Dynamic Defocus Deblurring with Depth Data

Eldrick Millares
Stanford University
eldrick@stanford.edu

Abstract

Consumer projectors have narrow depth of focus due to their large apertures optimized for display brightness. Out-of-focus blur thus limits the visual fidelity of fine details when the projection surface is not an optimal, fronto-parallel planar surface at the focal plane. We propose a technique that takes advantage of depth data from commercially available depth cameras to deblur projected images in real-time, making our approach suitable for dynamic scenes. Our approach uses continuously streaming depth data to estimate the blurring point spread function for a projection scene. Using these estimates, we can synthesize a pre-conditioned image using Wiener deconvolution which can preserve the visual fidelity of the original image after projection blur.

1. Introduction

Spatially augmented reality using projector-camera systems offer a unique visual experience for users. This technique enables the visual augmentation of everyday objects with computational techniques, with applications ranging from large-scale graphical art and advertisements to personal experiences such as gaming. Indeed, our project is motivated by the application of projector-camera systems to aid in surgical procedures. [4]

In order to facilitate a wide range of applications for this technique, the standard assumptions of projecting onto fronto-parallel planar surfaces with known or constant reflectivity and geometry must be lifted. Therein lies a plethora of distortions which negatively impact the visual experience of the spatially augmented reality technique. [1] One such distortion effect is out-of-focus deblurring. Consumer projectors have large apertures to optimize for the brightness of their projected image. Consequently, their depth-of-field is quite narrow, leading to significant blurring effects when the projected image does not land precisely at the focal plane.

This work seeks to address the challenges of correcting out-of-focus blur for projector-camera systems. Our approach utilizes continuously streaming depth data from a depth camera to estimate the severity of defocus blur on a projection surface. Using an estimate of the defocus point spread function (PSF), we apply Wiener deconvolution on our original image to pre-condition it before it is projected out onto the scene. Our technique can correct for defocus blurring in real-time, making it suitable for dynamic scenes.

2. Related Work

Previous works on projector deblurring have used coded feature markers to estimate the PSF from a projected image directly, with a variety of methods to produce a preconditioned image. State-of-the-art techniques pioneer the use of deep learning models.

Brown et al. [2] projects feature markers in the shape of crosses and estimates the PSF via the sharpness of the feature markers. They then precondition the image to be projected using a weighted interpolation of pre-computed "basis images". These basis images are a set of k preconditioned images computed with Wiener deconvolution using k distinct PSFs. This technique cannot easily be applied to dynamic scenes due to the need to project the feature markers when the projection surfaces change. Additionally, the need for pre-computed basis images imposes significant computational overhead when the image to project changes, increasing with k.

Zhang and Nayar [5] pursue a similar approach to Brown et al. but utilize an iterative gradient descent to compute the pre-conditioned image. Again, the same limitations for dynamic scenes apply - the need to re-project feature markers and the computational overhead of a descent algorithm negatively impact the effectiveness of this technique for real-time applications.

More recent work by Kageyama et al. [3] utilizes an end-to-end convolutional neural network to create a preconditioned image to minimize projection blur. Their technique foregoes the need for feature markers as they do not estimate the PSF directly. The technique is also suitable for real-time conditioning for dynamic scenes. Although they simplify training by proposing the use of synthetically generated data, the reliance on a training process and the depen-

dence overhead on deep learning libraries could be considered as a limitation for memory constrained scenarios. We consider this work as state-of-the-art.

3. Our Approach

Our approach aims to overcome the real-time limitations of previous work by using advances in the commercialization of depth sensors to allow continous estimation of the PSF directly from depth data - no feature marker projection required. Additionally, we aim to do this using classic computer vision and geometrical techniques, obviating the need for deep learning libraries or model training.

[Hardware] Our hardware setup consists of an Intel RealSense L515 lidar depth camera, a 1920x1080 consumer-grade projector, and a laptop with an Intel i7-10750H CPU, Nvidia RTX 2070 Max-Q mobile GPU, and 16 GB of RAM.

[Calibration] We first calibrate and register the RGB and depth data retrieved from the camera. Simple checkerboard calibration is sufficient to extract camera intrinsics and distortion coefficients. We then compute a 3×3 homography to transform the camera image to the projectable area using simple contour and corner detection when we project a fully white image with the camera. This homography allows us to align and rectify our camera and projector image spaces to simplify our downstream image processing and projection operations. We assume that we will be utilizing calibrated, registered, and rectified RGB and depth images from this point onward.

[Focal Plane Estimation] During the hardware setup process, we capture a depth image of a planar surface at our projector's focal plane. This simplifies the physical setup of our projector-camera system, eliminating the need for manually measured quantities and special orientations of our projector and camera.

[PSF Estimation] Our technique's pipeline begins with capturing a depth image of the scene. We then calculate the absolute difference between the captured depth and our reference focal plane depth captured during the setup process. We model the defocus blur PSF as a Gaussian kernel with a variance that increases linearly with reciprocal distance (diopters) from the focal plane. We can thus assign a variance value for the PSF applied to a pixel based on the absolute difference values computed above.

[Wiener Deconvolution] Given the estimated PSF, we can apply Wiener deconvolution for all pixels to yield our preconditioned image.

[Piecewise Approximations with Tiles] This spatially-varying deconvolution, however, is computationally expensive to do on a per-pixel basis. Additionally, our depth data contains noise and can contain spurious outliers that can introduce unwanted artifacts in the image. We decrease the computational expense of spatially varying deconvolution by splitting the image into tiles and deconvolving each tile with an estimated PSF based on the mean value of the depth data corresponding to the tile's neighborhood. This has the additional benefit of denoising the depth data via averaging. Tile size becomes a user-defined hyperparameter that can be tuned to trade-off run-time performance with the spatial resolution at which we can correct for different defocus PSFs. For this work, we utilize tile sizes of 120×120 pixels, since it is the greatest common divisor of our native projector resolution's dimensions of 1920×1080 pixels. We acknowledge the limitations of this approach in our Limitations and Future Work section.

After piecewise Wiener deconvolving the image, we obtain our pre-conditioned image.

4. Evaluation

We evaluate our technique with two approaches. First, we simulate the projected image formation process and compare our approach across PSF estimations. We use our simulated results to characterize our technique's robustness to mismatches between estimated and true PSFs. We report image PSNR to quantitatively characterize the performance of our approach. We also include small cutaways of relevant images for qualitative comparisons. Additionally, our simulation allows us to evaluate hyperparameters such as Wiener filter SNR, to yield high-fidelity results.

Second, we create a physical setup using available equipment to experimentally evaluate our technique in a real-world scenario. Due to COVID-19 constraints, our optics lab was limited to the author's bedroom. We use the sliding doors of the author's closet to serve as an in-focus plane and out-of-focus plane with a high degree of repeatability. Using a cellphone camera to capture the projected results, we can report images for qualitative comparison.

In addition to image quality metrics, we also report metrics related to run-time performance.

Compared to related work, we achieve defocus deblurring for dynamic scenes without the need to project feature markers or rely on deep learning models. However, our method results in images with less quality compared to previous work, as indirectly estimating the PSF with depth instead of direct estimation with feature markers proves to be difficult. We discuss the effects of improperly estimating the PSF in the following section.

5. Results

Fig.1 below shows our theoretical results for a variety of images when the estimate of the PSF is well-matched to the true projection blur. The first column shows the original, ground truth images. The second columns shows our preconditioned image after applying Wiener deconvolution with a 50×50 Gaussian kernel with the sigma value specified in the image title. Empirically, we observed good re-

sults when estimating the SNR of the image as the mean of its pixels divided by the quantity 0.01. Notice that this approach tends to emphasize the high frequency content in the image. We apply a full-image deconvolution for these comparisons to emphasize the difference between our result and an unconditioned blurred image. In a real-world scenario, we would apply deconvolution piecewise across tiles. In the third column, we show a simulated projection blur on the ground truth image with a 50×50 Gaussian kernel with the sigma value specified in the image title. In the last column, we show the same simulated projection blur on our preconditioned image. Notice that on average we have an increase of 0.9 dB of PSNR, with a max of 1.3 dB of PSNR gain for most natural images. An interesting failure mode of this technique is show in the last row of Fig. 1, where a monochrome image of text suffers a slight decrease in PSNR. Qualitatively however, it is easy to judge that our results are much clearer than the standard blurred image for all images.

When the PSFs are not well-matched, we introduce either insufficient pre-conditioning or image artifacts such as ringing corresponding to under- and over-estimated PSF sigmas. We show each failure mode in Fig 2 and Fig 3 respectively within our simulated framework. Observe that the underestimate results in a still-blurry image, while the overestimate results in an unnatural looking image.

We then present examples of our technique applied in real-time in Fig. 4. Notice that we can adjust for changes in the scene geometry as the focal plane is removed and the image lands on an out-of-focus plane. We achieve a frame rate of approximately 5 frames per second with our laptop setup. We exhibit some amount of ringing in our final result as well as some artifacts from our tiled approach. This points overall to a mismatch in estimated PSFs as discussed earlier, as well as a need to apply some windowing or smoothing to improve the seams between tiles. We discuss methods of improving this technique in the next section.

6. Limitations and Future Work

Due to the ill-posed nature of the deconvolution problem for most traditional optics systems, we cannot fully remove the projection blur from the resulting image. Additionally, both our simulated and physical results show the importance of accurate estimation of the PSF - mismatches degrade the quality of the deblurring result or introduce unwanted, visible artifacts in the projection. Future work will be dedicated to more sophisticated methods of estimating the PSF itself and developing more accurate parametrization of the PSF with regards to distance from the focal plane.

Our piecewise analysis is a compromise between fidelity and run-time efficiency. We observe artifacts at the borders between tiles as a result of this approach. Future work in addressing this limitation can take two directions. 1) We could improve the tiled approach with better stitching algorithms and decreasing the size of tiles to achieve finer resolutions. 2) We can replace the tiled approach with a new approach entirely for spatially varying convolutions. One such method could be better image segmentation which can accurately piecewise separate the image into homogenous areas for deconvolution.

Another source of error for our technique comes from the noise of our depth sensors. Applying denoising solutions such as bilateral filtering to the depth map could improve the quality of our technique.

Finally, the setup constraints needed for our approach can be considered a limitation. We require knowing the projectable, in-focus area beforehand for our rectification process. This limits the technique to the extent of that precomputed area. Dynamic scenes can very obviously change the nature of the projectable area and thus we can fail to capture the full geometry of the scene.

7. Conclusion

We have presented an approach for defocus deblurring for dynamic scenes using commercially available depth sensors, projectors, and compute hardware. Our approach is simple and computationally feasible to perform in real-time to adjust to new projection geometries. Our simulated results show a noticable quantitative and qualitative improvement, and with better PSF estimation techniques, we believe we can improve the quality of our physical results as well.

References

[1] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer. The visual computing of projector-camera systems. In *ACM SIGGRAPH 2008 Classes*, SIGGRAPH '08, New York, NY, USA, 2008. Association for Computing Machinery.

[2] M. S. Brown, P. Song, and T. Cham. Image pre-conditioning for out-of-focus projector blur. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2:1956–1963, 2006.

[3] Y. Kageyama, M. Isogawa, D. Iwai, and K. Sato. Prodebnet: projector deblurring using a convolutional neural network. *Opt. Express*, 28(14):20391–20403, Jul 2020.

[4] R. Vyas, L. R. Sayadi, D. Bendit, and U. Hamdan. Using virtual augmented reality to remotely proctor overseas surgical outreach: Building long-term international capacity and sustainability. *Plastic and Reconstructive Surgery*, 146:622e–629e, Nov 2020.

[5] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. *ACM Trans. Graph.*, 25(3):907–915, July 2006.

Figure 1. Theoretical Results for Matched PSFs

Figure 2. Underestimate of Blur

Figure 3. Overestimate of Blur

Figure 4. Dynamic image conditioning frame-by-frame. The black bar indicates the edge of the front plane, which is the focal plane. The focal plane is pulled back to the left to reveal a plane behind it that is out of focus. In the third and fourth image, the compensation effect can be seen activating for areas of the image which are on the out of focus plane.