixel-grids. These configurations are well suited for applications where a single projector cannot cover the entire display, an increase in total screen resolution is required (to project large content), and when the display surface is an abnormal shape (e.g, projection on a 3D model). Examples of this scenario include any display with an abnormal height-to-width ratio, planetariums, and projecting on the side of a building. The difference between projector properties is most apparent for non-overlapping and partially overlapping configurations. One important configuration characteristic is the inter-luminance difference between each projector; over time individual projectors can age differently resulting in some projectors becoming dimmer than others. Another difference is the colour response of a given projector. There are a variety of projector technologies and each has its own unique colour response. This can result in perceptually different colours despite the same intended colour (e.g., two projectors each displaying solid reds, the projected image from one projector appears red while the other projected image appears brownish in comparison). Another issue with partially overlapping regions is the increase in luminance. Superimposed light is additive; this creates a nonuniform luminance across the display surface. Reducing the light from each projector in the overlapping area is a common remedy.

The third type of configuration is where each projector is mostly overlapping, Figure 2.1c, with all other projectors. Projecting the same content with multiple overlapping projectors provides several advantages compared to using a single projector, such as increased brightness to overcome ambient light or display surface anomalies, redundancy in

case of projector failure, an increase in the area being projected on, and the possibility for increased content resolution. Like partially overlapping configurations content alignment is required. An issue for overlapping systems is the increase in back light bleed; this reduces how dark blacks are. Every additional projector added to the overlapping regions typically degrades the quality of edges within a pattern. This is a result of the pixel-grids between the projectors being unaligned.

2.2 Content Alignment

Both the partially overlapping and mostly overlapping configurations require the content to be aligned, shown in Figure 2.1b and Figure 2.1c respectively. This is required because the physical pixel-grids of each overlapping projector pair are unlikely to be physically aligned once setup in the desired location. The process of aligning the content can be broken down into two general stages: alignment calibration, and content transformation. Alignment calibration is the process of generating a model that relates every projector to a common frame-of-reference; namely, the space in which the content is placed. Content warping is the process of applying the alignment model to the desired content. That is, transforming an image from its native representation to a different representation. There are three different categories of 2D spaces required for these two stages of alignment: content space, projector spaces, and camera spaces. The camera space acts as an intermediate between the content space and the projector spaces. Note that this work assumes only one content space but possibly many camera and projector spaces. The remainder of this section defines each type of space, how they relate to each other, methods of space calibration, and the process of moving content between spaces.

2.2.1 Projector-to-Content Calibration

The goal of calibration is to generate a mapping from every point in each projector space to any equivalent point in content space. Let C be the content space. C is the native space of the images being projected. Let P_p be the space for projector $p \in \{1, 2, \dots, n\}$ for an n projector setup. Let V_v be the camera space for camera $v \in \{1, 2, \dots, m\}$ for an m camera setup. Let $H_{S_a \rightarrow S_b}$ model a mapping from space S_a to S_b of a set of points X

$$X^{S_b} = H_{S_a \rightarrow S_b} X^{S_a} \tag{2.1}$$

where X^{S_a} and X^{S_b} are the same set of points represented in spaces S_a and S_b , respectively. Mapping a set of points from some projector space P_p to content space C is defined as

$$X^C = H_{P_p \rightarrow C} X^{P_p} \quad (2.2)$$

Each projector is mapped to some portion of the content space. The camera spaces are used to connect the content space to each of the projectors spaces. Equation 2.2 can be decomposed into two separate mappings, a projector-to-camera mapping and a camera-to-content mapping.

$$X^C = H_{V \rightarrow C} H_{P_p \rightarrow V} X^{P_p} \quad (2.3)$$

Note that this equation assumes that every point in X^{P_p} can be mapped through some camera space V_v to the content space C . Calculating $H_{V \rightarrow C}$ and $H_{P_p \rightarrow V}$ each requires a set of pixel correspondences. These are determined through separate calibration steps. The result of the entire calibration process is to generate a set of pixel correspondences that map every point in each projector to some point in content space.

$$\{X^C, X^{P_p}\} \forall p \in P \quad (2.4)$$

Projector-to-camera calibration is the process of relating one or more world references (cameras) to every projector in the configuration. Structured light based methods [9, 10, 11, 12] are generally used to generate these correspondences. That is, for each projector-camera pair a series of predetermined patterns is projected which a given camera records. A set of pixel correspondences relating camera pixels to projector pixels can be determined using the set of pattern pairs (i.e., the displayed pattern and the captured pattern). The nature of the pixel correspondences depends on the patterns being used and the type of space transformation model being used (these models are discussed in subsection 2.2.2). Typically, there is a trade off between the number of pixel correspondences generated and the accuracy of any individual correspondence. For example, the method of Gray Coding [9] attempts to calculate a correspondence for every projector pixel without assuming a display surface model. Any one point may be noisy and post processing is required to improve the accuracy of such points; algorithms like Random Sample Consensus (RANSAC) [13] may be used.

It is generally required that enough cameras be used to cover the entire display surface. Specifically, enough world reference perspectives, not physical cameras, are required. This allows a single camera to have a non fixed position. A key requirement of multi-perspective world reference system is the ability to quantify the relationship between each perspective. This can be achieved through prior knowledge (predetermined perspective locations), or ensuring perspective overlap and applying a perspective-to-perspective calibration step.

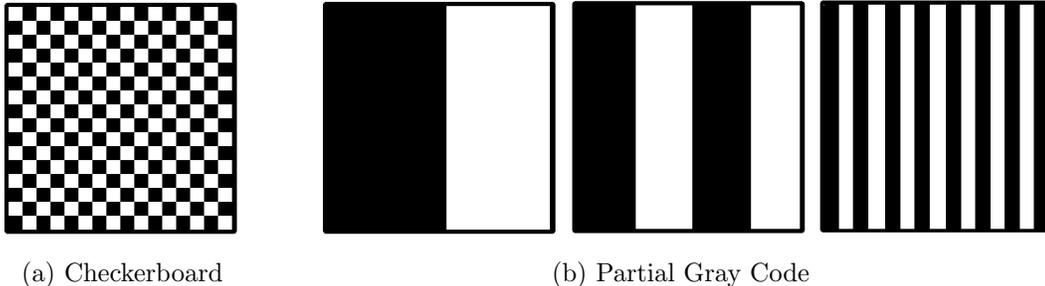


Figure 2.2: This figure shows two different calibration methods. Only a fraction of the total patterns used in the Gray Code algorithm are shown. Note that the black border on each image is for clarity.

The content-to-camera calibration step generates a mapping from the content space C to every camera space V_v . This process is heavily dependent on the application. Specifically, how important is the location and shape of the content on the display surface. Applications with complex screen geometries often require precise content placement, whereas applications with flat screen geometries have greater flexibility since the display surface is usually larger than the content. Correct placement on complex surfaces often requires prior knowledge of the display’s geometry, and a method of determining the display’s orientation relative to each camera. Typically some method of keypoint tracking [14, 15, 16] is required. For example, simple surfaces markers can be placed on the display surface to define the content shape, or just simply draw (in a graphical interface) the location and shape of the content in camera space on a computer monitor.

2.2.2 Space Mapping Models

There are several sources of possible noise during a space-to-space calibration step. This is especially true when physical objects, like a camera or a projector, are involved. Cameras have several properties that must be correctly configured to make an accurate calibration possible. The camera must be focused on the projector content, the correct exposure must be set, the camera sensor must be of sufficient resolution (ideally a many-to-one ratio between camera and projector pixels), and the camera must be synced with the content being displayed. Projectors are mainly required to have sufficient brightness compared to ambient light and to be focused across the projected content. The latter point can be difficult to achieve depending on the projector’s physical relationship to the screen.

Directly calibrating a dense correspondence map between content and each projector (via gray coding or similar methods) often leads to a noisy set of maps. Warping models H are often used to reduce noise. That is, use a set of raw calibration points to parameterize a warping model from which a dense set of point correspondences can be sampled. Noisy values that do not conform to the assumed model can be ignored. By removing noisy values a the sampled set of correspondences will be more accurate, but this only holds if there is no global bias present in the raw values. Note that the assumed model must be able to fit the display surface, otherwise the sampled values will be meaningless. A projective model will work for any flat surface and requires a minimum of four points [17]. Spline based models [17] will work of any continuous surface with sufficient control point density. The number of control points required increases with degrees of local freedom in the display geometry. Display surfaces with edges or discontinuities must be segmented into continuous segments where either a projective model [17] or spline model can be applied; another option is to not use a model. A projective model is defined by

$$\begin{bmatrix} x^{S_j} \\ y^{S_j} \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_3 & a_4 & a_5 \\ a_6 & a_7 & 1 \end{bmatrix} \begin{bmatrix} x^{S_i} \\ y^{S_i} \\ 1 \end{bmatrix} \quad (2.5)$$

where a_0 to a_7 are parameters that define the transformation. The 1's at the bottom of each matrix denote the multiplication being performed in homogeneous space [17].

2.2.3 Interpolation

Let the mapping of an image between spaces be defined by

$$I^{S_b} = I^{S_a}(H_{S_b \rightarrow S_a} X^{S_b}) \quad (2.6)$$

That is, a set of points X^{S_b} is mapped to S_a and is used to sample an image I^{S_a} . Images are typically stored in a quantized pixel space. A consequence of this is images only having values at integer coordinates. This is a problem for mapping between spaces when the mapping must accommodate sub-pixel shifts. Such mappings are prevalent in multi-projector configurations. A process called interpolation is to used fill in the gaps between pixels when sub-pixel shifts exist in the space mapping. There are many methods used to approximate the correct value at any sampling location. Interpolation methods including nearest neighbor, bilinear, convolutional bicubic, bicubic spline, lanczos, and sinc function [17]. These are listed in the order of least accurate and fast to accurate and slow.

For this research the bicubic spline function is used as it provides a balance of accuracy and computational speed. Bicubic spline interpolation is defined in equations (2.7) - (2.15)

for some given point (x, y) in an arbitrary space. The following equations define sampling points on a unit square:

$$\tilde{x} = \text{mod}(x, 1) \quad (2.7)$$

$$\tilde{y} = \text{mod}(y, 1) \quad (2.8)$$

where \tilde{x} and \tilde{y} are the decimal components of the sampling point. Only the decimal values of the sampling point are required as interpolation is performed using a unit square. That is, only the distance from neighbouring pixels is required. The four corners of the unit square within the content image are defined by the points

$$\bar{x}_0 = \lfloor x \rfloor \quad (2.9)$$

$$\bar{x}_1 = \bar{x}_0 + 1 \quad (2.10)$$

$$\bar{y}_0 = \lfloor y \rfloor \quad (2.11)$$

$$\bar{y}_1 = \bar{y}_0 + 1 \quad (2.12)$$

where $\lfloor \cdot \rfloor$ is the floored value of a given variable. The sampled value from I at point (x, y) is defined as

$$p(x, y) = [1 \ \tilde{x} \ \tilde{x}^2 \ \tilde{x}^3] A [1 \ \tilde{y} \ \tilde{y}^2 \ \tilde{y}^3]^T \quad (2.13)$$

where $\bar{x}_0, \bar{x}_1, \bar{y}_0, \bar{y}_1, I$ and its derivatives are omitted as parameters of $p(\cdot)$ for brevity. A is calculated as

$$A = B \begin{bmatrix} I_{00} & I_{01} & I_{y00} & I_{y01} \\ I_{10} & I_{11} & I_{y10} & I_{y11} \\ I_{x00} & I_{x01} & I_{xy00} & I_{xy01} \\ I_{x10} & I_{x11} & I_{xy10} & I_{xy11} \end{bmatrix} B^T \quad (2.14)$$

where I_x and I_y are the derivatives of I for the x and y axes respectively. I_{xy} is the second order mixed derivative in the x and y axis. These derivatives are determined by numerically calculating the slope of the four pixels surrounding the sampled point. The first and second subscript numbers of each I are the \bar{x} and \bar{y} coordinates on the unit square, respectively. B is a matrix containing polynomial coefficients [17], and has the values of

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 3 & -2 & -1 \\ 2 & -2 & 1 & 1 \end{bmatrix} \quad (2.15)$$

These values are derived from the definition of bicubic splines.

2.3 Image Comparison

Ideally, the superimposed image should match the content image (the image provided to a system to project). This is rarely the case in a multi-projector configuration since the physical projector pixel-grids do not perfectly overlap. As such, the ideal content image must be warped so the content between projectors are aligned. These transformations can degrade the quality of the original image. If I^C is the ideal projected image in content space then let \hat{I}^C be the realized superimposed image in content space. This can be expressed as

$$\hat{I}^C = \frac{1}{n} \sum_p^n I_p^C \quad (2.16)$$

where I_p^C is the image displayed by projector p mapped to content space C . Equation 2.16 assumes that each projector contributes equally to the superimposed image (i.e., all projectors share the same brightness). Dividing by n normalizes the brightness of the superimposed image to 1.

The quality of the approximation can be determined by comparing I^C to \hat{I}^C . Standard image comparison techniques include Mean Squared Error (MSE) [17] and Structural Similarity (SSIM) [8]. MSE measures the average squared difference between pixels across images. MSE is defined as

$$MSE(I^C, \hat{I}^C) = \frac{1}{N} \sum_i^N \|I^C(i) - \hat{I}^C(i)\|^2 \quad (2.17)$$

where N is the number of pixels in I^C and \hat{I}^C , and $I^C(i)$ and $\hat{I}^C(i)$ are the pixel values of the i^{th} pixel in I^C and \hat{I}^C , respectively. The lower the MSE the closer the two image are. A problem with MSE is that it does not quantify the visual appearance of an images from a human’s perspective. That is, the same MSE value can correspond to many visually different images. Consider an image that has had two separate adjustments applied to it such that they both have the same resulting MSE: a blurred image and an image with a global intensity offset. Visually, the latter image is almost identical to the original image where as the former image appears significantly degraded. This example is shown in Figure 2.3.

Several comparison metrics have been devised to overcome this pitfall in MSE. A commonly used metric is SSIM [8]. SSIM contains three different modalities of comparison: luminance, contrast, and structure. These metrics are designed to reflect different perceptual aspects of the human visual system. The three metrics on a window w of pixels in

images I^C and \hat{I}^C are defined as

$$l(I_w^C, \hat{I}_w^C) = \frac{2\mu_{I_w^C}\mu_{\hat{I}_w^C} + c_1}{\mu_{I_w^C}^2 + \mu_{\hat{I}_w^C}^2 + c_1} \quad (2.18)$$

$$c(I_w^C, \hat{I}_w^C) = \frac{2\sigma_{I_w^C}\sigma_{\hat{I}_w^C} + c_2}{\sigma_{I_w^C}^2 + \sigma_{\hat{I}_w^C}^2 + c_2} \quad (2.19)$$

$$s(I_w^C, \hat{I}_w^C) = \frac{\sigma_{I_w^C\hat{I}_w^C} + c_3}{\sigma_{I_w^C}\sigma_{\hat{I}_w^C} + c_3} \quad (2.20)$$

where $l(\cdot)$, $c(\cdot)$, and $s(\cdot)$ represent luminance, contrast, and structure, respectively. $\mu_{I_w^C}$ and $\mu_{\hat{I}_w^C}$ are the mean pixel values in window w for I^C and \hat{I}^C , respectively, $\sigma_{I_w^C}$ and $\sigma_{\hat{I}_w^C}$ are the sample standard deviations of pixel values in window w for I^C and \hat{I}^C , respectively, $\sigma_{I_w^C\hat{I}_w^C}$ is the covariance of pixel values between I^C and \hat{I}^C , and c_1 , c_2 , and c_3 are small constants used to prevent small denominators. The three metrics combine into the SSIM metric as

$$SSIM(I_w^C, \hat{I}_w^C) = l^\alpha c^\beta s^\gamma \quad (2.21)$$

where α , β , and γ are weights used to determine the importance of luminance, contrast, and structure, respectively. When $c_3 = \frac{c_2}{2}$ and $\alpha = \beta = \gamma = 1$ (a common setting) SSIM on a window of pixels is defined as

$$SSIM(I_w^C, \hat{I}_w^C) = \frac{(2\mu_{I_w^C}\mu_{\hat{I}_w^C} + c_1)(2\sigma_{I_w^C\hat{I}_w^C} + c_2)}{(\mu_{I_w^C}^2 + \mu_{\hat{I}_w^C}^2 + c_1)(\sigma_{I_w^C}^2 + \sigma_{\hat{I}_w^C}^2 + c_2)} \quad (2.22)$$

The SSIM score between two images is the average of all SSIM windows between the two images

$$SSIM(I^C, \hat{I}^C) = \frac{1}{N} \sum_w SSIM(I_w^C, \hat{I}_w^C) \quad (2.23)$$

where N is the number of local windows in each of the compared images. Typically, a sliding window approach is used when comparing images using SSIM. Note that the number of pixels and number of windows are the same when the window only moves over one pixel at a time. SSIM ranges from -1 to 1, where -1 represents a large dissimilarity between I_w^C and \hat{I}_w^C and 1 is when the two images are identical.



Figure 2.3: This figure compares MSE and SSIM in their ability to measure human visual perceptual differences between images. Two separate transforms are applied to the left most image (a): a Gaussian blur (b), and a global offset (c). Notice that the two transformed images have the same MSE but different SSIM. Comparatively, (c) is perceptually closer to the original image than (b). This is not reflected in MSE but is reflected in SSIM.

2.4 Other Aspects of Multi-Projector Configurations

Multi-projector configurations require many image processing operations to align and normalize (share the same properties) the content that each projector displays. So far only projector placement calibration, content warping, and interpolation have been discussed. However, there are other properties that must be considered to achieve improved image quality results in practice. Such topics are not required to understand the contributions in this work but serve to provide useful context for real world applications and various difficulties that must be overcome to achieve a superior viewing experience. Four areas of particular importance include projector colour response correction, projector edge blending and edge masking, lens Point Spread Function (PSF) [17] correction, and screen property correction. A generic pipeline for a multi-projector configuration is shown in Figure 2.4.

Projectors use a wide variety of methods to display light in the visible and non-visible spectrums. Such methods include: digital light processing (DLP) [18], liquid crystal display (LCD) [19], and light emitting diodes (LEDs) [20]. Each type of system has unique properties that often result in different colour responses. Even projectors that share colour technology can produce drastically different colour responses. For example, projecting a pure RGB *red* to two different projectors can result in two different reds on the same dis-

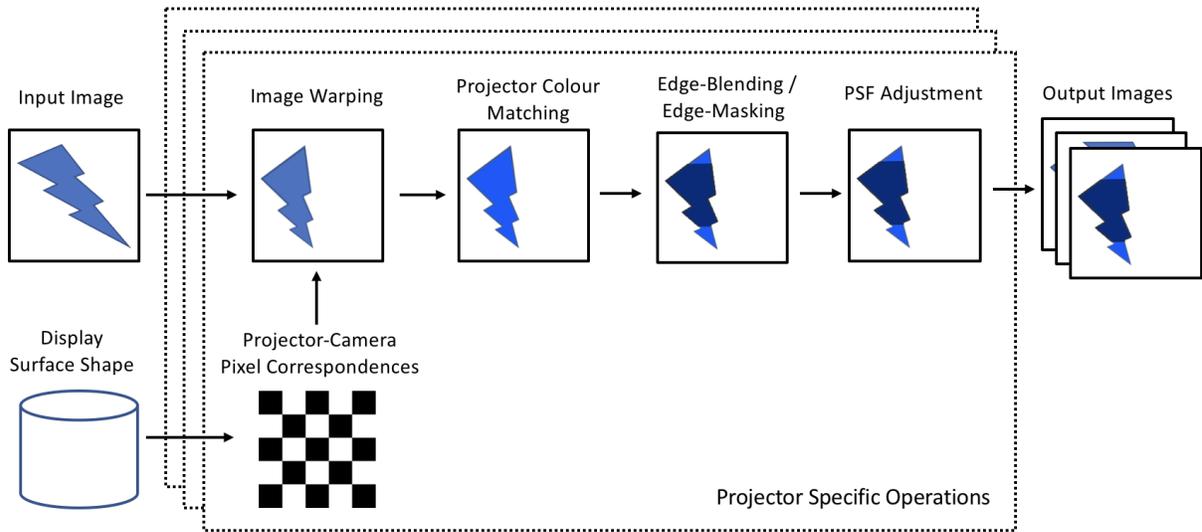


Figure 2.4: A generic pipeline for multi-projector configurations. A calibration step is launched after each projector in the configuration has been placed. Each projector’s position, relative to the projection surface, is determined through the use of structured light patterns. This allows a warping model between content space and projector space to be fit. These spacial models are combined in order to determine a boundary mask for each projector. In addition, relative projector colour responses are determined. A final calibration step involves determining the Point Spread Function (PSF) [17] of each lens in the projector configuration. During run time, for each projector, the input image is warped to the projector space, the image’s colour is corrected to match other projector responses, the overlapping and non-overlapping area brightness levels are harmonized, and the image is sharpened to correct for lens PSF.

play surface (typically one red will look brownish). Correcting for the differences between projectors in the same configuration can result in improved quality of the superimposed image.

There are many different types of multi-projector configurations with respect to overlapping projector patterns. The three general types are discussed in Section 2.1: non-overlapping, partially overlapping, and mostly overlapping. The non-overlapping regions in the overlapping configurations will have less luminance than the overlapping regions; this is because light is additive. Differences in luminance is not desirable for many applications. For example, planetariums are particularly dark as the projected stars cover only a small portion of the display being projected on. The overlap in back light bleeding becomes increasingly visible in this scenario. The brightness of the non-overlapping areas must be increased so that of the overlapping areas to hide the boundaries between the two types of regions (this results in a loss of contrast).

Lenses in projectors are rarely perfect and the light from the projectors may not hit the display surface at a consistent focus. These circumstances can result in a gradual blur across the content being displayed. Methods exist to compensate for this scenario by measuring the characteristics of the blur and pre-processing the content so once projected the blur will be diminished [21, 22, 23].

Many applications require projectors to display content on a non-ideal surface. That is, any surface that is not a white non-textured Lambertian surface [17] will degrade the quality of the displayed image. Several techniques have been developed to compensate for a non-ideal environment [12, 24, 25]. The main idea is to hide the irregularities of the background by pre-adjusting the content so when projected the image adjustments and the display surface irregularities cancel. Typically the average luminance of the projected image is lowered to hide non-reflective areas on the display. Projecting on the side of a building is an example of when the display surface must be compensated for.

Chapter 3

Quantitative Problem Formulation

Multi-projector configurations are built from a series of sub-components as illustrated in Figure 2.4. These systems include, but are not limited to, screen geometry modeling, projector-camera alignment, image warping, inter-projector colour correction, edge-blending, boundary masking, and PSF correction. Each component can add or subtract from the perceptual quality of the realized image on the display surface. This makes it difficult to design a system that optimizes for each component and their inter-dependencies simultaneously. The focus in this work is on the image warping component of multi-projector configurations.

The rest of the chapter is structured as follows Section 3.1 conceptualizes the task of inter-projector content alignment as generic filtering problem. Section 3.2 discusses the implications of overlaying quantized pixel spaces and describes the emergent Moiré patterns. Section 3.3 limits the set of multi-projector configurations under consideration. Finally, Section 3.4 discusses the reasoning for various limitations of the multi-projector configuration model used in the thesis.

3.1 Problem Conceptualization

The image warping component of a multi-projector image processing pipeline (described in Figure 2.4) is responsible for transforming an image I^C from the content space C to all projector spaces $\{P_p\}$. The transformation for moving from C to a specific projector space P_p is described in Equation 2.6. However, representing the process using this notation is limiting as it implies that the transformation is restricted to point-wise mapping and interpolation, and that each content-to-projector transformation is independent. This hobbles

the filtering process from using the increased pixel density, thus removing the potential for a gain in perceptual image quality.

Transforming an image between spaces is inherently a filtering process parameterized by a space transformation $H_{P_p \rightarrow C}$. That is, some set of operations are applied to an image I^C to produce a new image I^{P_p} . A generalized version of Equation 2.2 that allows for inter-projector pixel-grid conditioning within a space transformation can be defined as

$$I^{P_p} = f(I^C, \{H_{P_i \rightarrow C}\}; \theta_p) \quad (3.1)$$

where f is the space transformation filter function, $\{H_{P_i \rightarrow C}\}$ is the set of all space transformation models, and θ_p is a set of parameters used in the filtering process for the p^{th} mapping. For a camera-projector pair p , $H_{P_i=p \rightarrow C}$ is used as the base space transformation model and $\{H_{P_{i \neq p} \rightarrow C}\}$ is used to give f prior knowledge of the entire multi-projector configuration.

The focus of this research is the development of filtering techniques that use the relationship between pixel-grids to increase the perceptual quality of an image formed by superimposed sub-images in multi-projector configurations. The remainder of this chapter describes characteristics of overlapping pixel grids, limits the types of overlapping patterns under consideration, and lists assumptions used to narrow the research scope.

3.2 Moiré Interference Patterns

Understanding the consequences of different projector configurations is key in developing a framework that uses inter-projector pixel-grid relations to increase perceptual image quality. Each projector p in a multi-projector configuration has a different point of projection. This results in a misalignment between each pixel-grid on the display surface. That is, an integer translation between the given projector pair does not exist. No amount of image filtering can overcome this physical limitation. There are three properties of multi-projector configurations that affect the overlapping patterns of pixel-grids: the physical shape of the display surface (e.g., flat, wavy, has corners, etc.), the positioning of each projector, and the location on the display surface in which the content is placed.

The last property makes the placement of the content on the display surface independent of the projectors' physical locations. This distinction is important in a variety of situations. For example, consider a configuration with a single projector. This projector is placed close to the surface to which it is projecting. The projector is angled vertically so a larger image can be achieved. The desired shape of the content is rectangular. However, the vertical

angle of the projector produces a keystone effect. The content being projected must be warped such that the content when projected on the display surface appears rectangular and at the correct ratio. An example of this is shown in Figure 3.1.

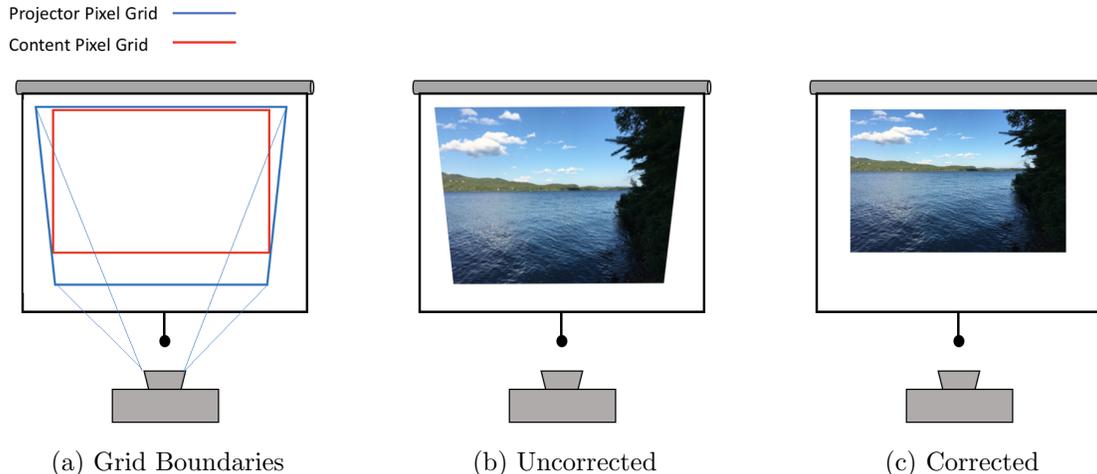


Figure 3.1: Single projector content pixel-grid correction. (a) A comparison between the physical projector pixel grid (blue) with an arbitrarily defined content pixel grid (red). (b) A projected image aligned to the projector pixel grid. (c) A projected image aligned to the content pixel grid.

A key distinction between the three properties (i.e., display shape, projector position, and content location) is that the first two points are physical properties, controlled by display geometry and projector placement, and the last point is a virtual property (digital), controlled by content placement. This requires a system to consider a separate pixel-grid for each projector in the system in addition to the pixel-grid of the content space. Distinguishing content pixel space and projector pixel space is a general requirement in a projection system whenever a projector is not serving as the systems frame-of-reference for content placement; as depicted in Figure 3.1.

A physical constraint to the content-projector relation is the location where the content is placed on the display surface. Another important characteristic of this relation is the pixel density of the content space. It is important to consider the relative density of the projector pixel-grids when deciding on the density of the content space. If the content-to-projector pixel ratio is too high then there is not enough physical projector pixels to sample the content, resulting in a degradation in image quality via image blurring. If the

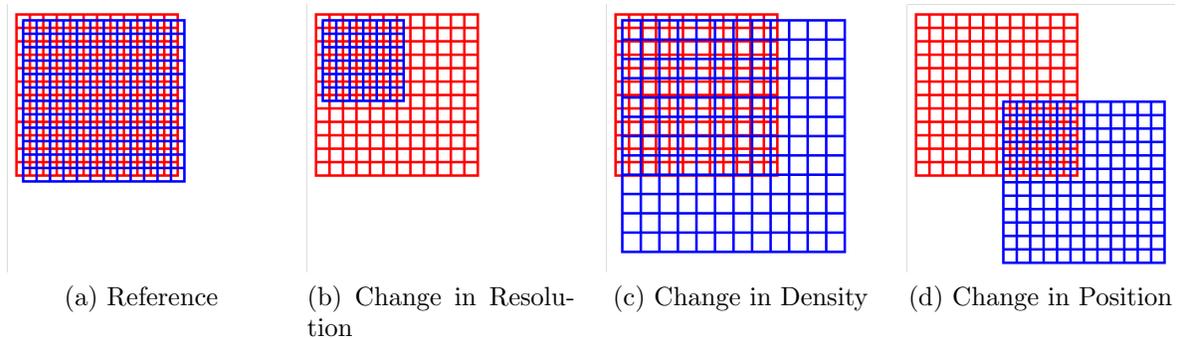


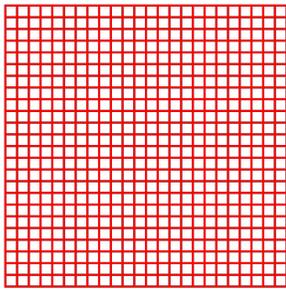
Figure 3.2: Two pixel-grids are shown, one red and one blue. Three different alterations are made to the blue pixel-grid to illustrate the difference between a change in pixel-grid resolution, pixel-grid density, and pixel-grid position.

ratio is too low then the aligned image quality suffers, resulting in a blocky (low resolution) looking image. A comparison between pixel-grid density, resolution, and position is shown in Figure 3.2.

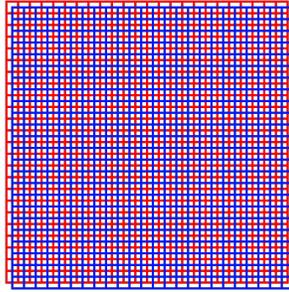
There are two aspects to the overlapping nature of the pixel-grids in a multi-projector configuration: the local overlapping pattern, and the global overlapping pattern. The local pattern is the manner in which pixels between overlapping pixel-grids are arranged. Given any location within an overlapping pixel-grid field, pixels can either completely overlap or only partially overlap. The latter pattern provides the most flexibility for image enhancement as frequencies beyond the Nyquist frequency of a single projector are realizable [4]. Both patterns appear when pixel-grids overlap (except when there is only a global translational shift between pixel-grids, see Figure 3.3b). Examples of various overlapping patterns are shown in Figure 3.3.

3.3 Multi-Projector Configuration Constraints

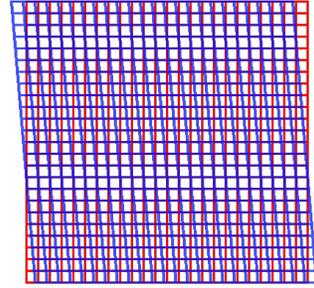
The goal of this research is to increase the perceptual quality of aligned superimposed images when compared to naive image warping techniques. Measuring the effectiveness of a perceptual quality enhancement system is difficult since the image quality loss is both intra-warp dependent and inter-warp dependent. That is, the quality loss of superimposed image is dependent on the loss from a single image warp (between content space and a given projector space), and how the loss between image warps stacks (how projector spaces



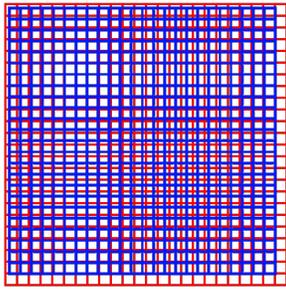
(a) No Warping



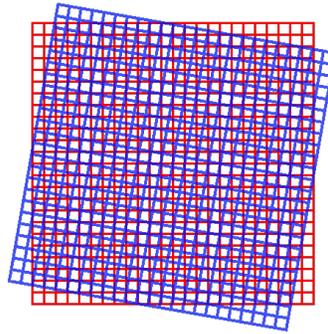
(b) Translation



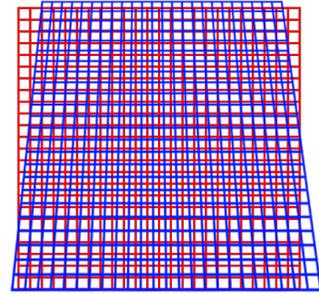
(c) Skew



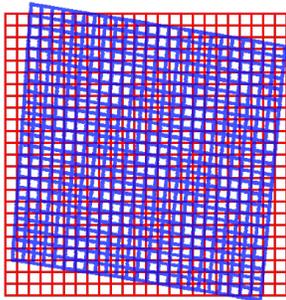
(d) Scale



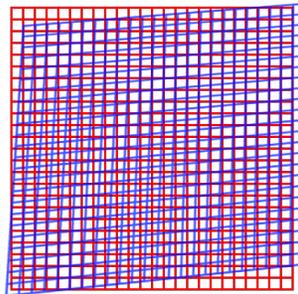
(e) Rotation



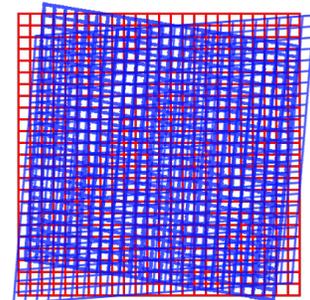
(f) Keystone



(g) Projective 1



(h) Projective 2



(i) Stacked Projective

Figure 3.3: Examples of various Moiré patterns resultant of overlapping pixel-grids. The red pixel-grid is static between examples. The blue pixel-grid has a specified transformation applied to it. Different Moiré patterns emerge as the type of transform applied to the blue pixel-grid is altered. Only the translation transformation has no Moiré pattern. Notice that a Moiré pattern is still present in example (i) while having two transformed pixel-grids superimposed.

relate). Properties that affect intra-warp and inter-warp loss, screen geometry and relative projector configurations, have infinite permutations.

The number of configurations possible given p projectors and an arbitrary display surface is potentially infinite. This makes testing all possibilities computationally intractable and requires restricting the domain of testing to a few variables. One way to restrict the domain of inquiry is to only test flat screens. This has a twofold effect first it reduces the number of possible surfaces to one, obviously, and second it allows for a simplified warping model to be used to map content space to projector space. A projective transform is the upper bound on complexity required to handle a flat surface [17]. This transform is described in Equation 2.5. This can be further restricted to a similarity transform [17], shown in Equation 3.2. Put simply, a projective transform only requires straight lines to remain straight after a transformation, and a similarity transformation requires the stricter condition of keeping angles between lines the same after a transformation.

Given a point (x^{S_i}, y^{S_i}) in 2D space S_i , a mapping to space S_j using a similarity warp is defined by

$$\begin{bmatrix} x^{S_j} \\ y^{S_j} \\ 1 \end{bmatrix} = \begin{bmatrix} s * \cos(\theta) & -s * \sin(\theta) & t_x \\ s * \sin(\theta) & s * \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^{S_i} \\ y^{S_i} \\ 1 \end{bmatrix} \quad (3.2)$$

where (x^{S_j}, y^{S_j}) is the corresponding 2D coordinate in space S_j , s and θ are the scale and rotation parameters, respectively, relating space S_i with space S_j , and t_x and t_y are the translation parameters. Note that this is in homogeneous coordinates, see Section 2.2.2.

Using a similarity model over a projective model reduces the search space from eight variables to four variables. Despite the reduction in search space not much is lost in terms of the transformation’s effect on the possible relation between content space pixels and projector space pixels. Remember that this research is interested in multi-projector configurations where there is substantial overlap between projectors with similar qualities (e.g., aspect ratio, and number of pixels); reflecting the circumstances in many practical applications. This requirement is only met, under the flat screen requirement, when the projectors are placed close to each other with similar axes of projection. The primary characteristic lost in the transformation simplification is the key-stoning effect. This effect under the limitations put forth (flat screen, with large projector overlap) is assumed have a minuscule effect on the measured perceptual quality of the superimposed image in any local region. That is, pixels under these conditions will be approximately square in any local region throughout the projector field. This should hold true for a wide variety of configurations with modern high definition projectors, and where each projector’s axis of projection does not approach a 90 degrees offset from the normal axis of the display surface.

The similarity warp defined in equation (3.2) has two parameters describing the translation within the transformation, one for the x-axis and one for the y-axis. The translation effect of the pixel overlapping pattern can be divided into two components: the integer component, and the decimal component. The integer component shifts the global pattern by a given number of pixels; the phase of the pixel overlap pattern stays in the same location relative to the content in the transformed space. The decimal component of translation shifts the phase of the pixel overlap pattern. Thus, neither the integer component nor the decimal component of the translation is important in describing the global pixel overlapping pattern since the average global overlapping pattern remains relatively constant. The only time this is not true is when there is no rotation and a scale value of 1; like in the case of wobulation [1]. In this case the decimal component of the translation has a global effect on the pixel overlap pattern. Both the translation components have been set to 0 for this research experiment since such a scenario is rare in practice. This limits the transformation model to two parameters, a scale difference, and a rotational difference. Using this limited similarity warping model allows for a more tractable and understandable method of measuring the effects of different content-to-projector and projector-to-projector relationships. Going forward the scale parameter of the limited similarity transform is referred to as the pixel-length-ratio. The pixel-length-ratio describes the length of a projector pixel relative to the length of a content pixel.

3.4 Non-Warping Assumptions

Various assumptions and restrictions must be placed on the other components, shown in Figure 2.4, in a multi-projector configuration to focus solely on image warping. Some of the assumptions have already been discussed but are repeated in this section for completeness. The display geometry is assumed to be flat with a white Lambertian [26] surface. Using a flat display surface allows a simple image warping model to be used resulting in a more tractable approach of testing different projector configurations. The Lambertian surface removes the need to quantify the superimposed image quality from multiple viewing perspectives. Ideal projector-camera pixel correspondences are assumed. This removes alignment uncertainty from all downstream stages. Lens and colour responses of each project are assume to be ideal. Issues of focus can be ignored with an ideal lens. An ideal colour response for each projector allows the research to ignore variations in lighting technology between projectors (i.e., the light source in each projector). For example, projectors cannot produce an impulse response at any given colour (frequency). In addition, projectors cannot typically project nothing (black), a small amount of light usually persists. The

non-illuminated spaces between pixels are ignored. The non-overlapping projector regions will not be considered for this research. This removes the need to match light levels of the non-overlapping regions. The superimposed image is assumed to be the average colour of each overlapping pixel in a given region (this is an implication of the above restrictions but worth pointing out).

The assumptions made for this research seem extensive but for the most part these assumptions do not remove major sources of perceptual quality degradation for a large portion of practical applications. For example, flat screens are one of the primary display surfaces used in practise, there exist projector-camera calibration methods that are sub-pixel accurate, using an application appropriate projector lens with a bore sight (a prism between the light source and the lens within a projector) allows many focusing problems to be remedied, and projectors used within a multi-projector configuration are usually of the same make and model which minimizes inter-projector colour variations. Assumptions that are potential problems include display surface properties vary and are rarely Lambertian, no projector has an ideal response at every (or any) frequency, discarding non-overlapping projector pixels is not done for every application.

Chapter 4

Sub-Pixel Integration

Ideally, the projected superimposed image \hat{I}^C will match the content image I^C . However, there are several transformation operations between each projected image I^{P_p} and the content image I^C . Each transformation potentially affects the content of the image; this makes it difficult to preserve the quality of edges within the transformed image. More importantly, the physical pixel-grids of the content space will not align perfectly with the pixel-grids of the projectors in most multi-projector configurations. This is a problem for standard image comparison metrics as they require a one-to-one mapping between pixels across images.

The following equation represents \hat{I}^C as an average of all the projected images.

$$\hat{I}^C = \frac{1}{n} \sum_p^n I_p^C \quad (4.1)$$

This is the same equation as Equation 2.16. Note that \hat{I}^C is never represented in any digital manner during the projection process. \hat{I}^C only *exists* once all the projectors are displaying the appropriate warped aligned images. This makes it difficult to compare a digital image I^C with \hat{I}^C as \hat{I}^C is only represented in the physical world. Thus, \hat{I}^C must be simulated to compare the two images, or an image must be taken using a camera.

Images are typically thought of in terms of quantized pixel units over a finite space. This representation is not suitable for \hat{I}^C since we want to simulate an accurate representation of the superimposed image on the display surface. More precisely, we want to simulate the overlap of unaligned pixel-grids where the individual pixels each have a physical height and width. Standard definitions of image comparison metrics (i.e. SSIM [8]), that typically

operate on pixels, must be generalized to accommodate comparing a *normal* pixel grid to a superimposed unaligned pixel grid.

The remainder of this chapter is structured as follows, Section 4.1 explores a method of representing an image formed from a set of unaligned pixel-grids, and then Section 4.2 develops a method of modifying a standard image comparison technique to accommodate the new method of image representation.

4.1 Sub-Pixel Calculation

Comparing the similarity of two unaligned pixel grids is not a straightforward task. Standard measures, such as SSIM [8] or MSE [17], require a one-to-one pixel mapping. In practice such a mapping is rarely present due to physical restrictions of overlaying multiple projectors. This makes using the aforementioned loss metrics impossible without some additional steps. One such approach would be to take a picture of the projected images, warp the captured image back to the content space, and then compare. This approach has several issues with it: first, having the projectors and a camera in the measurement loop introduces several sources of noise. Also, this method requires the use of an additional transformation which would further reduce the quality of the comparison.

Another option is to calculate each of the projected images, warp them back to the original content space, and average them to approximate what the superimposed image would look like. This method removes both the projector and camera from the loop. However, this method still uses an inverse warping to move back to the image content space. The loss introduced by the inverse transformation can be reduced by increasing the pixel densities of the original image by upscaling the projected images before applying the inverse transformation. The greater the upscaling the smaller the loss from the inverse transformation. Unfortunately, the computational requirements grow with the size of the upscaling. If this approach is taken to the limit the loss from the inverse transformation reduces to zero. This is equivalent to calculating the individual areas of sub-pixels formed from unaligned overlapping pixel-grids. A depiction of two grids overlapping is shown in Figure 4.1. This figure shows the results of averaging two checkerboard patterns and counts the sub-pixels formed by the overlap.

The sub-pixels that form \hat{I}^C , Equation 4.1, have several properties that affect the perception of \hat{I}^C including: shape, position, and area. All of these properties are affected by the number of projectors in the configuration, and the relative position of each projector with the content space. The shape of each sub-pixel will be a simple polygon assuming

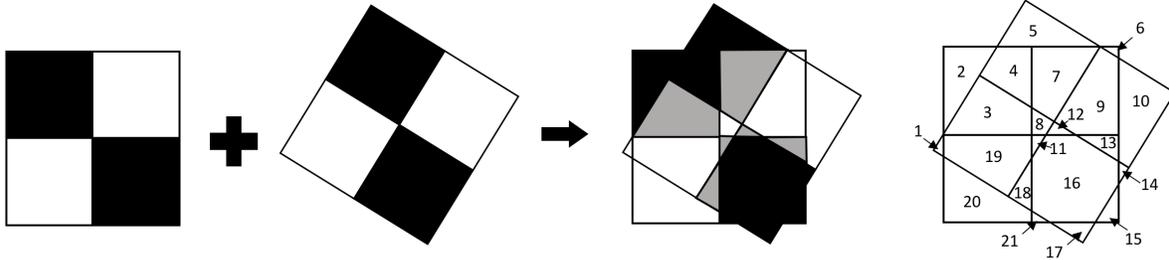


Figure 4.1: The general process of overlapping two separate pixel grids. Two grids with the same checkerboard pattern are offset by a small rotation and translation. Sub-pixels are formed when the two grids are overlaid. When a black pixel from one grid overlaps with a white pixel from the other grid a grey sub-pixel is created. The right most image counts the number of sub-pixels created by the overlap of the two pixel grids.

there are no discontinuities with respect to the projection field of a single projector (i.e., there is a continuous path between all pixels on the display surface for every projector). This allows the shoelace algorithm [27] to be used to calculate the area of every sub-pixel efficiently. Note that this requires pixel edges to be straight lines. This condition is upheld under a similarity transform.

For the purpose of this research the process of representing an image \hat{I}^C using sub-pixels is known as Sub-Pixel Integration (SPI). The following general steps are performed for SPI:

1. Move each projector pixel grid to content space.
2. Overlay all pixel grids. Each simple polygon formed is considered a sub-pixel.
3. Calculate the area of each sub-pixel.
4. Use the set of sub-pixels to represent \hat{I}^C .

After SPI, the shape, the position, and the area of each sub-pixel are known in content space. Note the area of each sub-pixel is with respect to the area of a content pixel. For example, if a sub-pixel has an area of 0.25 then it is 0.25 times the size of a content pixel regardless of the sub-pixel shape.

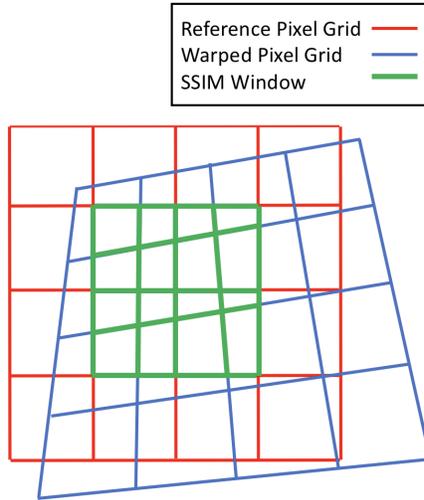


Figure 4.2: This figure shows an example of using SPI-SSIM to compare two unaligned pixel-grids. The original content space is shown in red and the warped content space is shown in blue. The green window is the region under comparison. In this case the SSIM [8] window has a size of 2. Each red pixel in the SSIM window is divided into four sub-pixel sections due the nature of the blue warped grid. The area of each sub-pixel is represented by α_{ab} in Equations 4.5 to 4.7.

4.2 Loss Function Adjustment

After SPI, Section 4.1, is performed on \hat{I}^C , both I^C and \hat{I}^C have a numerical representation in the content space and can be directly compared. The position property of each sub-pixel in \hat{I}^C is used to determine to which pixel in I^C it should be compared. An important consideration is the size of the sub-pixel. That is, there can be a large size difference between sub-pixels. The larger the sub-pixel the more effect it is assumed to have on the perceptual quality of \hat{I}^C . Thus, the effect of a sub-pixel comparison will be weighed by its area. Note that no sub-pixel that overlaps with a content pixel can have an area larger than 1 content pixel squared. The final property extracted during SPI is the shape of each sub-pixel. Humans have a varying sensitivity to lines at different orientations [28]. However, no method of integrating the orientation of each sub-pixel into standard loss metrics is explored in this work.

The standard image comparison metric SSIM [8] operates on a window of pixels and uses mean, variance, and covariance statistics. To apply SPI to SSIM each statistic involving

\hat{I}^C must be modified. Let SPI-SSIM denote that SSIM operates under the SPI constraint

$$SPI-SSIM(I_w^C, \hat{I}_w^C) = \frac{(2\mu_{I_w^C}\mu_{\hat{I}_w^C} + c_1)(2\sigma_{I_w^C\hat{I}_w^C} + c_2)}{(\mu_{I_w^C}^2 + \mu_{\hat{I}_w^C}^2 + c_1)(\sigma_{I_w^C}^2 + \sigma_{\hat{I}_w^C}^2 + c_2)} \quad (4.2)$$

here the terms α , β , and γ are all set to one, and the $c_3 = \frac{c_2}{2}$. Note that Equation 4.2 is the same as Equation 2.22. This results in the same general expression shown in Equation 2.22. The mean value of window w in I^C is

$$\mu_{I_w^C} = \frac{1}{W} \sum_a^W I_{wa}^C \quad (4.3)$$

where W is the number of pixels in window w relative to the original image, and I_{wa}^C is the a^{th} pixel value in window w . The sample variance of I_w^C is:

$$\sigma_{I_w^C}^2 = \frac{1}{W-1} \sum_a^W (\mu_{I_w^C} - I_{wa}^C)^2 \quad (4.4)$$

The mean of a window w of sub-pixels in \hat{I}_w^C is defined as

$$\mu_{\hat{I}_w^C} = \frac{1}{W} \sum_a^W \sum_b^{\|\hat{I}_{wa}^C\|} \alpha_{ab} \hat{I}_{wab}^C \quad (4.5)$$

where α_{ab} is the area of sub-pixel w_{ab} , and $\|\hat{I}_{wa}^C\|$ is the number of sub-pixels in pixel w_a . Note that $\sum_b^{\|\hat{I}_{wa}^C\|} \alpha_{ab} = 1$. The sample variance of \hat{I}_w^C is defined as

$$\sigma_{\hat{I}_w^C}^2 = \frac{1}{W-1} \sum_a^W \sum_b^{\|\hat{I}_{wa}^C\|} \alpha_{ab} (\mu_{\hat{I}_w^C} - \hat{I}_{wab}^C)^2 \quad (4.6)$$

and the sample covariance between the two images about w is defined as

$$\sigma_{I_w^C \hat{I}_w^C} = \frac{1}{W-1} \sum_a^W (\mu_{I_w^C} - I_{wa}^C) \sum_b^{\|\hat{I}_{wa}^C\|} \alpha_{ab} (\mu_{\hat{I}_w^C} - \hat{I}_{wab}^C) \quad (4.7)$$

The size of the SSIM window matters because it determines the balance between local edges and global structure. With resolution enhancement we are interested in preserving

content at the sub-pixel level. Focusing only on the smallest detail may introduce other artifacts that are undesirable. Having a SSIM window that is too large will focus too much on global similarity and adds additional computational costs. It is not inherently clear what size of SSIM to use. Subjectively a SSIM window size of 7 works well and will be used for the remainder of this work. Figure 4.2 shows an example of SPI-SSIM being applied across two unaligned pixel grids.

Chapter 5

Quantitative Properties of Multi-Projector Configurations

SPI-SSIM, Section 4.2, provides a way of quantitatively measuring the perceptual quality of a superimposed image produced by a multi-projector configuration. The only prior information SPI, Section 4.1, requires is the location of each projector relative to the content space. In this chapter SPI-SSIM is used for a variety of tests to determine how different relative projector placement affects the perceptual quality of the superimposed projected image. First, Section 5.1 uses SPI-SSIM to evaluate a single projector and a two projector configuration across a range of pixel-length-ratio and rotation configurations. Section 5.2 expands the two projector analysis by comparing the rotation between projector spaces with the average rotation between the projector spaces and the content space. Finally, Section 5.3 investigates pixel-length-ratio effects on SPI-SSIM on a wider range of values.

5.1 SPI-SSIM Pixel-Length-Ratio vs. Rotation

Establishing a baseline for SPI-SSIM is important before developing models that will attempt to improve upon it. In this section SPI-SSIM is tested on a single projector configuration and a two projector configuration. Remember that even in a single projector case the projector space and content space may not be the same, see Figure 3.1.

The limited similarity warp is used for both configurations. In the one projector configuration the rotation is varied from 0 degrees to 45 degrees. The warp will be centered on the center of the original image. This range covers all possible pixel overlap patterns

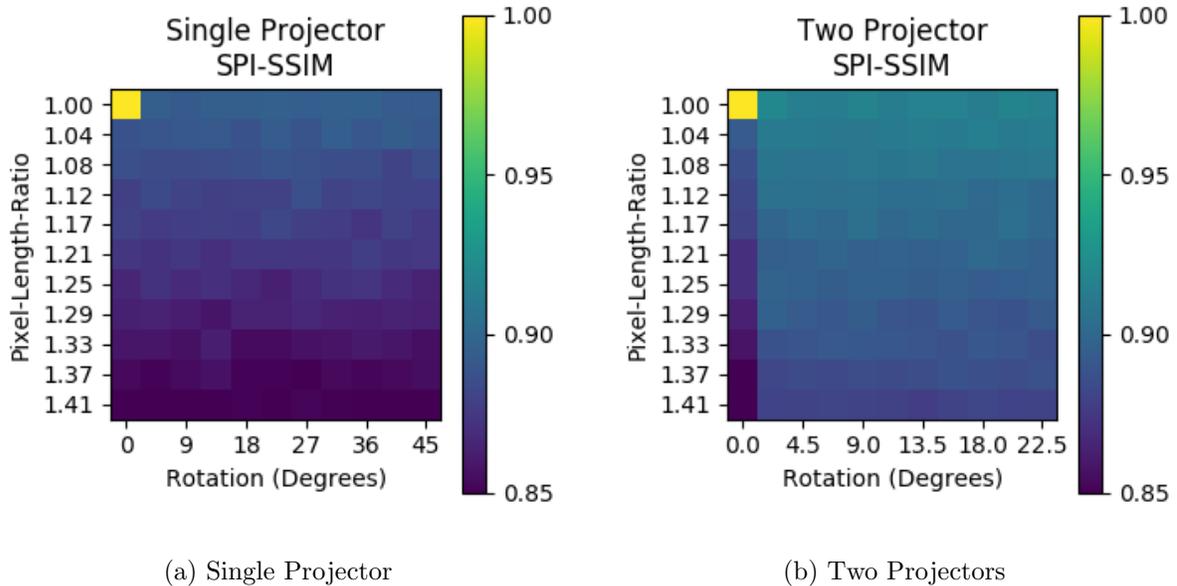


Figure 5.1: This figure shows the average quality loss incurred across a set of images after a similarity transform is applied. The loss is measured using SPI-SSIM. The vertical axis notes the pixel-length-ratio difference between the original content and the warped content. The horizontal axis notes the rotation between the sets of pixel grids. The optimal SPI-SSIM is shown in the top-left corner when an identity warp (no warp) is applied. The loss increases as scale increases and is relatively consistent across rotations.

for the similarity warp at a give scale. The pixel-length-ratio will be varied from 1.0 to $\sqrt{2}$. Remember that pixel-length-ratio is equivalent to the scale parameter in the limited similarity transform model.

In the two projector configuration the rotation for each projector is varied from 0 to 22.5 degrees; one projector will be rotated clockwise and the other projector will be rotated counter-clockwise. This is done to keep the effects of rotation on the content image the same for each projector. These ranges test all possible overlapping pixel patterns between the projector pixel-grids but do not test all the overlapping patterns between each projector pixel-grid and the content pixel-grid. The same pixel-length-ratio range that is used in the one projector test is used for the two projector configuration test.

The average SPI-SSIM is measured across a set of 1000 images randomly selected from the ImageNet dataset [29]. Areas where content is lost due to the warp will not be included

in the average. Note that natural images have specific statistical qualities to them. Namely that a large percentage of natural image content consists of low frequency patterns. The perceptual change to such patterns, under the limited similarity warp, will be minor and will not be reflected in the SPI-SSIM. This makes the amount of SPI-SSIM persevered hard to interpret.

Figure 5.1a shows the results of the single projector configuration test. Both the pixel-length-ratio and the rotation effect of the quality of the image. In general, SPI-SSIM decreases as the pixel-length-ratio increases and is consistent across rotations. Changing rotations only has an effect when the rotation is small. The loss measured at a pixel-length-ratio of $\sqrt{2}$ and a rotation of 0 degrees (bottom-left corner) is less than the loss at a pixel-length-ratio of 1 and a rotation of 45 degrees (top-right corner). The steepest change in SPI-SSIM is when either the pixel-length-ratio or rotation is changed from identity.

Figure 5.1b shows the results of the two projector configuration test. Much like the single configuration test, a change in pixel-length-ratio causes a greater change to the perceptual quality of superimposed image then a change in rotation. A rotation of zero degrees produces a noticeable loss in SPI-SSIM compared to all other rotations. At this rotation both projector pixel-grids are always on top of one another which causes pixels between projectors to average with only one other pixel. Thus, no sub-pixels are formed and the resolution of the superimposed image is effectively diminished.

Overall the two projector configuration test has a higher average SPI-SSIM across the tested parameter ranges. However, directly comparing a specific pixel-length-ratio and rotation setting between tests is not particularly useful because the addition of a projector has a larger effect on the characteristics of the pixel-grid overlapping pattern.

Furthermore, notice that the SPI-SSIM does not drop below 0.85, on average. The limited use of SPI-SSIM’s range, under a similarity transform, will make any improved SPI-SSIM hard to interpret as the maximum loss is not known. Determining maximum loss would allow for the measured loss to be normalized, making SPI-SSIM more interpolatable. Figure 5.2 compares superimposed images with each individual projected image that forms them.

5.2 Inter-Rotation vs. Intra-Rotation

A primary objective in any multi-projector configuration is to minimize the difference between the projected content and the original content. Results from Subsection 5.1 show that relative pixel-length-ratio and rotation have a large effect on the preserved image

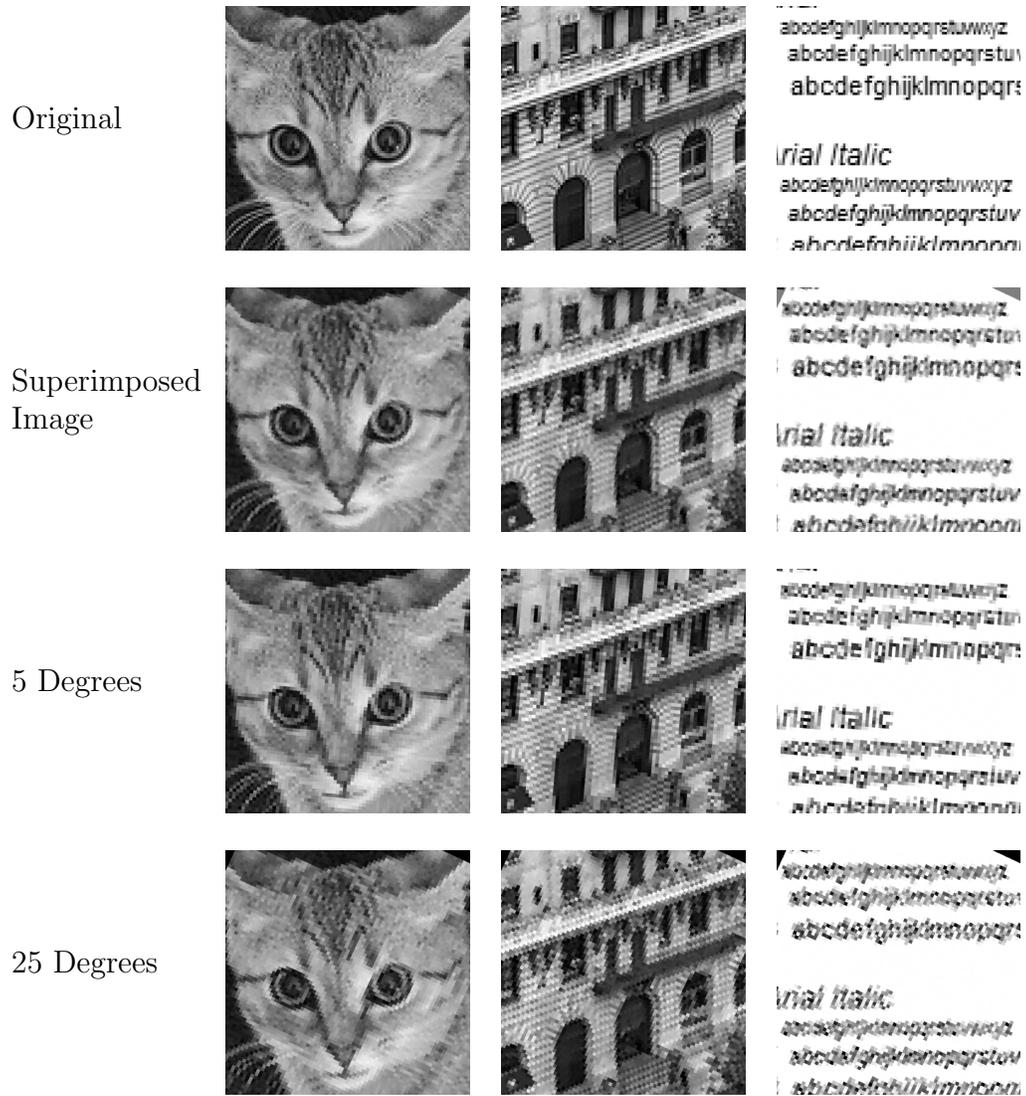


Figure 5.2: A comparison between the original images, the superimposed images, and the individual projected images that when stacked form the superimposed image. The PLR of the projected images is $\sqrt{2}$. Two different rotations are used for demonstrative purposes. Note that the superimposed image and each projected image have their brightness normalized to 1.0 so the content is visible. In reality the superimposed image would be twice as bright as each individual projected image.

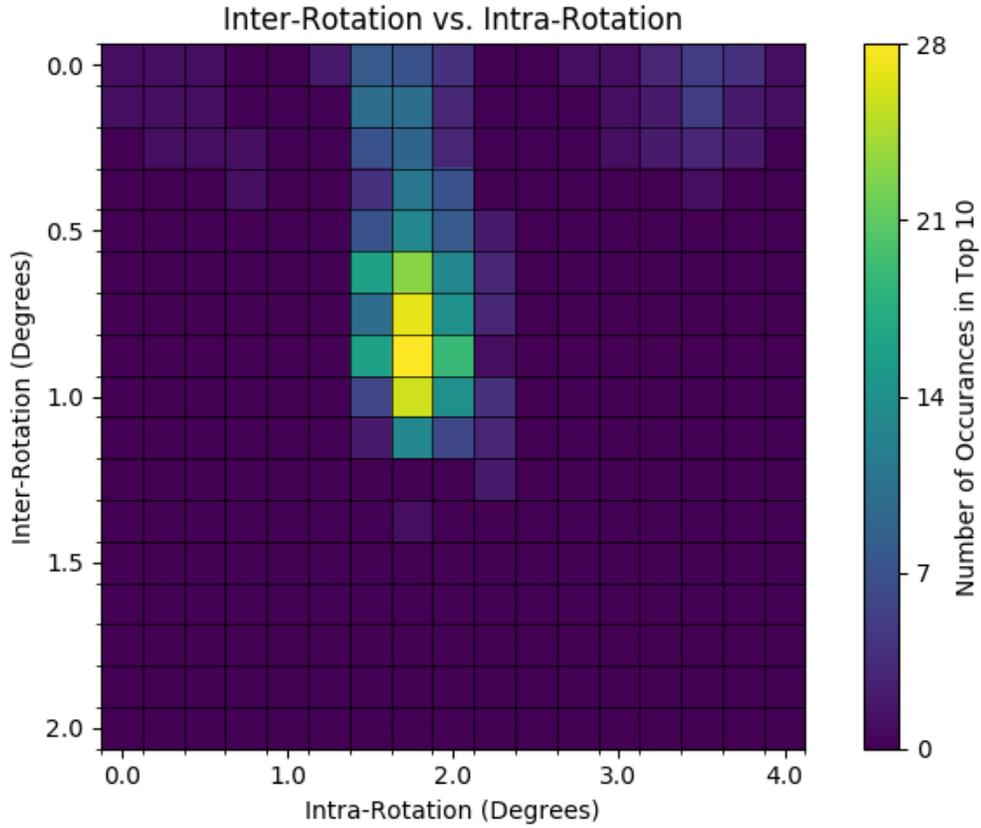


Figure 5.3: This figure shows the results of the inter-rotation vs. intra-rotation test. The rotation between two projectors (intra-rotation) and the average rotation between the projectors and the content space (inter-rotation) are varied across a range of scales. The resulting images is compared to the original image using SPI-SSIM. The top ten location combinations per scale are scored in the histogram. The greatest cluster of maximal SPI-SSIM is focused around an intra-rotation of 1.75 degrees and an inter-rotation of 0.875 degrees.

quality. The smooth gradient in preserved image quality in the test implies some optimal orientation to align projectors. The purposed SPI-SSIM metric, Section 4.2, allows for an automated approach in determining optimal projector placement.

The purpose of this test is to demonstrate that an optimal alignment exists but not how to efficiently find it. Brute force checking all possible alignments for an arbitrary

number of projectors is computationally expensive in general but is tractable given a two projector setup under the defined limited similarity transform. Two parameters will be varied to find a near-optimal configuration in the two projector configuration across a range of pixel-length-ratios. The two parameters are intra-rotation and inter-rotation. The intra-rotation defines the rotation between the two projectors, the inter-rotation defines the average rotation between the projectors and the content. The test from Section 5.1 only varied the intra-rotation; the inter-rotation is kept constant at zero. This could be limiting the proposed system’s capabilities as both projectors suffer the same type of loss at the same time.

Section 5.1 shows the a near-optimal configuration will be found when both the inter-rotation and intra-rotation are low. The inter-rotation is varied from 0 degrees to 2 degree in increments of 0.125 degrees, the intra-rotation is varied from 0 degrees to 4 degrees in increments of 0.25 degrees, and the pixel-length-ratio is varied from 1.0 to $\sqrt{2}$ in increments of approximately 0.01. A total of $17 * 17 * 41 = 11849$ combinations are tried. The results of this test are shown in Figure 5.3.

The greatest clustering of maximal SPI-SSIM is focused around an intra-rotation of 1.75 degrees and an inter-rotation of 0.875 degrees. This centroid is located along the top-left to bottom-right diagonal. This diagonal marks when one of the two projector pixel grids has the same rotation as the original content pixel grid. The majority of the maximal SPI-SSIM locations lie about an intra-rotation of 1.75 degrees and above 1 degree inter-rotation. Why this is occurring will require further investigation but it is likely associated with the relationship between the morié pattern created by the unaligned projector pixel-grids and statistical qualities of natural images. Figure 5.4 compares original images with two different two projector configurations, the optimal configuration and a non-optimal configuration.

5.3 Pixel-Length-Ratio Exploration

The previous tests have limited the pixel-length-ratio range from 1 to $\sqrt{2}$. At a ratio of 1 the pixel length of the content and each projector are equal. In other words, each projector space has the same number of pixels as the content space. At a ratio of $\sqrt{2}$ there are, approximately, the same number of pixels between the two projector spaces as there is in the content space. This counts the projector pixels that overlap with the content pixels.

The goal of this test is to determine how a two projector configuration behaves when pushed beyond a pixel-length-ratio of $\sqrt{2}$. In this test a two projector configuration is



Figure 5.4: A comparison between an original image, a two projector system at the optimal configuration, and a two projector configuration at a non-optimal configuration. Both configurations have a PLR of $\sqrt{2}$. The non-optimal configuration has one projector rotated 22.5 degrees clockwise and the other projector rotated 22.5 degrees counter-clockwise. Notice that the sub-pixels in the non-optimal images do not resemble squares.

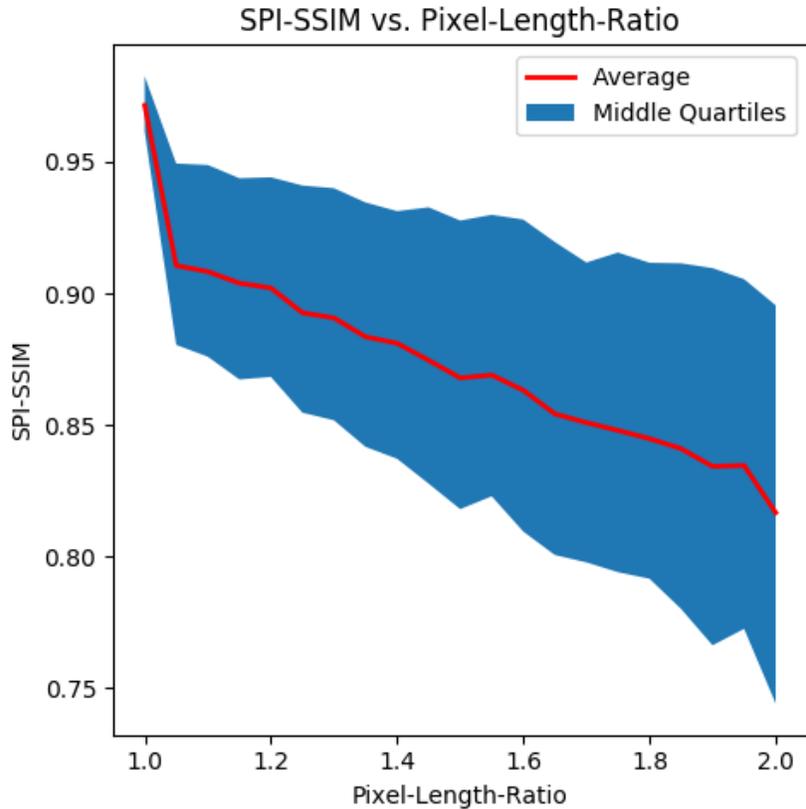


Figure 5.5: Comparing SPI-SSIM to an increase in the pixel-length-ratio of a two projector configuration. Generally, the SPI-SSIM decreases as the pixel-length-ratio increases. There is a major decline in performance after a pixel-length-ratio of 1. The decline in performance is approximately linear. The middle quartile range approximately doubles from a pixel-length-ratio of 1.05 and a pixel-length-ratio of 2.0.

tested at the optimal configuration determined in Section 5.2; an intra-rotation of 1.75 and an inter-rotation of 0.875. The pixel-length-ratio goes from 0 to 2 in increments of 0.05. At 2 the content space has approximately 2 times the number of pixels as there are between the projectors combined. Like the other tests, SPI-SSIM is used as the evaluation metric on a 1000 random images sampled from the ImageNet dataset.

The results of this test are shown in Figure 5.5. A pixel-length-ratio of 1 has the highest achieved SPI-SSIM. This reflects the findings in the previous tests. At this setting there are twice as many pixels on the display as required for the given content. A perfect SPI-SSIM

is not achieved because one of the projector pixel-grids does not align with the content pixel-grid. A steep decline in performance results when a pixel-length-ratio of 1.05 is tested. The decline in performance continues linearly at a rate of approximately 0.04 SPI-SSIM per pixel-length-ratio. The linear decline in perceived quality after a pixel-length-ratio of $\sqrt{2}$ is unexpected as there are no longer enough addressable locations to contain the original content. Calculating effective pixel density in the overlapping projector regions may aid in understanding the relationship between SPI-SSIM and pixel-length-ratio. The middle quartile bounds also follow a linear pattern. The quartile range is the smallest at a pixel-length ratio of 1.0. The range grows significantly at a pixel-length-ratio of 1.05. The quartile range approximately doubles from a pixel-length-ratio of 1.05 to a pixel-length-ratio of 2.0. The range increase is likely cause by the pixel-length-ratio having a non-linear effect across a range of higher frequency signals, but this requires further investigation. Figure 5.6 shows examples of simulated superimposed images at different pixel-length-ratios.

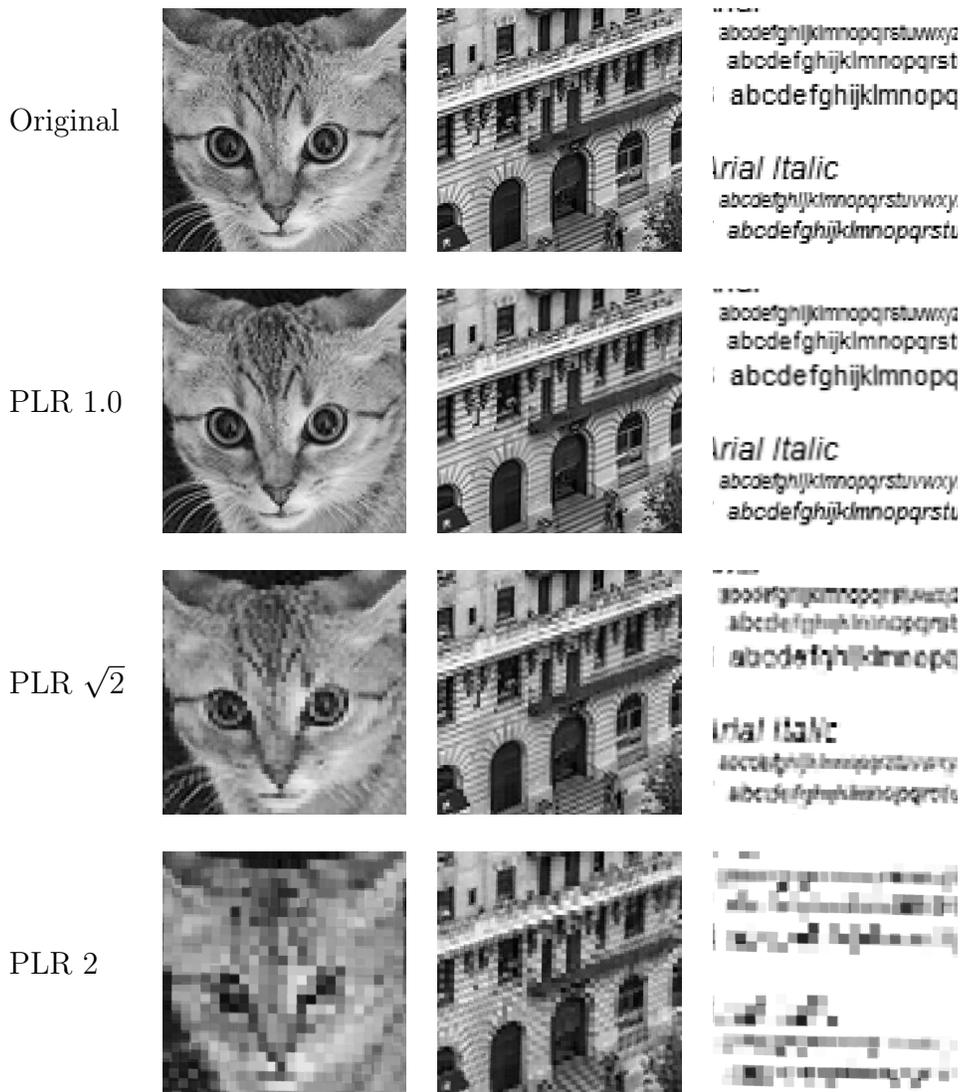


Figure 5.6: A comparison of three simulated superimposed images at different pixel-length-ratios. The images are simulated using a two projector configuration with an intra-rotation of 1.75 and an inter-rotation of 0.875 . From left to right, there is a picture of a cat, a building, and some letters. Notice that the text (a small structured pattern) becomes unreadable before a pixel-length-ratio of $\sqrt{2}$. By a pixel-length-ratio of 2 the text is unrecognizable. The pictures of the animals remain distinguishable across all three pixel-length-ratios. If you look closely at one of the simulated images you will see two sets of overlapping pixel-grids and the emergent Moiré pattern.

Chapter 6

Content Improvement Models

The tests in Chapter 5 show that SPI-SSIM, Section 4.2, is capable of quantitatively measuring the perceptual quality of superimposed images produced by a multi-projector configuration in a variety of configurations. Any difference between the content pixel-grid and any of the projector pixel-grids can result in a non-optimal SPI-SSIM (i.e., a value of 1.0). This is true, in general, if there is any non-DC signal in the content image. This is a result of interpolation and the pixels having a discrete size.

Standard interpolation functions used during space transformations, such as bilinear [17] and bicubic [17], typically operate with knowledge of only the spaces involved, the original space and the mapped space. This is an issue for multi-projector configurations as pixels between projectors have significant overlap. The type of inter-projector pixel overlapping pattern range from constructive interference (a pixel in one space perfectly overlaps with only one other pixel per projector space) to destructive interference (a pixel in one space overlaps with many other pixels in other spaces); examples of these patterns are shown in Figure 3.3.

The destructive interference patterns causes blurring when standard interpolation functions are applied. This is because interpolation *assumes* that this is the optimal approach of signal reconstruction. Building knowledge of other pixel-grids into the transformation process has shown to result in improved superimposed image quality, as discussed in Section 1.2. However, none of these systems use a quantitative method to directly optimize a transformation process for any generic multi-projector configuration. The rest of this section explores three different methods of conditioning each content-to-projector transformation on every other projector in a given multi-projector configuration to improve the perceptual quality of the superimposed image. The system designed in this section is

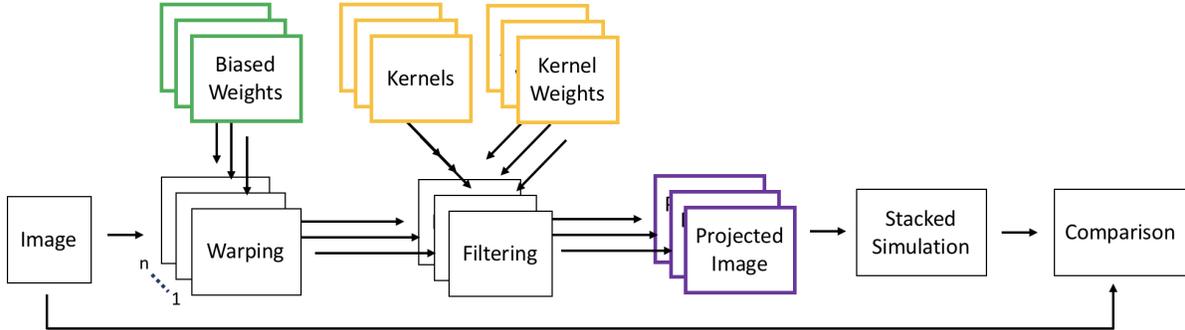


Figure 6.1: The general structure of the proposed model. There are four main sections to the model, warping the original image I^C to the projector space, filtering each warped image, stacking the projected images so comparison may take place, and comparing the superimposed image with the original image. There are three sets of parameters that are learned in the model: the biased interpolation weights (green), used from biased interpolation, the kernel weights (orange), used both the kernel-space and kernel-limited model configurations, and the projected images (purple), used for the optimal model. Note that the projected images are only parametrized for the optimal model and are otherwise the output of the filtering stage.

shown in Figure 6.1.

Transforming between spaces is essentially a filtering process; an output image is produced by applying a set of operations to an input image. Operations applied on the input image can be divided into content specific operations, and content independent operations. For this research, content specific operations implies the filtering operations, for a given spatial region of the content space, have potentially different behaviour for different content. For example, the desired operations for low frequency content and operations for high frequency content may be different. Content independent operations are where filtering operations are static regardless of the content.

The remainder of this chapter is structured as follows, Section 6.1 introduces a filtering method that is integrated directly into the interpolation process between space. Section 6.2 explores a model that applies a post interpolation filter operation. Section 6.3 proposes a model that directly optimizes the projected images. Finally, Section 6.4 analyzes the proposed models.

6.1 Biased Interpolation

Information about high frequency patterns can be lost during interpolation. Previous systems use filtering to minimize the loss introduced and to maximize the the quality of the superimposed image. Determining the number of filters to use and where to use them is a complex problem due to the non-uniformity in pixel overlapping patterns. Past research has not explored integrating filtering into the interpolation process directly. Doing so may allow a greater flexibility for perceptual enhancement as the learned filters will be able to directly leverage the sampling curves. In a sense, more information survives the transformation process.

One type of traditional spatial filtering uses convolutional kernels to enhance content at any given location. Kernel based filtering adjusts a given value based on values about a given location. Interpolation algorithms approximate an unknown value given surrounding values. Thus, any filtering based interpolation method must be able to incorporate both the approximation and adjustment properties. A model is proposed in which an interpolation algorithm can shift along the interpolation curve (i.e., the approximated values between point) by a learned amount.

Common interpolation functions include: nearest neighbor, bilinear, convolutional bicubic, bicubic spline, lanczos, and sinc function [17]. For this research the bicubic spline function is used. The standard bicubic spline function is defined in Equations (2.7) - (2.15). For this research a slight modification is made to the definition of \tilde{x} and \tilde{y} original defined in Equations 2.7 and 2.8, respectively. \tilde{x} and \tilde{y} become

$$\tilde{x} = \text{mod}(x, 1) + \omega_x \tag{6.1}$$

$$\tilde{y} = \text{mod}(y, 1) + \omega_y \tag{6.2}$$

$$|\omega_x|, |\omega_y| \leq \delta_\omega \tag{6.3}$$

where ω_x and ω_y are sampling offsets for the x and y axes respectively. ω_x and ω_y are limited by δ_ω to prevent sampling occurring beyond neighboring pixels. This bicubic interpolation function is the same as the standard bicubic interpolation function when ω_x and ω_y are set to zero. Note that ω_x and ω_y only change where sampling occurs and has no effect on the interpolation curve itself. This method of altering the bicubic spline function can be used in conjunction with most other interpolation functions.

Every pixel in each projector has an ω_x and an ω_y associated with it. Each set of omegas can have different values. These values will ideally allow any configuration of stacked pixel-grids to achieve improved perceptual quality by shifting the sampling point

along the interpolation curve. These weights are dependent on the nature of the local overlapping pattern.

6.2 Kernel Models

Kernel based models learn a set of linear convolutional kernels that can be applied post interpolation to improve the perceptual quality of \hat{I}^C , Equation 4.1. There are many different ways to parameterize and learn kernel based models. The two types of kernel models explored are distinguished by the number of kernels that each model is *allowed* to use. The first model learns as many kernels as needed to achieve an optimal result. The second model is limited to a predetermined number of kernels.

6.2.1 Kernel-Space Model

The Kernel-Space (KS) model has the freedom to learn a unique linear kernel for each pixel in every projector. This is achieved by two sets of parameters per projector: a spanning set of kernels and a kernel activation map. The spanning set of kernels is defined by $K^{P_p} = \{k_{s1}^{P_p}, \dots, k_{sq}^{P_p}\}$ consisting of q kernels where q is the dimensionality of the kernels beings used. s indicates kernels used for the Kernel-Space model. For this model assume each kernel has the same dimensionality and that the kernels are square. Each kernel has a real number associated activation map m_s . Each activation map has a value for every pixel in I^{P_i} . The Kernel-Space filtered image $I_s^{P_p}$ is calculated as

$$I_s^{P_p} = \sum_j^q m_{sj}^{P_p} (I^{P_p} * k_{sj}^{P_p}) \quad (6.4)$$

There are two different methods to parametrize the spanning set of kernels used for this model: use a fixed one hot initialization, or have a tunable set of kernels. The one hot approach ensures that the spanning set is orthonormal and that the only trainable parameters are the kernel activation maps. This effectively makes the kernel activation maps the kernels used on a given pixel. The second approach allows both the kernels and the kernel activation maps to be trainable. Both techniques are described below. Both methods allow the same degree of freedom for learning. The former approach is computationally cheaper to train since the kernels used on any pixel do not need to be sampled from kernel space. The latter approach removes independence between the kernels during training since every pixel effects the delta change in the spanning set.

6.2.2 Kernel-Limited Model

The second model under consideration is the Kernel-Limited (KL) model. This model learns a finite set of kernels $K_l^{P_p} = \{k_{l1}^{P_p}, \dots, k_{ld}^{P_p}\}$ consisting of d kernels for each projector P_p . l indicates kernels used for the Kernel-Limited model. A kernel activation map m_l is learned for each kernel in every projector. For the Kernel-Limited model only a single kernel will be used per pixel. The output of a single projector is defined as

$$I_l^{P_p} = \sum_j^d m_{lj}^{P_p} (I^{P_p} * k_{lj}^{P_p}) \quad (6.5)$$

m_{lj} has the same dimensionality as I^{P_p} and every value in m_{lj} is either 0 or 1 (i.e., off or on). Since only one kernel can be active at time

$$\sum_j m_j = \mathbf{1} \quad (6.6)$$

There are two distinct methods to learn $K_l^{P_p}$, directly train the kernels using gradient descent [30], or use kernels learned for the kernel-space model as a starting point for a clustering based approach. The gradient descent approach requires a relaxation of only one kernel being active (on) at a time so that the system is completely differentiable. Every value of m_l will be allowed to have a continuous value in the range $[0, 1]$. A KullbackLeibler divergence [31,32] based regularizer will be used to promote a sparse distribution for kernel activations for a given pixel in I^{P_p} . During inference the kernel with the highest activation for a given pixel will be used.

The clustering based approach makes use of pre-trained kernel activation weights learned for the Kernel-Space model. The kernel activation maps for a specific pixel in a projector represents the coordinate of the kernel, in kernel space as defined by the kernel spanning set, being used at the pixel’s location. Assuming similar kernels (kernels that produce similar effects such as directional sharpening or blurring) have similar coordinates in kernel space, clustering may be used to directly reduce the number of kernels.

The key to using a limited number of kernels is the fact that any Moiré pattern present in the pixel overlap between projectors is not locally unique. That is, the characteristics of any potential perceptual improvement is dependent on the Moiré pattern and that the local Moiré pattern is repeating throughout a configuration.

6.3 Optimal Model

Both the Biased Interpolation model and the Kernel models are both parameteric and content independent. That is, they try to model a set of operations that will enhance the superimposed image \hat{I}^C using a set of pre-learned parameters. Biased Interpolation learns a set of offsets used during the interpolation process, and the Kernel models learn a set of filters that are applied post interpolation. The ability of both of these models to improve \hat{I}^C 's perceptual quality is restricted by the limited ability of the operations in their respective models.

A method that does not have this limitation is one that does not try fit a set of operations. This can be achieved by optimizing the projected images of each projector directly. In effect, the limitation on the perceptual quality improvement is moved from the model parameterization to the comparison metric being used (namely SPI-SSIM, Section 4.2). The only thing limiting this model is the ability of the comparison metric being used. This model is call the Optimal model and is formally defined as

$$\hat{I}^C = \underset{\hat{I}^C}{\operatorname{argmax}} \operatorname{SPI-SSIM}(I^C, \hat{I}^C) \quad (6.7)$$

In this equation \hat{I}^C , and more importantly its sub-images $\{\hat{I}_p^C\}$, is a flexible set of variables rather than a fixed set of images. There are two options to initialize each sub-image in $\{\hat{I}_p^C\}$ using a random initialization, or using the naive sub-image as initialization (i.e., the image that would be produced by each projector in a naive mutli-projector configuration). For this research the second approach is used. Note that the pixels of each sub-image $\{\hat{I}_p^C\}$ are directly trained in this model.

6.4 Results

The three models proposed are the Biased Interpolation model, Section 6.1, Kernel models, Section 6.2, the Optimal Model, Section 6.3. First two tests are performed on Kernel models, then all models are compared.

6.4.1 Kernel-Space Spanning Set

This test compares the two discussed methods of parameterizing the Kernel-Space model's kernel spanning set, Section 6.2.1. The first method uses a fixed Euclidean spanning set,

and the second method learns the spanning set that maximizes SPI-SSIM, Section 4.2. Both methods are tested using a two projector configuration on a range of pixel-length-ratios and intra-rotations. In this test the inter-rotation is set to zero. To reiterate, the purpose of this test is to determine which of the two parameterization methods is superior for improving SPI-SSIM, and not to measure the effectiveness of the Kernel-Space at improving SPI-SSIM.

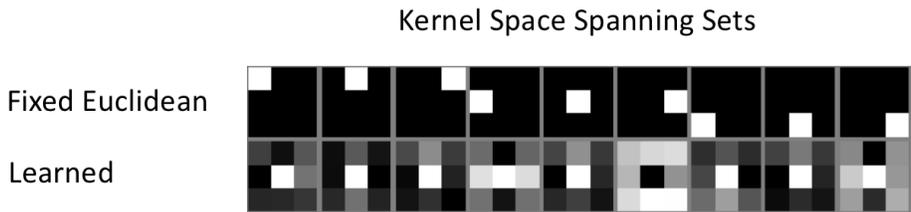


Figure 6.2: This figure shows two sets of kernel spanning sets. The top row of 3x3 kernels is the spanning set used when the spanning set is fixed. Only one value each of these kernels is set to 1, the rest are set to 0. This spanning set is the natural basis of Euclidean space. The bottom row of 3x3 kernels are an example of a learned spanning set. The values in these kernels may be set to any real number. Note that the values between the spanning sets are not normalized in this figure.

Figure 6.2 visualizes one spanning set for each approach. The learned spanning set approach proved a more effective training approach for training the Kernel-Space model. The model converges quicker and with a greater increase in average SPI-SSIM. This is not surprising as the kernels learned for the free spanning model individually resemble useful sharpening and blurring filters. This most likely allows the model to move along the high dimensional manifold of useful kernels more efficiently. For the remainder of the tests in this paper the learned spanning set of kernels is used.

Notice that the third, fifth, and eighth kernels for the learned spanning set are almost identical. This redundancy indicates that useful linear kernels for perceptual quality enhancement are limited to a plane in kernel space. The complexity of the overlapping pixel-grid patterns increases as more projectors are added. Thus, the redundancy in the kernel spanning set may not hold for configurations with more than two projectors.

6.4.2 Kernel-Space vs. Kernel-Limited

The Kernel-Space model, Section 6.2.1, and the Kernel-Limited model, Section 6.2.2, are very similar. In fact, the Kernel-Space model is equivalent to the Kernel-Limited model when the number of kernels is limited to the number of pixels within the image being filtered. Limiting the number of kernels has two primary advantages: reducing the computational cost of the model, and reducing the memory required to store the kernels.

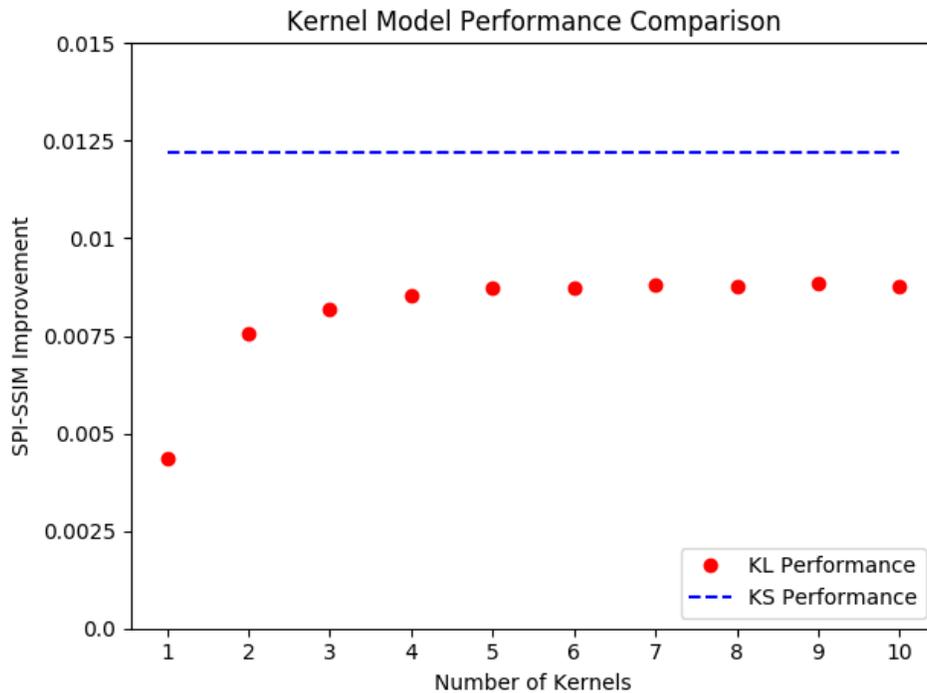


Figure 6.3: This figure shows the performance of the Kernel-Limited model as the number of kernels is increased. The top line indicates the average performance of Kernel-Space model from which the Kernel-Limited model uses as a starting point during clustering. A range of 1 kernel to 10 kernels is shown as performance saturates as the number of kernels increases.

The Kernel-Limited model for this test uses the clustering based approach to generate the set of kernels. Directly learning the kernels through gradient descent proved too unstable to produce reliable results across different initializations and projector configurations. More work is required to determine if direct kernel learning in the Kernel-Limited model

is feasible. For this test the number of clusters is varied between 1 to the number of pixels in the projected images.

Figure 6.3 demonstrates the trade-off between the number of kernels used and the achievable SPI-SSIM improvement. The improvement in SPI-SSIM, Section 4.2, increases as the number of kernels increases with diminishing returns. The improvement saturates around 5 kernels, on average, reaching approximately 75% of the Kernel-Space model’s performance. Achieving 75% with less than one percent of the total number of kernels indicates that there is a limited variety of distortions that must be compensated for (at least in the two projector configuration). The performance of the Kernel-Limited model continues to increase as the number of kernels approaches the number of pixels in each of the projected images.

6.4.3 Model Comparison

This section compares the Biased Interpolation model, Section 6.1, the Kernel-Space model (using the learned spanning set), Section 6.2.1, and the Optimal model, Section 6.3. The first two models are designed to be content independent. Testing the BI model and the KS model in a content specific way will provide an upper limit of performance if these models contained a content dependent mechanisms of adjusting the underlying filtering process. As such, both these models will be trained using a content independent manner (optimize the model across a set of images) and a content dependent manner (optimize the model for every image). The content independent models are referred to as general models and the content dependent models are referred to as particular models.

For the sake of comparison, the Biased Interpolation model is combined with the Kernel-Space model to create a Fusion model. Biased Interpolation allows a model the ability to reduce the quality loss of the interpolation process. The Kernel Space model allows for explicit sharpening or blurring of content within a region to improve the superimposed image quality. Fusing these two different approaches may allow of a more effective system. Figure 6.1 illustrates fusion model.

A gradient ascent [30] approach is taken to optimize the SPI-SSIM, Section 4.2, for each model using a two projector configuration. The Tensorflow [33] package is used for gradient ascent in this work. Momentum is used in conjunction with gradient ascent [34]. Each model is tested using the same pixel-length-ratio and rotation ranges used in Section 5.1 for the two projector configuration. Note that the general models are trained until performance improvement saturates. The average improvement across 1000 images is shown in Figure 6.4. Note that the baseline SPI-SSIM of the dataset across all tested

configurations is 0.897. The numerical improvement, regardless of the model, will be small since the maximum value of SPI-SSIM is 1.0.

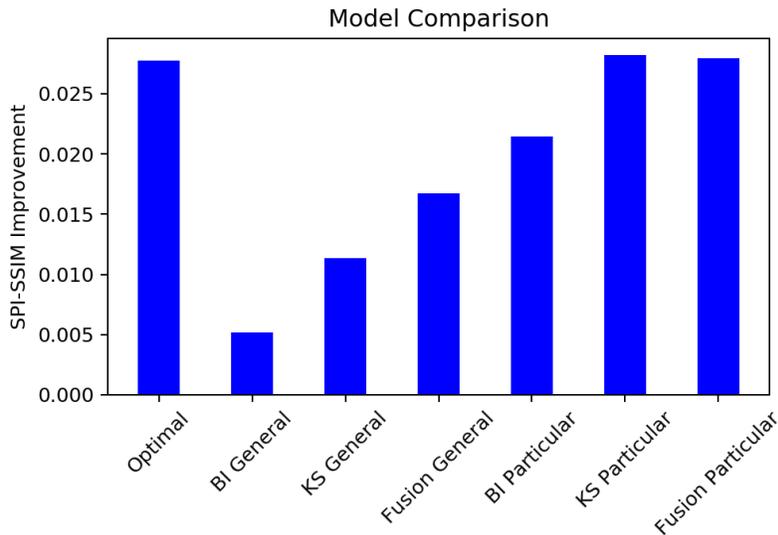


Figure 6.4: This figure shows the relative performance of seven different models. Notice that the Fusion General model is approximately the addition of the BI General model and the KS General model. The KS Particular model and the Fusion Particular model achieve the same improvement as the Optimal model as they are capable of learning unique pixel offsets on a per pixel basis.

The Optimal model, the Kernel-Space particular model, and the Fusion particular model all achieve the same SPI-SSIM improvement. The Kernel-Space particular model learns a unique kernel for every pixel in every image. The learned kernels effectively act as optimal offsets for the pixel being filtered. This allows the Kernel-Space model to match the Optimal model’s performance. The particular Kernel-Space model converges in about 100 iterations per image, on average. The same applies to the Fusion model since it contains the Kernel-Space based filtering.

The next best model is the Biased Interpolation particular model. BI should not be able to achieve the same performance as the Optimal model because it is limited to the sampling curves present in the interpolation performed during space warping. Despite the reduction in relative performance the Biased Interpolation model only takes 10 iterations per image to converge on average. All the particular models are not practical to use as they require more steps to converge and require more computations per step then the Optimal

model. However, they indicate an upper bound to the performance of the general models.

The three general models from worst to best performance are the Biased Interpolation model, the Kernel-Space model, and the Fusion model. The Biased Interpolation model achieves one quarter and the Kernel-Space model achieves a half of the performance when trained across a set of images compared to their respective particular based training. Unlike the particular case, the Fusion model has superior performance to both the Biased Interpolation model and Kernel-Space model. The performance of the Fusion model is approximately the addition of the individual models. The Biased Interpolation model, the Kernel-Space model, and the Optimal model are compared in Figures 6.5 to Figure 6.8. From a visual perspective filtered images produced by the BI general model may not be noticeably different from the original image. The noticeable differences only occur around high contrast areas.

Text - Optimal Two Projector Configuration Example

Original

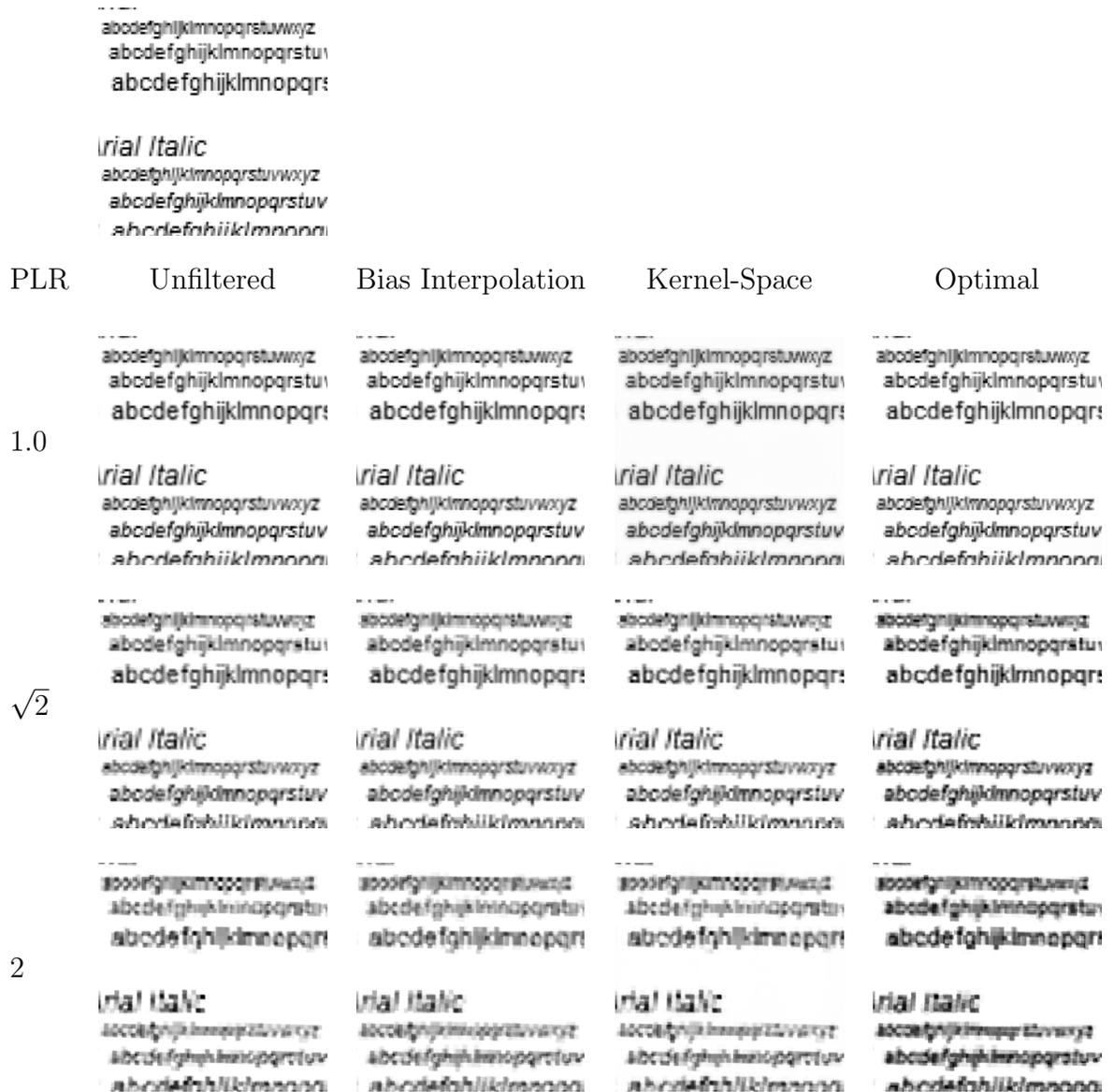


Figure 6.5: An example of the proposed filtering techniques applied to a two projector configuration at the optimal inter-rotation and intra-rotation setting. Each models' ability to improve the quality of the text image increases as the PLR increases. The Optimal model has the largest effect on the quality of the text. Perceptually the Optimal model increases the contrast in the image but numerically it increases the structure in the image. Note that the models have little effect on a PLR of 1.0 since there are twice as many projector pixels as content pixels within each simulated image.

Text - 22.5 Degrees Two Projector Configuration Example



Figure 6.6: An example of the proposed filtering techniques applied to a two projector configuration where the projectors are rotated 22.5 degrees away from the content space. One projector is rotated clockwise and the other projector is rotated counter-clockwise. Note that the black marks in the top corners of each image are caused by each simulated projector not having pixels at those locations due to the applied rotation. The simulated unfiltered images in this example are all blurrier compared to the optimal configuration in Figure 6.5. This enables the filtering models to have a greater impact on improving the perceptual quality.

Building - Optimal Two Projector Configuration Example

Original



PLR

Unfiltered

Bias Interpolation

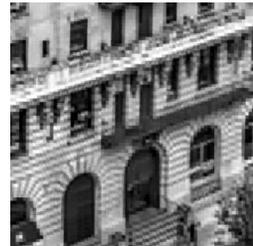
Kernel-Space

Optimal

1.0



$\sqrt{2}$



2



Figure 6.7: An example of the proposed filtering techniques at the optimal two projector configuration. In this example the difference between the original image and the unfiltered images is obscured in comparison to Figure 6.5. This also makes it more difficult to see the improvement provided by each model. Notice that none of the proposed models can compensate for the Moiré effect on the building’s staircase at a PLR of 2.

Building - 22.5 Degree Two Projector Configuration Example



Figure 6.8: An example of the proposed filtering techniques applied to a two projector configuration where the projectors are rotated 22.5 degrees away from the content space. The orientation of the projectors is perceptually difficult to determine (as opposed to Figure 6.6) because of the high frequency content. One of the projector pixel-grids now aligns (more or less) with the grain of the building's staircase causing the Moiré pattern to no longer be present. Like the other examples the Optimal model has the greatest effect on quality of the image; the build's features become sharper with the appearance of an increase in global contrast.

Chapter 7

Conclusions

This chapter summarizes this thesis, reviews the contributions made, discusses the impact of this work, and lists potential areas for future research.

7.1 Summary of Thesis and Contributions

In this thesis, a framework for quantitatively evaluating the perceptual quality of a multi-projector configuration is proposed. This framework is used to evaluate the multi-projector configurations across a wide range of physical setups. Finally, the proposed framework is used to train three different models that improve the perceptual quality of image produced by multi-projector configurations.

Chapter 3 introduces the problems encountered when designing a system to improve the perceptual quality of superimposed images produced by multi-projector configurations. First, the process of aligning content between spaces is conceptualized as a filtering problem, Section 3.1. The emergence of Moiré interference patterns in multi-projector configurations are reviewed and their implications for perceptual quality improvement is discussed, Section 3.2. In addition, the constraints, Section 3.3, and assumptions, Section 3.4, on the type of multi-projector configurations considered in this work are stated.

Chapter 4 proposed a framework to evaluate the perceptual quality of the multi-projector configurations. A technique called Sub-Pixel integration (SPI) is proposed to numerically represent superimposed images produced by multi-projector configurations, Section 4.1. SPI integrates the concept of sub-pixels into the structural similarity (SSIM) [8] metric to produce a new metric called SPI-SSIM, Section 4.2. This new metric

allows a method of quantitatively comparing the ideal projected image with the superimposed image produced by a mutli-projector configuration.

Chapter 5 uses SPI-SSIM to evaluate a variety of projector configurations. First SPI-SSIM is used on a single projector configuration and a two projector configuration to serve as a baseline, Section 5.1. Then SPI-SSIM is used to determine the optimal configuration for naive two projector configurations. It is shown that perceptual quality is most preserved in a two projector configuration when there is a intra-rotation of 1.75 degrees and an inter-rotation of 0.875 degrees, Section 5.2. Finally, a range of pixel-length-ratios are tested on the optimal two projector configuration, Section 5.3.

Chapter 6 uses the proposed framework to train three separate models capable of improving the perceptual quality of the superimposed image. The Bias Interpolation model integrates filtering directly into the interpolation process used during the content-to-projector transformation process, Section 6.1. The Kernel based models learn a set of projector specific linear convolution kernels that are applied after the space transformation process, Section 6.2. Finally, the Optimal model directly learns the best possible combination of sub-images for each projector, Section 6.3.

The models are compared in Section 6.4. The Optimal model is most able to preserve perceptual quality of the superimposed image. This is followed by the Kernel Space model, then the Biased Interpolation model. Fusing the KS model and the BI model showed greater performance then either individually. However, the Fused model is still outperformed by the Optimal model. Using a trainable kernel spanning set for the kernel space in the KS model both improved performance and allowed for a faster convergence. Every additional kernel used in the Kernel Limited model greatly improves the its performance. This trend continues until approximately five kernels where every additional kernel is met with diminishing returns.

7.2 Impact

The novel contributions of this work can be grouped into two categories: a method of comparing a superimposed image with the ideal content, and a framework for training differentiable models in an end-to-end manner using the proposed comparison metric. A process call Sub-Pixel Integration (SPI), Section 4.1, is used to represent a superimposed image as a collection of sub-pixels and allows standard image comparison metrics to quantitatively compare the ideal projected image with the actual projected image. An important part of SPI is that it can be used for any multi-projector configuration. Previous quantitative measures only work on configurations where the projectors differ by translation shifts,

or methods of approximating the superimposed image are used. The framework developed is a general method of using an SPI altered comparison metric to optimize a model that improves the superimposed image quality. This framework only requires that the model under consideration is differentiable. In addition, the proposed framework makes it easier to iterate and improve upon model designs as a model can be quickly trained and evaluated on actual multi-projector configurations.

7.3 Future Work and Preliminary Results

The framework and models introduced in this work are effective tools for measuring and improving upon the perceptual quality of superimposed images. For the most part, the proposed framework and models are assembled using existing techniques. These techniques may not be optimal for measuring or improving upon the perceptual image quality in an unaligned pixel-grids environment. The remainder of this section lists possible directions for future research related to the contributions made in this work; some of the discussed directions also include a discussion on preliminary results. This list is in no way extensive.

7.3.1 Improved Comparison Metric

SSIM [8] was chosen as the base comparison metric because it has been a standard image comparison technique for the last 15 years. As discussed in Section 2.3, SSIM [8] is formed from three separate statistical measure, one for luminance, one for contrast, and one for structure. Each of these measures behaves differently when a limited similarity transform is applied. Luminance has little to no change, contrast is reduced by at most 5%, and structure can be drastically effected, especially for high frequency content. This observation implies, at the very least, that two of the three components of SSIM are not critical to measuring the perceptual degradation of images when a limited similarity warp is applied. A possible area of research would be to verify this pattern on other warping models and to explore other comparison techniques that measure the local structure different between images.

Another observation regarding SSIM is that it is not sensitive to the perceptual quality difference across a range of rotations. Consider two single projector systems, one where there is a 5 degree difference between the content space and projector space, and another system where there is a 45 degree difference between the content space and projector space. SPI-SSIM will quantitatively evaluate the appearance of both systems similarity.

However, in reality the former system appears much closer in quality to the original image, on average. This is because the orientation of the 5 degree system is more capable of preserving both vertical and horizontal frequencies of the original content. This inability to distinguish quality between rotations has only been observed in rotations greater than 5 degrees, approximately. Developing a metric that is sensitive to relative pixel-grids orientation would be a useful area of advancement.

7.3.2 Complex Surfaces

Only flat surfaces were used throughout this work. This is was done to enable a systematic way of testing a variety of projector configurations and allowed a warping model with only two parameters to be used. Limiting the warping to model to only vary scale and rotation was made under an assumption that the key stoning effect would have a minor influence on the perceptual quality of superimposed images. This assumption should be verified. In addition, the proposed comparison framework and models should be tested on more complex surfaces and more complex warping models. Note that SPI, Section 4.1, assumes that a configuration only contains a set of pixel-grids where each pixel-grid in contiguous on the display surface (i.e., neighbouring pixels are beside each other). This condition will not hold for all complex 3D models. Thus, SPI must be generalized to allow for discontinuities between pixels and within a single pixel (i.e., a single projector pixel lays on disconnected regions of the display surface).

7.3.3 Testing More than Two Projectors

The methods proposed in this work are limited to no more than two projectors. This restriction helped in limiting the number of tests required to validate the proposed models. Initial tests suggest that increasing the number of projectors is beneficial but with diminishing returns for each new projector. An important factor in the potential for perceptual improvement lies with the difference between the content pixel density and each projectors pixel density; much like with the two projector configurations.

7.3.4 Parametric Content Dependent Models

The three types of model proposed in this work are either feed-forward and content independent (Biased Interpolation and Kernel models, Section 6.1 and Section 6.2 respectively),

or iterative and content dependent (Optimal model, Section 6.3). The feed-forward models are fast but lacked performance during runtime, and the iterative model is slow but provides superior results. Developing a model that is capable of being feed-forward and content dependent would be beneficial where pre-processing a video sequences is not desirable. Initial attempts at developing such a model proved unfruitful. However, some tests indicate that a two step training process may be beneficial. First generate a dataset of optimal sub-images using the Optimal model. Then train the feed-forward content dependent model using the generated dataset as the targeted output. In a sense, this method of training provides such a model with a *more realistic* goal.

References

- [1] W. Allen and R. Ulichney, “47.4: Invited paper: Wobulation: Doubling the addressed resolution of projection displays,” *SID Symposium Digest of Technical Papers*, vol. 36, no. 1, pp. 1514–1517, 2005.
- [2] B. Sajadi, D. Qoc-Lai, A. H. Ihler, M. Gopi, and A. Majumder, “Image enhancement in projectors via optical pixel shift and overlay,” in *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10, April 2013.
- [3] C. Jaynes and D. Ramakrishnan, “Super-resolution composition in multi-projector displays,” in *IEEE Int’l Workshop on Projector-Camera Systems*, vol. 8, 2003.
- [4] N. Damera-Venkata and N. L. Chang, “On the resolution limits of superimposed projection,” in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 5, pp. V–373, IEEE, 2007.
- [5] S. A. J. Hansen, J. Y. Hardeberg, and M. Nadeem Akram, “Resolution enhancement through superimposition of projected images—how to evaluate the quality?,” *Electronic Imaging*, vol. 2017, no. 12, pp. 141–146, 2017.
- [6] D. G. Aliaga, Y. H. Yeung, A. Law, B. Sajadi, and A. Majumder, “Fast high-resolution appearance editing using superimposed projections,” *ACM Transactions on Graphics (TOG)*, vol. 31, no. 2, p. 13, 2012.
- [7] E. Barshan, M. Lamm, C. Scharfenberger, and P. Fieguth, “35.3: Resolution enhancement based on shifted superposition,” in *SID Symposium Digest of Technical Papers*, vol. 46, pp. 514–517, Wiley Online Library, 2015.
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

- [9] G. Frank, “Pulse code communication,” Mar. 17 1953. US Patent 2,632,058.
- [10] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1, pp. I–I, IEEE, 2003.
- [11] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong, “Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 74–83, 2017.
- [12] J. Salvi, J. Pages, and J. Batlle, “Pattern codification strategies in structured light systems,” *Pattern recognition*, vol. 37, no. 4, pp. 827–849, 2004.
- [13] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [14] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.
- [15] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *European conference on computer vision*, pp. 404–417, Springer, 2006.
- [16] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “Brief: Binary robust independent elementary features,” in *European conference on computer vision*, pp. 778–792, Springer, 2010.
- [17] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [18] L. J. Hornbeck, “Digital light processing for high-brightness high-resolution applications,” in *Projection Displays III*, vol. 3013, pp. 27–41, International Society for Optics and Photonics, 1997.
- [19] J. A. Van Raalte, “Reflective liquid crystal television display,” *Proceedings of the IEEE*, vol. 56, no. 12, pp. 2146–2149, 1968.
- [20] T. J. Brukilacchio and C. A. Demilo, “Light emitting diode projection system,” Nov. 16 2010. US Patent 7,832,878.

- [21] M. S. Brown, P. Song, and T.-J. Cham, “Image pre-conditioning for out-of-focus projector blur,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, pp. 1956–1963, IEEE, 2006.
- [22] Y. Oyamada and H. Saito, “Focal pre-correction of projected image for deblurring screen image,” in *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pp. 1–8, IEEE, 2007.
- [23] S. Ladha, K. Smith-Miles, and S. Chandran, “Projection defocus correction using adaptive kernel sampling and geometric correction in dual-planar environments,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pp. 9–14, IEEE, 2011.
- [24] A. Grundhöfer and D. Iwai, “Robust, error-tolerant photometric projector compensation,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5086–5099, 2015.
- [25] Q. Jia, H. Xu, J. Song, and X. Gao, “Research of color correction algorithm for multi-projector screen based on projector-camera system,” in *Intelligent System Design and Engineering Application (ISDEA), 2012 Second International Conference on*, pp. 1285–1288, IEEE, 2012.
- [26] R. Basri and D. Jacobs, “Lambertian reflectance and linear subspaces,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, pp. 383–390, IEEE, 2001.
- [27] B. Braden, “The surveyor’s area formula,” *The College Mathematics Journal*, vol. 17, no. 4, pp. 326–337, 1986.
- [28] A. Slavutskaya, N. Y. Gerasimenko, and E. Mikhailova, “Mechanisms of orientation sensitivity in the human visual system: Part i. behavioral characteristics of orientation sensitivity. influence of the task type, experimental conditions, and gender,” *Human Physiology*, vol. 40, no. 6, pp. 660–668, 2014.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [30] S. Ruder, “An overview of gradient descent optimization algorithms,” *CoRR*, vol. abs/1609.04747, 2016.
- [31] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

- [32] J. M. Joyce, “Kullback-leibler divergence,” in *International encyclopedia of statistical science*, pp. 720–722, Springer, 2011.
- [33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [34] N. Qian, “On the momentum term in gradient descent learning algorithms,” *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999.
- [35] E. Barshan, M. Lamm, C. Scharfenberger, and P. Fieguth, “35.3: Resolution enhancement based on shifted superposition,” *SID Symposium Digest of Technical Papers*, vol. 46, no. 1, pp. 514–517, 2015.
- [36] A. Said, “Analysis of subframe generation for superimposed images,” in *2006 International Conference on Image Processing*, pp. 401–404, Oct 2006.
- [37] F. W. Campbell and J. G. Robson, “Application of fourier analysis to the visibility of gratings,” *The Journal of Physiology*, vol. 197, no. 3, pp. 551–566, 1968.
- [38] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on Image Processing*, vol. 15, pp. 430–444, Feb 2006.
- [39] T. Okatani, M. Wada, and K. Deguchi, “Study of image quality of superimposed projection using multiple projectors,” *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 424–429, 2009.
- [40] F. L. Bookstein, “Principal warps: Thin-plate splines and the decomposition of deformations,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [41] A. Majumder, C. Browne, and M. S. Brown, *Practical multi-projector display design*. AK Peters/CRC Press, 2007.