

Multi-Projector Content Preservation with Linear Filters

Andrew Hryniowski
Zohreh Azimifar
Mark Lamm
Paul Fieguth
Alexander Wong

University of Waterloo, ON, Canada
Shiraz University, Fars, Iran
Christie Digital Systems Inc., ON, Canada
University of Waterloo, ON, Canada
University of Waterloo, ON, Canada

Abstract

Using aligned overlapping image projectors provides several advantages when compared to a single projector: increased brightness, additional redundancy, and increased pixel density within a region of the screen. Aligning content between projectors is achieved by applying space transformation operations to the desired output. The transformation operations often degrade the quality of the original image due to sampling and quantization. The transformation applied for a given projector is typically done in isolation of all other content-projector transformations. However, it is possible to warp the images with prior knowledge of each other such that they utilize the increase in effective pixel density. This allows for an increase in the perceptual quality of the resulting stacked content. This paper presents a novel method of increasing the perceptual quality within multi-projector configurations. A machine learning approach is used to train a linear filtering based model that conditions the individual projected images on each other.

1 Introduction

Using multiple image projectors to display content on an overlapping area of a screen's surface is a common practice to achieve a larger area of projection, to increase brightness, to allow for projector failure, etc. Aligning the content from each projector is typically done by warping the content to a common space as opposed to physically aligning the pixel-grids of each projector. In practice, it is impossible to align all projector pixel-grids within a configuration because each projector has a different physical point of projection. Overlapping unaligned pixel-grids produces a moiré pattern because some pixels between projectors mostly overlap while other pixels between projectors do not. Standard practice in multi-projector systems is to use a warping model such that the content of each projector is warped between spaces independently [1].

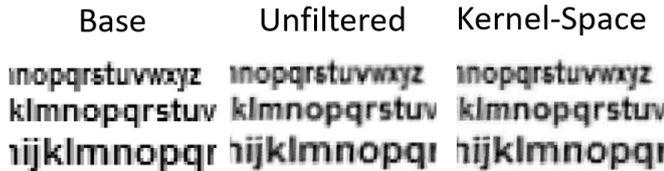


Fig. 1: An example result of the kernel-space model developed in this paper. From left to right, the first image is the ideal projected image, the second image is a simulation of the stacked unfiltered images, and the third image is a simulation of stacked filtered images. Note that improvement does not occur evenly in the kernel-space image. This is a physical property of two projector configurations that cannot be overcome with image processing.

Several approaches have been used to measure the quality loss in a multi-projector configurations. Standard approaches include formulating a theoretic model in 1D space and assume it applies in 2D space [2], taking a high-resolution picture of the stacked image, warping it to content space, and performing a qualitative comparison, or simulating the resulting image in a Wobulation type configuration [3] and using MSE to train the model [4]. None of these methods produce a quantitative number that can be used to optimize a model for any generic multiple-projector configuration. Several approaches have used kernel-based filtering approaches [2, 5, 6, 7, 8, 4], but none directly trained them for a generic multi-projector configuration using human perceptual metrics, such as Structured Similarity (SSIM) [9].

To overcome this problem [10] uses a technique called sub-pixel integration (SPI) so SSIM can accommodate unaligned pixel-grids. SPI calculates the area of each sub-pixel created by over-

lapping multiple pixel-grids. This method calculates the sub-pixel formed by overlapping pixel-grids and biases SSIM based on the various sub-pixel sizes.

The rest of the paper is structured as follows: Section 2 defines the basic notation used to describe the different coordinate spaces used, and the calculation of the stacked image. Section 3 describes the different models developed in this paper. Section 4 describes how SPI is used to modify SSIM. Section 5 discusses the experiments performed. Section 6 draws conclusions and future research directions are suggested.

2 Multi-Projector Configurations

This section describes the various spaces used and how the stacked image is calculated. Let the content space C be the native space of the image I^C being projected. Let every projector space be defined by P_p . Mapping a set of points $X^C = x_1^C, \dots, x_{h*w}^C$ from C to P_p is defined by

$$X^{P_p} = H_{C \rightarrow P_p} X^C \quad (1)$$

where $H_{C \rightarrow P_p}$ models the relationship between the two spaces, and h and w are the height and width of the image I^C , respectively. An image I^{P_p} in projector space P_p is calculated as

$$I^{P_p} = I^C (X^{P_p}) \quad (2)$$

I is quantized in its native space C and must remain quantized in all projector spaces. This requires interpolation for points in X^{P_p} that lie between the sample points of I^C . The realized image seen by a viewer in content space is the normalized sum of the individual projected images

$$\tilde{I}^C = \frac{1}{n} \sum_p^n I^{P_p} (H_{P_p \rightarrow C} X^{P_p}) \quad (3)$$

where \tilde{I}^C is the stacked image, n is the number of projectors, and $H_{P_p \rightarrow C}$ is a mapping from projector space P_p to content space C . Note that n is used to normalize the range of pixel intensity values to a finite range and allows \tilde{I}^C to be directly compared to I^C . Equation 3 can be modified to use a weighted sum approach when stacking projectors to allow each projectors to vary in brightness. To better approximate the stacked image in reality, the mapping used in equation 3 does not treat the set of projector images $\{I^{P_p}\}$ as entities that each have a finite number of sampling points. Instead, the images consist of *physical* pixels that have a height and width, just as projector pixels have a physical height and width. This means that every location within the bounds of \tilde{I}^C has a value, as opposed to taking the value of the nearest integer point. Note that the screen door effect, the non-illuminated space between pixels, is ignored and each pixel is assumed to be ideal and that the pixels are perfect squares with a uniformly illuminated surface.

3 Kernel-Space Model

This section describes two different models. The first subsection explores a model in which separate kernels are learned for every single pixel in every projector. The second subsection discusses the combination the proposed kernel model with a previously developed approach.

3.1 Kernel-Space Model

Learning a unique linear kernel for every single pixel in each projector requires two sets of parameters per projector: a spanning set of kernels and a kernel activation map. The spanning set of kernels is defined by $K^{P_p} = \{k_1^{P_p}, \dots, k_q^{P_p}\}$ consisting of q kernels where q is

the dimensionality of the kernels beings used. For this research, assume each kernel has the same dimensionality and that the kernels are square. Let $m_j^{P_p}$ be an activation map for projector P_p . Each activation map has q values for every pixel in I^{P_p} representing the activation strength of the q^{th} kernel for a given pixel. Let the Kernel-Space (KS) model be defined by

$$I^{P_p} = \sum_j^q m_j^{P_p} (I^{P_p} * k_j^{P_p}) \quad (4)$$

There are two different methods to parameterize the set of kernels used for this model: use a fixed Euclidean initialization or have a tunable set of kernels. The Euclidean approach ensures that the spanning set is orthonormal and that the only trainable parameters are the kernel activation maps. The Euclidean initialization 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 explored below.

3.2 Fusion

For the sake of comparison, the Biased Interpolation (BI) model developed in [11] is combined with the KS model. BI provides the ability to reduce the quality loss from the interpolation process when transforming an image between spaces. The KS model allows for explicit sharpening or blurring of content within a region to improve the stacked image quality. Fusing the BI model and the KS model will be tested to determine if they compliment one another.

4 Image Comparison

Ideally the original image I^C and its approximation \hat{I}^C should equal each other. In practice, the original image and the approximation is only possible when the content pixel-grid is perfectly aligned with each of the projector pixel-grids. Such a scenario is nearly impossible to achieve in practice with multi-projector configurations. Traditional loss functions in image processing require sample points to be uniform across the two pieces of data being compared. Unlike standard images \hat{I}^C is the average of multiple unaligned pixel-grids and results in a non-uniform set of sub-pixels.

SPI [10] can be applied to one-to-one based loss functions to accommodate for a non-uniform sub-pixel space. SPI divides a given pixel in I^C to match the local sub-pixels shapes found in \hat{I}^C . [10] uses SPI on SSIM [9] and is called SPI-SSIM. SSIM 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. For brevity, only the parts of SSIM that SPI adjusts are listed. The mean of a window r_w of sub-pixels in \hat{I}^C centered on pixel w is defined as

$$\mu_{\hat{I}^C} = \frac{1}{W} \sum_a^W \sum_b^{||\hat{I}_{wa}^C||} \alpha_{ab} \hat{I}_{wab}^C \quad (5)$$

where W is the number of pixels in r_w , α_{ab} is the area of sub-pixel w_{ab} , \hat{I}_{wab}^C is the sub-pixel intensity for sub-pixel w_{ab} , and $||\hat{I}_{wa}^C||$ is the number of sub-pixels in pixel w_a . Note that $\sum_b^{||\hat{I}_{wa}^C||} \alpha_{ab} = 1$ since the total area of a given pixel in content space is equal to 1. The sample variance of r_w is defined as

$$\sigma_{\hat{I}^C}^2 = \frac{1}{W-1} \sum_a^W \sum_b^{||\hat{I}_{wa}^C||} \alpha_{ab} (\mu_{\hat{I}^C} - \hat{I}_{wab}^C)^2 \quad (6)$$

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

$$\sigma_{I^C \hat{I}^C} = \frac{1}{W-1} \sum_a^W (\mu_{I^C} - I_{wa}^C) \sum_b^{||\hat{I}_{wa}^C||} \alpha_{ab} (\mu_{\hat{I}^C} - \hat{I}_{wab}^C) \quad (7)$$

The placement of α_{ab} in Equations 5, 6, 7 biases the mean, variance, and covariance relative to each sub-pixel's area, respectively. Note that the maximum value (i.e., the optimal value) of SPI-SSIM is 1.0. The models in this paper are trained such that

$$\theta = \underset{\theta}{\operatorname{argmax}} \operatorname{SPI-SSIM}(I^C, \hat{I}^C; \theta) \quad (8)$$

where θ is the set of trainable parameters for any given model. Gradient ascent is used to optimize Equation 8.

5 Experiments

This section examines the proposed Kernel-Space model. The first test compares two different methods of parameterizing the set of kernel activation maps $m_j^{P_p}$. The Kernel-Space model is compared to two other models, a Biased Interpolation model [11], and the Fusion model.

The tests below use a two-projector configuration where each projector is offset from the content by a given scale and rotation. For each section below, the scale range $[1.0 - \sqrt{2}]$ and the rotation range $[0.0^\circ - 22.5^\circ]$ are considered. For a given rotation, one projector is rotated clockwise and the second projector is rotated counter-clockwise. The images used for training are sampled from the ImageNet dataset [12].

5.1 Kernel Representation

There are two ways to train the Kernel-Space model: learn the kernels at every pixel directly or let the model learn a spanning set and a coordinate in the spanning set. Technically, the former approach uses the natural basis for Euclidean space. Both methods allow the same degrees of freedom for learning. The former approach is computationally cheaper to train since the kernels 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. Figure 2 visualizes two spanning sets, one for each approach.

The learned spanning set method proved a more effective training approach for training the Kernel-Space model as the learned spanning set method converges quicker and with a greater increase in average SPI-SSIM. The kernels in the learned spanning set individually resemble useful sharpening and blurring filters. This most likely allows the model to move along the high dimensional manifold of sharpening/blurring kernels more efficiently. For the remainder of the tests in this paper the learned spanning set method is used for the KS model.

5.2 Model Comparison

This section compares the Kernel-Space model, the Biased Interpolation model, and a Fusion of the two approaches. Two training approaches are taken for each model: a particular method and a generalized method. The particular method optimizes a given model on a per image basis resulting in a uniquely parameterized model per-image. The generalized training method trains a given model to perform well across a dataset of images. Each model is trained using SPI-SSIM on a dataset of 8000 images. The models are trained until improvement saturates. Note that the average baseline SPI-SSIM across the tested configurations is 0.897. The average improvement produced by each model after training is shown in Figure 3.

The Kernel-Space Particular model, and the Fusion Particular model both achieve the same SPI-SSIM improvement. The Kernel-Space Particular model learns a unique kernel for every given pixel in every given image. The learned kernels effectively act as optimal offsets for the pixel being filtered. The Particular Kernel model converges in about 100 iterations of gradient ascent 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 other particular models because it is limited to the interpolation sampling curves during the space transformation process. Despite the reduction in relative performance the Biased Interpolation model only takes 10 iterations of gradient ascent to converge on average. All the particular models are not well suited for real-time scenarios as each projected frame would have to be optimized prior to projection. This would be a difficult task at high resolutions and high frame rates. 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 BI General model achieves one quarter of the performance compared to the BI Particular model. KS General

Kernel Space Spanning Sets



Fig. 2: 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 for each kernels is set to 1, the rest are set to 0. The bottom row of 3x3 kernels show 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. Notice that there is redundancy in the learned spanning set of kernels. Starting on the left, the third, fifth and eighth kernel resemble one another.

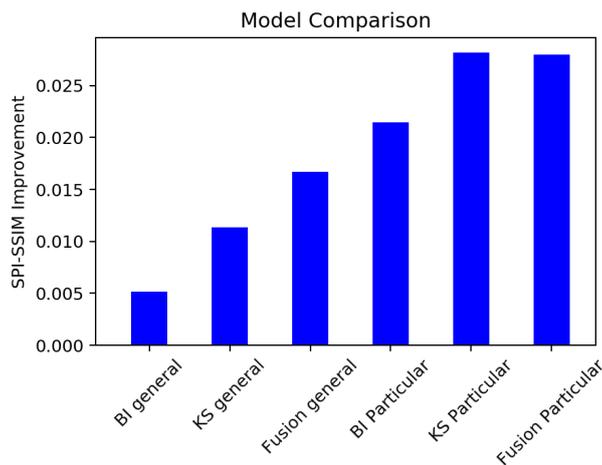


Fig. 3: The relative performance of six different models. Notice that the Fusion General model is approximately the addition of the BI General model and the KS General model. The KS Particular model and the Fusion Particular model achieve the same improvement as they are both capable of learning unique pixel offsets on a per pixel basis.

model achieves a half of the performance when compared to the KS Particular mode. The Fusion General model's performance is approximately the addition of the KS General model and BI General model. This indicated that the Kernel-Space model and the Biased Interpolation model are independent with regards to their filtering abilities.

6 Discussion

A new technique of learning sharpening kernels for multi-projector content enhancement is developed. The proposed Kernel-Space model learns a spanning set of kernels and a kernel activation map indicating the strength of each kernel for a given pixel. Two different methods of parameterizing the Kernel-Space model are tested: using a fixed spanning set, or using a trainable spanning set. The trainable spanning set model converges quicker and is capable of achieving better results for a fixed amount of training. Finally, the Kernel-Space model is compared to two other models using a general training strategy and an image specific strategy. The two other types of models used for comparison are a Biased Interpolation model, and a Fusion model. The Kernel-Space model outperformed the Biased Interpolation model. Combining the two models proved an effective approach at increasing the achieved SPI-SSIM improvement when using the general training strategy.

Acknowledgments

We would like to thank Christie Digital Systems Inc., Natural Sciences and Engineering Research Council (NSERC), Canada Research Chairs program, and Ontario Centres of Excellence (OCE)

for supporting this research.

References

- [1] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [2] 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.
- [3] W. Allen and R. Ulichney, "Wobulation: Doubling the addressed resolution of projection displays," *SID Symposium Digest of Technical Papers*, vol. 36, no. 1, pp. 1514-1517, 2005.
- [4] E. Barshan, M. Lamm, C. Scharfenberger, and P. Fieguth, "Resolution enhancement based on shifted superposition," *SID Symposium Digest of Technical Papers*, vol. 46, no. 1, pp. 514-517, 2015.
- [5] 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.
- [6] C. Jaynes and D. Ramakrishnan, "Super-resolution composition in multi-projector displays," in *IEEE Int'l Workshop on Projector-Camera Systems*, vol. 8, 2003.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] A. Hryniowski, "A content enhancement framework for multi-projector systems," Master's thesis, University of Waterloo, 2018.
- [11] A. Hryniowski, I. Ben Daya, A. Gawish, M. Lamm, A. Wong, and P. Fieguth, "Multi-projector resolution enhancement through biased interpolation," in *Conference on Computer and Robot Vision*, 2018.
- [12] 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.