

An inverse problem approach to depth from defocus via active quasi-random colored point projections

Avery Ma and Alexander Wong

Vision and Image Processing Lab, Systems Design Engineering, University of Waterloo, East Campus 4, 295 Phillip Street, Waterloo, Ontario, Canada, N2L 3G1

E-mail: b24ma@uwaterloo.ca, a28wong@uwaterloo.ca

Abstract. A novel methodology for inferring depth measurements using active quasi-random colored point projection patterns is presented by recourse to computational inverse modelling within the context of depth from defocus. The forward problem is considered that when capturing objects using a camera, light reflected off of objects at different depths will hit the lens at different degrees outside of focal range which results in blurry observations in the acquired image, whereas the reverse analysis seeks to estimate the depth from the knowledge of the blurry observations. In the proposed approach, a colored quasi-random point projection pattern is projected onto the scene of interest, and each projection point in the image captured by a cellphone camera is analyzed using an inverse model to estimate the depth at that point. The proposed method has a relatively simple setup, consisting of a camera and a projector, and enables depth inference from a single capture. Furthermore, the use of a quasi-random projection pattern can allow us to leverage compressive sensing theory to produce full depth maps in future applications. We conclude with qualitative and quantitative results that highlight the strong potential for enabling active depth sensing in a simple, efficient manner.

1. Introduction

The problem of robust and fast 3D capture and reconstruction of objects and scenes has always been one of the core research tasks for image processing and robotic applications. Depth information is crucial for many applications including 3D reconstruction, action recognition, human-computer interfacing, etc. As such, depth cameras have received much attention both academically and in industry with constant advancements to depth camera technology.

Traditionally, active depth sensing techniques include methods such as projector-camera structured light systems, laser scanners, ToF cameras, and IR projected point patterns such as the Microsoft Kinect [5]. While recent efforts have lowered the cost of depth camera sensors, depth sensing still requires specialized complex hardware that are not compact. Motivated by these challenges, in this paper we present a system for inferring sparse depth measurements from a single camera capture by leveraging active quasi-random point projections and camera defocus.

We consider a model of forward problem, when capturing objects using a camera, light reflected off of objects at different depths will hit the lens at different degrees outside of focal range which results in blurry observations in the acquired image. The corresponding inverse problem can be established where an estimation of the depth is sought from the knowledge of the blurry observations. Specifically, the forward and inverse model can be formulated

mathematically as:

$$m = f(d) \quad (1)$$

$$d = f^{-1}(m) \quad (2)$$

where m is the camera measurement data of the projected pattern and d is the depth of the surface where the pattern is projected onto. The proposed method of inferring sparse depth has a relatively simple setup, thus can potentially lead to very compact and low cost active depth sensing systems.

The proposed method involves projecting a colored quasi-random point pattern onto the scene of interest. The colored quasi-random point pattern can for instance be generated by using a light source with a point pattern mask, resulting in much simpler and cheaper hardware configuration. The detected projected points on the imaged scene are analyzed to determine their minimum eigenvalue of the covariance matrix, which are used to estimate depth via an inverse model.

2. Related Work

Depth from defocus (DFD) methods generally estimate depth by analyzing the difference in blurriness of two images captured at different focal lengths [12, 8, 11], with different methods using different filters for determining the measure of blur. A major drawback to such DFD methods is the unreliable detection of blur, especially in untextured areas of the image. This problem is mitigated in active depth sensing, where an optical projection is used to find correspondences for triangulation. A review of structured light patterns for depth measurement is provided by Salvi *et al.* [10]. Therefore, using active projection patterns does not depend on the objects in the scene and is also effective in untextured regions. On the other hand, active depth sensing systems suffer from occlusion and require complex hardware. As such, we are motivated in the proposed system to takes advantage of the benefits from both DFD and active depth sensing methods to design a system that has a simple setup yet reliable in the depth measurements.

The concept of using DFD using active projections have also been explored in literature. Pentland *et al.* [9] used evenly spaced line projections to determine depth from line spread. This simple method is able to create low resolution depth maps. Nayar *et al.* [7] used a dual sensor plane with optimized projection and camera setup to produce a dense depth map and reduce front/back focal ambiguity. Ghita *et al.* [2] used a dense projected pattern with a tuned local operator for finding the relationship between blur and depth. Moreno *et al.* [6] used an evenly spaced point pattern with defocus to obtain an approximate depth map used for automatic image refocusing. These methods use a high density projection pattern which require either a projector or more specialized calibrated hardware.

Promising results were shown in the preliminary approach by Ma *et al.* [4]. In the previous work, blurry projected pattern as captured by the camera was modelled using the standard deviation of a circularly-symmetric 2D Gaussian which did not account for the distortion of the pattern after projection. Also, the projected quasi-random pattern from the previous approach only consists of white dots, which could be enhanced by projecting colored pattern with different wavelengths.

3. Proposed Method

The proposed system can be described as follows. A colored quasi-random point pattern is projected onto the scene, which is then captured by a camera. The camera's focus is fixed such that the degree of focus of each point in the quasi-random point pattern as it appears in the captured image is dependent on the depth of the surface. The inverse model between depth and the blurriness of the projected point pattern is characterized in a calibration step, and used to

infer the depth at each point to produce sparse depth measurements. A one-time calibration step is required to learn the inverse model.

3.1. Calibration

The purpose of the calibration procedure is to characterize the relationship between point projection measurement and depth and establish the mathematical formulation for inverse model $f^{-1}(m)$. When out of focus, a projected point will appear blurred, with the degree of blurriness correlated with the depth of the scene at that point. The blurry projected point pattern as captured in the image can be approximately modelled by an elliptical 2D Gaussian with mean and covariance, and the minimum eigenvalue of the covariance matrix is computed to characterize the relationship between the depth and the blurriness of the point pattern. This model can be used as an inversion operator mapping the range of minimum eigenvalues to a unique depth estimation corresponding to each projected point and the sparse depth map can be obtained. The blur effect of a projected point is visualized in Figure 1.

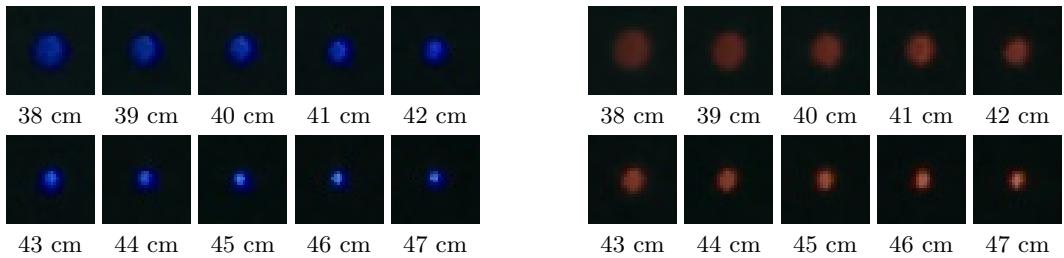


Figure 1: Projected blue and red points on an object at various distances (38cm to 47cm) away from the setup, as captured by the camera

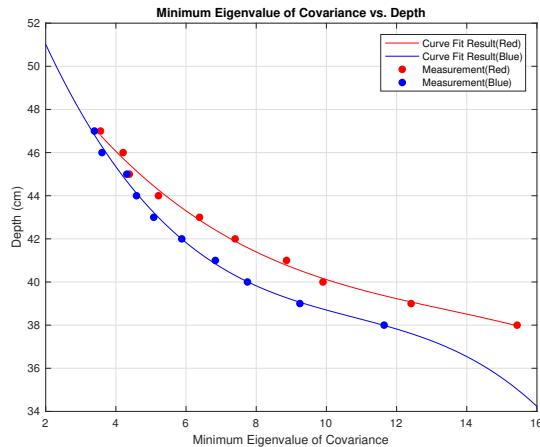


Figure 2: Inverse model characterizing minimum eigenvalue of covariance matrix vs. distance

Therefore, to learn the inverse model, a point pattern is projected onto various planes with known distances away from the projector-camera setup. The measured data points corresponding to minimum eigenvalue of the covariance matrix vs. depth can then be used to construct the inverse model characterizing the relationship between point projection measurement and the distance away from setup. To obtain a continuous curve for the inverse

model, regression with a third order polynomial function is used to fit the data points:

$$f^{-1}(m) = a * \lambda(m)^3 + b * \lambda(m)^2 + c * \lambda(m) + d \quad (3)$$

where $\lambda(m)$ is the minimum eigenvalue of the covariance matrix given the camera measurement at corresponding location. A visualization of the inverse model is shown in Figure 2. The third order polynomial regression is sought by fitting the data to the model:

$$\text{argmin}(\sum(d_i - f^{-1}(m_i))) \quad (4)$$

3.2. Sparse Depth Estimation Pipeline

With the calibrated model, the proposed system can then be used to estimate sparse depth of the scene. To this end, the proposed depth recovery method can be divided into 4 main stages outlined in Figure 3 and described as follows.

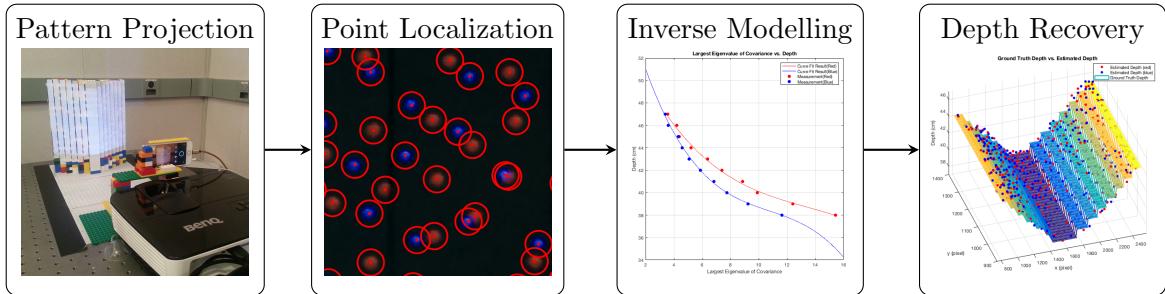


Figure 3: Flowchart of the depth recovery pipeline

Active Quasi-random Point Projection: A colored quasi-random point pattern is projected onto the scene. In the current system, a Poisson Disc Sampling method was utilized to generate the quasi-random point pattern such that the random points are tightly packed together, but no closer than a specified minimum distance [1]. Compared to other random sampling methods, Poisson-Disc method significantly reduces the chances of having overlaps between blurred projected points of the same wavelength, which would result in erroneous depth recovery.

Point Localization: After the projected point pattern has been captured by the camera, colored pattern are separated by their corresponding wavelength. Otsu's method is used to obtain a binary map consisting of regions corresponding to the projected points [3].

Estimation of Minimum Eigenvalue of Covariance Matrix: Based on the detected regions of interest of the projected points, the minimum eigenvalue of covariance matrix of each projected point is estimated by using an intensity-weighted approach, where the number of samples at each location is approximated by multiplying the pixel intensity by a factor of 10,000. An elliptical Gaussian can be formulated to model the generated sample populations:

$$p(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} ((\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})) \right\} \quad (5)$$

where $\underline{\mu}$ is the sample mean and Σ is the sample covariance. The minimum eigenvalue of covariance matrix can be computed as follows:

$$\lambda = \min \{ \lambda_{1,2} : \det(\Sigma - \lambda_{1,2} I) = 0 \} \quad (6)$$

Recovery of Sparse Depth Map: Based on the minimum eigenvalue of covariance matrix of a projected point, the inverse model can then be used to infer the depth corresponding to that projected point. By performing this on all projected points in the quasi-random point projection pattern, the sparse depth map can be obtained.

4. Experimental Setup

The main goal of this current realization of the proposed technique is to build a compact and portable system to obtain depth information of the scene. For this purpose, the scene is imaged using an iPhone 5S camera (resolution: 3264×2448) and the quasi-random point pattern is projected using a BENQ MH630 Digital Projector (resolution: 1920×1080).

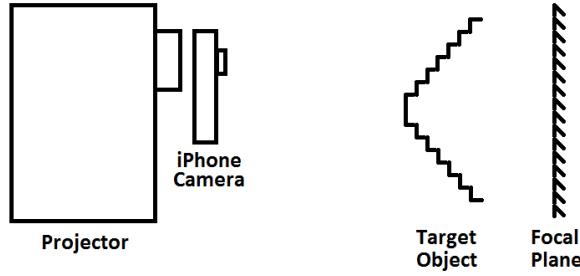


Figure 4: The diagram of the experimental setup

The experimental setup is described in Figure 4. The iPhone camera is mounted 3 cm before the projector. The camera is in-line with the projection lens. The focal plane of both projector and the camera are on a plane 50 cm away from the projector.

5. Results

The test scene was a two-way staircase with 1 cm step-size away from the setup, and the quasi-random point density is approximately 0.20% of the projector resolution.

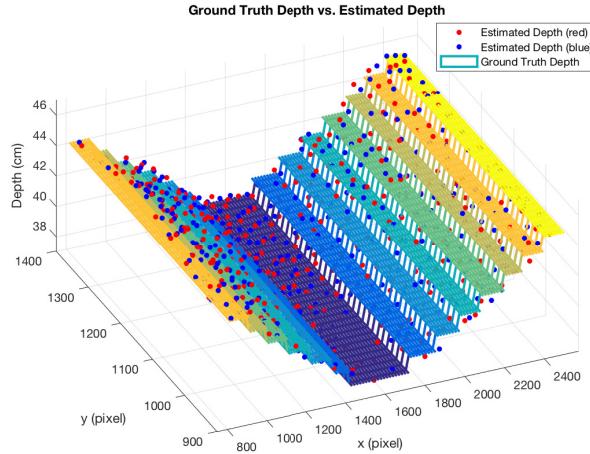


Figure 5: Sparse depth estimation of the staircase with ground truth surfaces

Figure 5 shows the depth recovery of the object compared to the ground truth surfaces. To evaluate the performance of the proposed method, the root mean square error (RMSE) and correlation coefficient were computed between the estimated depth and the ground truth depth at corresponding locations.

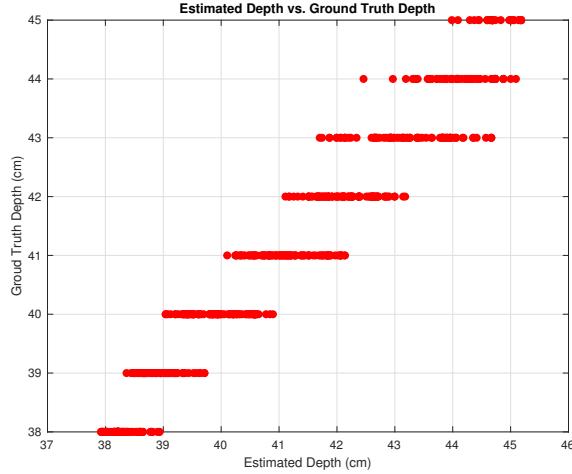


Figure 6: Estimated depth vs. ground truth depth

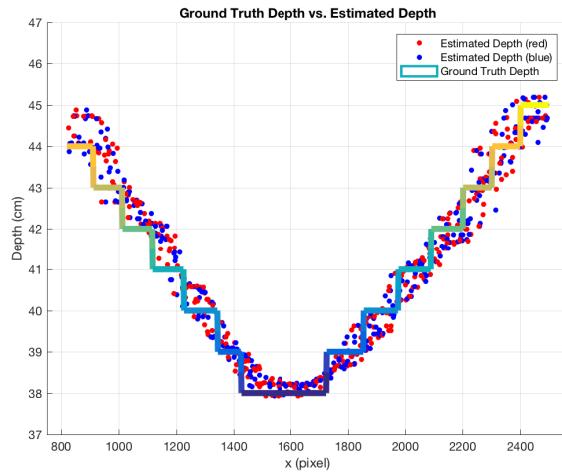


Figure 7: Top view of the reconstructed sparse depth estimation of the staircase

To evaluate the reconstruction errors, the RMSE value is calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \hat{d}_i)^2} \quad (7)$$

over the n captured dot pattern. The RMSE value of the system is $\pm 0.51\text{cm}$, which demonstrates that the proposed depth recovery pipeline is capable of achieving 3D reconstruction with high accuracy. Figure 7 shows the top view of the sparse depth recovery of the test scene.

The correlation coefficient was computed to be **0.9726**, which demonstrates that a strong positive correlation between the ground truth depth and the estimated depth. This strong correlation between the ground truth depth map and the estimated depth can also be observed in the sample point visualization shown in Figure 6.

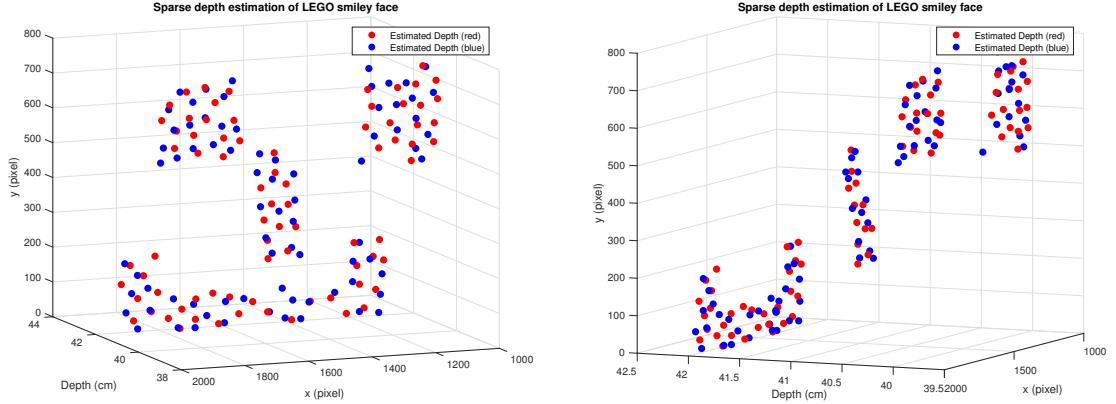


Figure 8: Sparse depth estimation of the LEGO smiley face

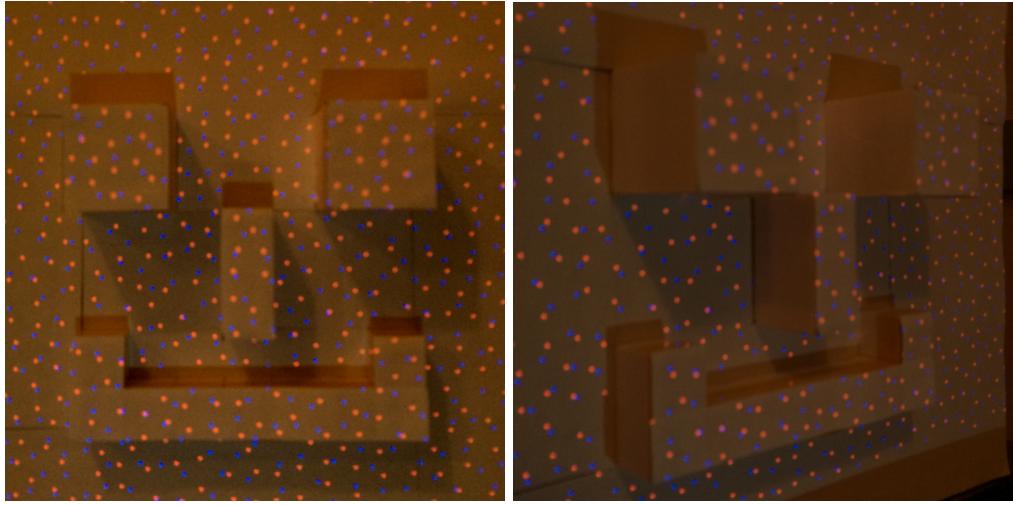


Figure 9: Actual scene of the LEGO smiley face

Figure 8 illustrates the depth estimation of a smiley face made of LEGO. It can be observed that the proposed method can also achieve great results when imaging complex geometric shapes and objects. Figure 9 shows the actual scene of the smiley face imaged by the camera.

6. Conclusion

This work presented a sparse depth measurement system based on active quasi-random colored point projection and depth from defocus. The new enhanced method extends upon previous approach by leveraging an extended inverse model that models the projected pattern as an elliptical Gaussian function as opposed to a symmetric Gaussian, to better account for the distortions of the pattern after projection. Colored point projection drastically increases the

density of the reconstructed sparse map. The main advantage of the proposed method is its simplicity in hardware and computation, requiring merely a camera and projected dot pattern. There are many additional future directions that can be explored. Deep learning approach can be investigated to better model the relationship between the point projection measurement and its corresponding depth. In addition, different spatial-temporal, data-guided quasi-random point projection patterns at various wavelengths can be examined to optimize reconstruction accuracy while facilitating for fast acquisitions.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs Program, and the Ontario Ministry of Research and Innovation.

References

- [1] R. Bridson. Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH 2007 sketches*, page 22. ACM, 2007.
- [2] O. Ghita, P. F. Whelan, and J. Mallon. Computational approach for depth from defocus. *Journal of Electronic Imaging*, 14(2):023021–023021, 2005.
- [3] L. Jianzhuang, L. Wenqing, and T. Yupeng. Automatic thresholding of gray-level pictures using two-dimension otsu method. In *Circuits and Systems, 1991. Conference Proceedings, China., 1991 International Conference on*, pages 325–327. IEEE, 1991.
- [4] A. Ma, F. Li, and A. Wong. Depth from defocus via active quasi-random point projections. *Journal of Computational Vision and Imaging Systems*, 2(1), 2016.
- [5] Microsoft. Kinect hardware (online). <https://developer.microsoft.com/en-us/windows/kinect/hardware>, 2016.
- [6] F. Moreno-Noguer, P. N. Belhumeur, and S. K. Nayar. Active refocusing of images and videos. *ACM Transactions On Graphics (TOG)*, 26(3):67, 2007.
- [7] S. K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1186–1198, 1996.
- [8] A. Pentland, T. Darrell, M. Turk, and W. Huang. A simple, real-time range camera. In *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR'89., IEEE Computer Society Conference on*, pages 256–261. IEEE, 1989.
- [9] A. Pentland, S. Scherock, T. Darrell, and B. Girod. Simple range cameras based on focal error. *JOSA A*, 11(11):2925–2934, 1994.
- [10] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004.
- [11] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, 1998.
- [12] Y. Xiong and S. A. Shafer. Depth from focusing and defocusing. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 68–73. IEEE, 1993.