Deep Learning-driven Depth from Defocus via Active Multispectral Quasi-random Projections with Complex Subpatterns

Avery Ma
Dept. of Systems Design Engineering
University of Waterloo
Waterloo, Canada
avery.ma@uwaterloo.ca

Alexander Wong
Dept. of Systems Design Engineering
University of Waterloo
Waterloo, Canada
a28wong@uwaterloo.ca

David A Clausi
Dept. of Systems Design Engineering
University of Waterloo
Waterloo, Canada
dclausi@uwaterloo.ca

***Abstract*—A promising approach to depth from defocus (DfD) involves actively projecting a quasi-random point pattern onto an object and assessing the blurriness of the point projection as captured by a camera to recover the depth of the scene. Recently, it was found that the depth inference can be made not only faster but also more accurate by leveraging deep learning approaches to computationally model and predict depth based on the quasi-random point projections as captured by a camera. Motivated by the fact that deep learning techniques can automatically learn useful features from the captured image of the projection, in this paper we present an extension of this quasi-random projection approach to DfD by introducing the use of a new quasi-random projection pattern consisting of complex subpatterns instead of points. The design and choice of the subpattern used in the quasi-random projection is a key factor in the ability to achieve improved depth recovery with high fidelity. Experimental results using quasi-random projection patterns composed of a variety of non-conventional subpattern designs on complex surfaces showed that the use of complex subpatterns in the quasi-random projection pattern can significantly improve depth reconstruction quality compared to a point pattern.**

***Keywords*-active depth sensing; depth from defocus; deep learning; computational modelling;**

I. Introduction

Depth is one of the fundamental cues in enabling robots and machines to understand their environment. Depth sensor plays an important role in computer vision tasks such as object recognition and scene understanding, especially in real-life complex environment.

There are many camera-based depth-sensing techniques in the literature [1]. Active systems based on triangulation have been gaining popularity due to its superior performance and ease of application [2]. For example, stereo structured-light systems measure the distance to the object by analyzing the disparity between the coded light pattern and the captured image of the pattern from the illuminated scene. However, a major disadvantage of the triangulation system is the necessary baseline to operate, which induces a minimum size constrain on the system that makes it ineffective to utilize for in certain scenarios. In many cases, it also requires the scene to be illuminated with a projected pattern that is high-powered and well-focused in order to apply triangulation, and hence it relies on specialized hardware that increases cost and complexity. As such, alternative active depth-sensing techniques that address these challenges are required.

Recent developments in active depth-sensing methods has led to a promising approach that involves actively projecting a quasi-random point pattern onto an object and assessing the blurriness of the point projection as captured by a camera to recover the depth of the scene [3]. Specifically, the projected pattern is captured using a RGB camera, and an ensemble of deep neural networks is used to estimate point-wise depth based on the subpattern in the captured image. Throughout this paper, we use the term subpattern to refer to the individual element that forms the overall projection pattern. A computational reconstruction method is used to generated a final depth map from the sparse depth estimation results from the deep neural networks. Preliminary results demonstrate that such depth inference approach can effectively address the aforementioned shortcomings of the triangulation-based stereo structured light system: it eliminates the need for a baseline (the projector can be completely in-line with the camera), and the need of a well-lighted pattern. Most importantly, it has a relatively simple setup which would be expected to lead to very compact and low-cost active depth-sensing systems.

Additionally, it was discovered that the depth inference can be made not only faster but also more accurate by leveraging deep learning approaches to computationally model and predict depth based on the captured images of the point subpattern. Motivated by the fact that deep learning techniques can automatically learn useful features from the captured image of the projection, in this paper we present an extension of this quasi-random projection approach to DfD by introducing the use of a new quasi-random projection pattern consisting of complex subpatterns instead of points.

II. Related Work

Active DfD methods generally estimate depth by analyzing the visual variation of the projected pattern captured

at different focal lengths. Pentland *et al.* [4] proposed a low-resolution depth estimation method based on the line spread of the evenly-spaced line projections. Ghita *et al.* [5] suggested projecting a dense projected pattern onto the scene and using a tuned local operator designed for finding the relationship between blur and depth. Moreno *et al.* [6] proposed the use of an evenly spaced point pattern with defocus to approximate depth in the context of automatic image refocusing. Traditional active DfD is elegant and holds a lot of promise due to its simplicity. However, a major drawback to such methods is its complex hardware requirements to simultaneously capture far and near focused image with different blurriness in order to estimate depth, which makes them ineffective to utilize for in practical scenarios.

Ma *et al.* [3], [7], [8], [9] have recently developed an alternative approach based on computational active DfD. The method leveraged the active projection of multispectral quasi-random point subpatterns and non-parametric modelling of the level of blurring associated with the projected subpattern as captured by a camera. This approach shows considerable promise as it does not require custom projectors or specialized calibration hardware, and thus can enable low-cost, compact depth-sensing systems.

However, previous studies by Ma *et al.* [3], [7], [8], [9] have almost exclusively focused on exploring different modelling methods to characterize the blurring of the projected subpattern at different depth levels. While deep convolutional neural networks were adopted as a powerful tool for computational modelling [3], one area that was not well explored is how to leverage subpattern with complex shapes to fully take advantage of the benefit from deep convolutional neural networks. As such, the aim of this study is to evaluate the efficacy of the computational active DfD approach with quasi-random projection patterns consisting of complex subpatterns.

III. Method

In the extension of the quasi-random projection approach to DfD, the scene is illuminated by a quasi-random projection pattern consisting of numerous complex subpatterns in blue and red wavelengths, and then captured by a RGB camera. The camera's focus is fixed such that the level of blurring associated with the projection pattern as it appears in the captured image is dependent on the depth of the surface. Next, the subpatterns are extracted from the acquired image and then passed into an ensemble of deep convolutional neural networks, with each network responsible for estimating the depth of a projected subpattern at a different spectral wavelength. As such, the ensemble of deep convolutional neural networks produces sparse depth measurements at different wavelengths, and a training process is required to learn the ensemble of deep convolution neural networks. The final depth map is reconstructed via triangular-based interpolation based on the sparse depth measurements.

A. Subpattern Designs for Active Scene Illumination

The key idea behind the active DfD approach is that out-of-focus projection pattern will appear blurred, with the degree of blurriness correlated with the depth of scene at that point. In computer vision literature, 2D Gaussian has been widely employed to approximate the point spread function of the blurring effect [10]. For example, the blurring effect of the conventional point subpattern can be ideally represented using the standard deviation of the Gaussian operator, but the main constrain is that there are only very limited amount of features in the point subpattern. On the other hand, subpatterns with complex designs can contribute to increased variation in the camera measurement data, and the use of deep convolutional neural networks as an non-parametric model can enable automatic feature learning from the complex subpatterns, leading to a significant increase in the number of useful features. As such, by leveraging deep convolutional neural networks with complex subpatterns, one can achieve higher fidelity in the depth reconstruction results.

One consequence from having non-conventional subpattern design is that it greatly increases the chance of having overlapped blurring subpatterns with the same wavelength, resulting in erroneous depth estimate. Therefore, we limit the size of the subpattern to be within a 3×3 region, so the quality of the captured images of subpatterns can be retained. The proposed subpattern designs are illustrated in Figure 1.

(a) (b) (c) (d)

Figure 1: Proposed subpattern designs: a) Square, b) Cross, c) Triangle, d) X

B. Ensemble of Deep Convolutional Neural Networks for Depth Inference

Instead of using a parametric model to approximate the blurring of the subpattern, the use of an ensemble of deep convolutional neural networks is proposed to learn an non-parametric mapping from the camera measurement of the blurring subpattern at different wavelengths directly to the depth of scene at that point. Each deep convolutional neural network in the ensemble is responsible for performing depth inference at a particular wavelength.

In our work, each proposed subpattern is trained using an independent ensemble of deep convolution neural network. To train each network in the ensemble, a quasi-random

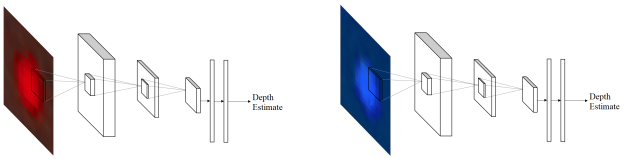

Figure 2: Architecture of the proposed ensemble of convolutional neural networks for depth inference using red projection pattern(left) and blue projection pattern (right)

projection pattern is projected onto a vertical surface placed at known distances away from the projector-camera setup. The projected subpatterns are extracted from the acquired images and a 30×30 image patch of pixels is formed at the location of each subpattern. The subpattern-depth model is trained with supervised learning using ground truth depth label, and the same training procedure was repeated for all proposed subpatterns.

Having a sufficient number of captured images of the blurring subpattern at each depth level greatly generalizes the learning process, and is the key to an accurate depth inference model. As such, we augment the dataset by projecting a total of four quasi-random projection patterns for each subpattern at every depth level. The four projection patterns consist of the actual quasi-random projection pattern, and three one-pixel-shifted versions (horizontal, vertical, and diagonal) of the actual pattern which closely resembles the blurriness of the original pattern.

Table I: Summary of the architecture for one network in the ensemble of convolutional neural networks for depth inference

	Layer Description	Output Tensor Dim.
	Input image	$30 \times 30 \times 1$
1	5×5 conv, 16 filters	$26 \times 26 \times 16$
2	5×5 conv, 32 filters	$22 \times 22 \times 32$
3	5×5 conv, 64 filters	$18 \times 18 \times 64$
4-5	Fully-connected	20736×1
	Output depth	1×1

Subpatterns captured from the projection of the three shifted versions of the quasi-random pattern are used to train the deep convolutional neural networks in the ensemble, and the actual quasi-random pattern is used for quantitative evaluation. There are 3,883 subpatterns in the original quasi-random pattern. The three shifted versions of the projection pattern result in a total of 11,649 30×30 images for each depth label. In our model, we learn a deep representation through numerous 2D convolutional operations. Each convolutional layer is followed by a rectified linear nonlinearity. We append 2 fully connected layers at the end of the network. The network architecture is illustrated in Fig 2, with a more detailed layer-by-layer definition in Table I.

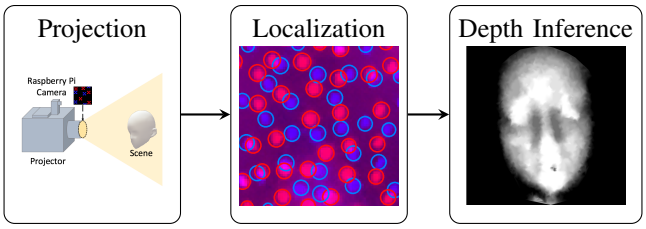

Figure 3: Illustration of the proposed extension of the quasi-random projection approach to DfD. The scene is actively illuminated with a multispectral quasi-random pattern consisting of numerous complex subpatterns, and a RGB camera is used to capture images of the projected pattern. The ensemble of deep convolutional neural networks then analyses the level of blurring associated with the captured images of the subpattern, and predict depth at each location in the projected pattern. With depth measurements at all locations predicted using the ensemble, triangulation-based interpolation is performed to generated the final depth map.

C. Depth Inference Pipeline

With the ensemble of deep convolutional neural networks, the depth of the scene can be estimated. To this end, the depth recovery method can be divided into 3 main stages outlined in Fig 3 and described as follows.

Stage 1: Multispectral Active Quasi-random Pattern Projection: A multispectral quasi-random pattern consisting of the proposed subpattern is projected onto the scene. Poisson-disc sampling (PDS) method was utilized to generate the location for the subpatterns in each projection pattern such that the subpatterns are tightly packed together, but no closer than a specified minimum distance [11]. Given projector resolution $[x, y]$, the PDS algorithm $\phi(\cdot)$ can be expressed as:

$$P = \phi(x, y, \rho, d) \tag{1}$$

where ρ is the desired subpattern density, d is the minimum distance between subpatterns and P is a quasi-random point map. The operation to generate the light pattern $\hat{P}$ with the subpattern S can be formulated as:

$$\hat{P} = P \oplus S \tag{2}$$

where $\oplus$ denotes the image dilation operation and S serves as a structuring element to expend the point contained in the point map. Compared to other random sampling methods such as Sobol sequence and Halton sequence [12], PDS method significantly reduces the chances of having overlaps between blurred projected subpattern, which would result in erroneous depth recovery. To generate the multispectral light pattern, PDS is performed once for each wavelength and the results are concatenated into a single projection pattern.

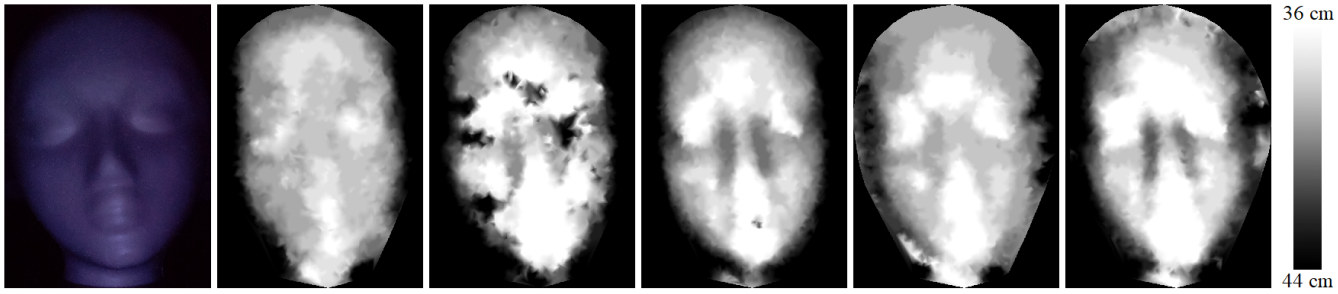

Figure 4: A grayscale representation of the reconstructed depth maps using different subpatterns. From left to right: Styrofoam mannequin head as captured by camera, depth map generated using the point subpattern, depth map generated using the square subpattern, depth map generated using the cross subpattern, depth map generated using the triangle subpattern, depth map generated using the X subpattern.

Stage 2: Point Localization: After the projected pattern has been captured by the camera, subpattern corresponding to the same wavelength can be effectively separated by taking the single channel measurements from the camera. We use Otsu's method to obtain a binary map consisting of regions of the projected subpatterns [13]. The centroid of each region is computed and a 30×30 image patch is formed at each subpattern location.

Stage 3: Depth Inference and Depth Image Reconstruction: After identifying the projected subpattern in the acquired scene, the ensemble of deep convolutional neural networks can then be used to predict the depth corresponding to that projected subpattern. By performing this on all projected subpatterns in the quasi-random projection pattern, the sparse depth estimation can be obtained. With depth measurements at all detected locations, triangulation-based linear interpolation is performed to reconstruct the final depth map.

IV. Results

The performance of the depth inference model with proposed subpatterns was evaluated by means of quantitative and qualitative tests. For comparison purposes, we compare with the conventional point subpattern used in a published method in [3]. A low-cost active DfD system was used in the experiment. The multispectral quasi-random projection pattern is projected using a BENQ MH630 digital projector, and the scene is imaged using a software-controlled Raspberry Pi camera.

A. Quantitative Evaluation

When we train the ensemble of the deep convolutional neural network, the captured images of the three shifted versions of the actual projection patterns are used as the training set. To quantitative evaluate the performance of the depth inference model, we use the captured images of the actual quasi-random pattern projected at various vertical surfaces. The same test is repeated for each subpattern. The results of the experiment are shown in Table II. The table includes the mean square error, in cm, of the depth inference results for each subpattern, and the % of the captured subpatterns that are correctly predicted. It is apparent from Table II that the use of subpatterns with complex designs leads to a significant improvement in the accuracy of the depth inference model, which can be seen from the increase in the inference accuracy and the decrease in the mean square error.

Table II: Quantitative results of the ensembles of deep convolutional neural networks for different subpatterns

Subpattern	Inference accuracy (%)	MAE (cm)
Point	70.28	0.32
Square	77.19	0.26
Cross	79.80	0.21
Triangle	77.33	0.24
X	76.25	0.26

B. Qualitative Evaluation

In order to evaluate the performance of depth inference model with different subpatterns, it is necessary to observe the reconstruction of certain surfaces and analyze from a qualitative point of view. In Fig 4, we illustrate the difference between depth maps generated using different subpatterns. The test surface is a Styrofoam mannequin head of dimensions $30 \times 15 \times 15$ cm, placed at a distance about 36 cm to the camera. The depth inference results are presented as a rendered depth map. The result from qualitative evaluation is consistent with the findings from quantitative tests. It is obvious that subpattern with complex designs enable details of the mannequin head to be distinguished, while the traditional one-pixel subpattern is only able to obtain the basic geometry of the mannequin head. It is interesting to note that the depth reconstruction with square subpattern did not perform as well compare to the other proposed subpatterns. One possible explanation for this result may

be that the square subpattern occupies the most number of pixels among the proposed subpatterns, resulting in extra brightness when being projected. This greatly increases the chance of having overlapped burring subpatterns, and can directly lead to erroneous depth inference results as shown in the black and white spots in the reconstructed depth map.

V. Conclusion

In this paper, we present an extension of the quasi-random projection approach to DfD by introducing the use of a new quasi-random projection pattern consisting of complex subpatterns. We leverage an ensemble of deep convolutional neural networks to automatically extract optimal features in complex subpatterns, leading to improved fidelity of the 3D reconstruction result than previous implementations with point subpatterns. Experimental results using quasi-random projection patterns composed of a variety of non-conventional subpattern designs on complex surfaces showed that the use of complex subpatterns in the quasi-random projection pattern can significantly improve depth reconstruction quality compared to a point pattern.

Acknowledgment

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

References

[1] J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern recognition*, vol. 37, no. 4, pp. 827–849, 2004.

[2] 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. IEEE, 2003, pp. I–I.

[3] A. Ma, A. Wong, and D. Clausi, "Depth from defocus via active multispectral quasi-random point projections using deep learning," *Journal of Computational Vision and Imaging Systems*, vol. 3, no. 1, 2017.

[4] A. Pentland, S. Scherock, T. Darrell, and B. Girod, "Simple range cameras based on focal error," *JOSA A*, vol. 11, no. 11, pp. 2925–2934, 1994.

[5] O. Ghita, P. F. Whelan, and J. Mallon, "Computational approach for depth from defocus," *Journal of Electronic Imaging*, vol. 14, no. 2, pp. 023 021–023 021, 2005.

[6] F. Moreno-Noguer, P. N. Belhumeur, and S. K. Nayar, "Active refocusing of images and videos," *ACM Transactions On Graphics (TOG)*, vol. 26, no. 3, p. 67, 2007.

[7] A. Ma, F. Li, and A. Wong, "Depth from defocus via active quasi-random point projections," *Journal of Computational Vision and Imaging Systems*, vol. 2, no. 1, 2016.

[8] A. Ma, A. Wong, and D. Clausi, "Depth from defocus via active quasi-random point projections: A deep learning approach," in *Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings*, vol. 10317. Springer, 2017, p. 35.

[9] A. Ma and A. Wong, "Enhanced depth from defocus via active quasi-random colored point projections," in *9th International Conference on Inverse Problem in Engineering*, 2017.

[10] A. Mennucci and S. Soatto, "On observing shape from defocused images," in *Image Analysis and Processing, 1999. Proceedings. International Conference on*. IEEE, 1999, pp. 550–555.

[11] R. Bridson, "Fast poisson disk sampling in arbitrary dimensions," in *ACM SIGGRAPH 2007 sketches*. ACM, 2007, p. 22.

[12] H. Niederreiter, "Point sets and sequences with small discrepancy," *Monatshefte für Mathematik*, vol. 104, no. 4, pp. 273–337, 1987.

[13] 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*. IEEE, 1991, pp. 325–327.