An inverse problem approach to computational active depth from defocus

A Ma and A Wong
Vision and Image Processing Lab, Systems Design Engineering, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1
E-mail: avery.ma@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 result in blurry observations in the acquired image, whereas the inverse problem 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. 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 [1]. 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 result 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:

\[ m = f(d) \]
\[ d = f^{-1}(m) \] (2)

where \( m \) is the camera measurement data of the projected pattern and \( d \) is the depth of the surface on which the pattern is projected. 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 a more compact and self-contained depth sensing device. The detected projected points on the imaged scene are each analyzed to determine the minimum eigenvalue of the related covariance matrix, and the set of these minimum eigenvalues is 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 [2, 3, 4], 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. [5]. 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 take advantage of the benefits from both DfD and active depth sensing methods to design a system that has a simple setup yet is reliable in the depth measurements.

The concept of DfD using active projections has been explored in the literature. Pentland et al. [6] 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. [8] used a dense projected pattern with a tuned local operator for finding the relationship between blur and depth. Moreno et al. [9] 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. [10]. In the previous work, blurry projected points as captured by the camera were modelled using the standard deviation of a circularly-symmetric 2D Gaussian. However, in traditional projectors, it occurs that one-pixel point source from the projector does not converge into an ideal point after transmission through the projector-lens system. This causes skewness of the projected point, which leads to erroneous depth inference results. As such, an enhanced active DfD technique that addresses this issue is highly desired.

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 relating the 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 the 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 points as captured by the camera are modeled using an elliptical Gaussian intensity map, and the minimum eigenvalue of the elliptical Gaussian covariance is used to characterize the depth. It was noticed that the maximum eigenvalue of the covariance matrix corresponds to the magnitude of the skew caused by projector distortion, whereas the minimum eigenvalue is significantly less affected under the distortion. As such, the minimum eigenvalue can better preserve the geometric information of the actual projected points. The blur effect of a projected point is visualized in 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](image1)

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.

![Inverse model characterizing minimum eigenvalue of covariance matrix vs. distance](image2)

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

This eigenvalue-depth 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. 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:

\[
f^{-1}(m) = a_3 \lambda(m)^3 + a_2 \lambda(m)^2 + a_1 \lambda(m) + a_0
\]

where \( \lambda(m) \) is the minimum eigenvalue of the covariance matrix given the camera measurement at the corresponding location. The third order polynomial function is used to fit the data points because it is the lowest order of a polynomial function to provide best fit to the series of data points. It can be observed from Figure 2 that the eigenvalue-depth curve is dependent on the color of the projection pattern. This is due to the discrepancy in the spectral response between channels in the camera sensor. The third order polynomial regression is sought by fitting the
data to the model:
\[
\arg\min \left( \sum (d_i - f^{-1}(m_i))^2 \right) \tag{4}
\]

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.

Active Quasi-random Point Projection: A colored quasi-random point pattern is projected onto the scene. The use of projection pattern with different wavelengths can provide increased spatial resolution beyond single-wavelength approaches. The camera sensor captures images in three unique ranges of the visible spectrum: red, green and blue. The red and blue channels lie the furthest apart in the spectrum among the three channels, which makes them easily separable from each other. As such, the red and blue projection patterns are selected to achieve improved reconstruction results with an accurate point localization. In the proposed approach, Poisson-disc sampling method was utilized to generate the quasi-random point pattern. Poisson-disc sampling method produces random points that are tightly packed together, but no closer than a specified minimum distance [11]. Compared to other random sampling methods, Poisson-disc method significantly reduces the chance of having overlaps between blurred projected points of the same wavelength, which would result in erroneous depth recovery. Moreover, the size of the projection point is chosen to minimize overlapping of the blurred projected points at the closest depth level.

Point Localization: After the projected point pattern has been captured by the camera, colored patterns are separated by their corresponding wavelength. Otsu’s method is used to obtain a binary map consisting of regions corresponding to the projected points [12].

Estimation of Minimum Eigenvalue of the Covariance Matrix: Based on the detected regions of interest of the projected points, the minimum eigenvalue of the 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:
\[
p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right\} \tag{5}
\]
where \( \mu \) is the sample mean and \( \Sigma \) is the sample covariance. The minimum eigenvalue of the covariance matrix can be computed as follows:

\[
\lambda = \min \{ \lambda_{1,2} : \det(\Sigma - \lambda_{1,2}I) = 0 \}
\]  

(6)

**Recovery of Sparse Depth Map:** Based on the minimum eigenvalue of the 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: \( 3264 \times 2448 \)) and the quasi-random point pattern is projected using a BENQ MH630 Digital Projector (resolution: \( 1920 \times 1080 \)).

![Figure 4: The diagram of the experimental setup](image)

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 the projector and the camera is 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. Also, the possible depth range of the inference system is limited to the calibrated depth values.

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.

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

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

(7)

over the n captured point 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 6 shows the top view of the sparse depth recovery of the test scene.
Figure 5: Sparse depth estimation of the staircase with ground truth surfaces

Figure 6: Top view of the reconstructed sparse depth estimation of the staircase

Figure 7: Estimated depth vs. ground truth depth

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 7.

Figure 8 illustrates the depth estimation of a smiley face made of LEGO. The RMSE value between the ground truth and the estimated depth of the LEGO smiley face is ±0.14 cm, with a high correlation coefficient of 0.9803. Both visual appearance and quantitative results demonstrate that the proposed method can also achieve great performance when imaging complex geometric shapes and objects. Figure 9 shows the actual scene of the smiley face imaged by the camera. It is worth mentioning that the above results are limited to surfaces that are parallel to the image plane, and the results are expected to degrade for angled surfaces since they are not considered in the calibration stage.

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 point 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. Different surface orientations can be considered during the calibration stage to improve the robustness of the method when imaging angled planes. 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] Microsoft 2016 Kinect Hardware (online) https://developer.microsoft.com/en-us/windows/kinect/hardware
[2] Xiong Y and Shafer S A 1993 Depth from focusing and defocusing Proc. CVPR’93 pp 68-73
[3] Pentland A, Darrell T, Turk M and Huang W 1989 A simple, real-time range camera Proc. CVPR’89 pp 256-261
[4] Watanabe M and Nayar S K 1998 Rational filters for passive depth from defocus International Journal of Computer Vision 27 pp 203-225
[5] Salvi J, Pages J and Batlle J 2004 Pattern codification strategies in structured light systems Pattern recognition 37 pp 827-849
[6] Pentland A, Scherock S, Darrell T and Girod B 1994 Simple range cameras based on focal error JOSA A 11 pp 2925-2934
[7] Nayar S K, Watanabe M and Noguchi M 1996 Real-time focus range sensor IEEE Transactions on Pattern Analysis and Machine Intelligence 18 pp 1186-1198
[8] Ghita O, Whelan P F and Mallon J 2005 Computational approach for depth from defocus Journal of Electronic Imaging 14 023021
[9] Moreno-Noguer F, Belhumeur P N and Nayar S K 2007 Active refocusing of images and videos ACM Transactions On Graphics 26 pp 67
[10] Ma A, Li F and Wong A 2016 Depth from Defocus via Active Quasi-random Point Projections Journal of Computational Vision and Imaging Systems 2 1
[11] Bridson R 2007 Fast Poisson disk sampling in arbitrary dimensions SIGGRAPH sketches pp 22
[12] Liu J, Li W and Tian Y 1991 Automatic thresholding of gray-level pictures using two-dimension Otsu method Proc. International Conference on Circuits and Systems pp 325-327