# Rectification Based Single-Shot Structured Light for Accurate and Dense 3D Reconstruction

Sina Farsangi      Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Mohamed A. Naiel      Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Mark Lamm      Christie Digital Systems Canada Inc., Kitchener, ON, Canada
Paul Fieguth      Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Email:{sfarsangi, mohamed.naiel, paul.fieguth}@uwaterloo.ca,    mark.lamm@christiedigital.com

## Abstract

Structured Light (SL) patterns generated based on pseudo-random arrays are widely used for single-shot 3D reconstruction using projector-camera systems. These SL images consist of a set of tags with different appearances, where these patterns will be projected on a target surface, then captured by a camera and decoded. The precision of localizing these tags from captured camera images affects the quality of the pixel-correspondences between the projector and the camera, and consequently that of the derived 3D shape. In this paper, we incorporate a quadrilateral representation for the detected SL tags that allows the construction of robust and accurate pixel-correspondences and the application of a spatial rectification module that leads to high tag classification accuracy. When applying the proposed method to single-shot 3D reconstruction, we show the effectiveness of this method over a baseline in estimating denser and more accurate 3D point-clouds.

## 1 Introduction

Structured Light (SL) is one of the most widely used methods for 3D reconstruction. A SL-based system usually consists of a projector and camera, where SL image(s) is(are) projected onto the surface that its shape is required to be reconstructed. The purpose of using a SL image is to encode the projector space in some unique or robustly interpretable way [1]. SL encoding methods can be divided into multi-shot [2, 3] and single-shot [4–7] methods, where the locations of the projector pixels are encoded using a sequence of images and a single image, respectively, in order to allow finding the correspondences between the projector and camera pixels.

Although multi-shot SL methods offer dense pixel correspondences, they cannot be used for dynamic (time-varying) surfaces, cameras and/or projectors [1, 8], motivating a variety of single-shot SL methods to address this problem. Among single-shot SL approaches, most widely used have been those based on pseudo-random arrays [4, 6, 7], where tags with different textures are embedded in the SL image using non-repeated random codewords to ensure the uniqueness of every overlapping block of tags.

An important step with this single-shot SL approach is to extract feature points in the camera and projector SL images along with the codeword decoding to construct the pixel correspondences [4, 6, 7]. The aim of our research is to address how to better detect the feature points of single-shot SL patterns. Having this same goal, the method in [4] utilized the centroids of the embedded tags as feature points, and the work in [6] introduced grid lines between tags and used grid detectors to find the grid intersections in the SL camera image. However, these methods [4, 6] extract only one pixel correspondence per detected tag. In order to have a denser set of correspondences, the method in [7] used pseudo-random arrays with color tags and grid lines to find four corner points per each detected tag by leveraging an iterative process using the Ramer-Douglas-Peucker algorithm [9]. However, colored tags in SL patterns are not necessarily easily perceived when projected onto colored surfaces, and also the method uses hyper-parameters that need to be fine tuned.

In this paper, we tackle the problem of rectifying single-shot SL patterns for more accurate and densely reconstructed 3D point clouds. The proposed method is based on estimating the corner points of quadrilaterals that better represent the input SL tags by applying an iterative optimization process of a simple cost function. The method allows the corner points for each of the tags in the camera image to be found which results in a higher number of pixel correspondences. The extracted points offer more accurate pixel-correspondences, but also allow for tag rectification to achieve improved classification accuracy, resulting in further improvements in precision. The rectification is of particular importance and interest for scenarios involving significant tag skew due to surface geometry.
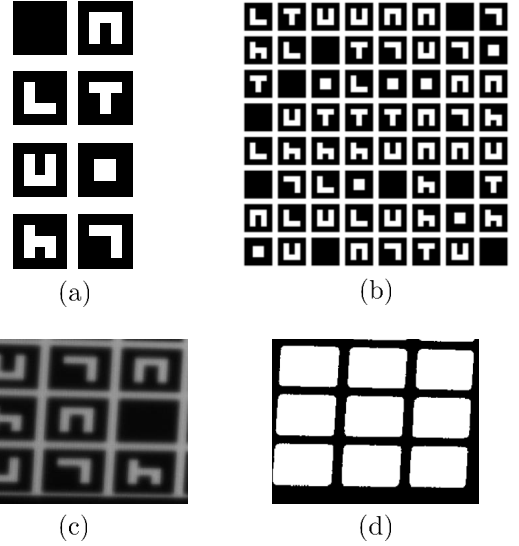


Fig. 1: (a) A set of eight 2D tags, (b) a sample pseudo-random based SL projector image, (c) a sample camera captured SL image, and (d) the mask image for the detected tags using the image in (c).

## 2 Proposed Method

**Overview of the Proposed Method:** In this paper, we use a pseudo-random based SL single-shot image with 8 different tags each of size $20 \times 20$ pixels, gird lines to separate the tags of thickness 4 pixels, and sliding window of size $3 \times 3$ tags as shown in Figures 1(a) and 1(b). After creating the SL image, this image is projected onto the surface and a camera image, $I_C$, is captured. In order to obtain the pixel correspondences between the projector and the camera spaces, first, the global thresholding method is used to threshold the captured image, then a binary connected-components analysis is employed to create the mask image of the detected tags, $I_M$. As an example, a subset of the captured camera image and the mask image are shown in Figures 1(c) and 1(d).

Let us now assume that there are $N_c$ connected components in $I_M$, which indicate the locations of $N_c$ corresponding tags in $I_C$. We aim to find $N_c \times r$ pixel correspondence pairs between the projector and the camera as:

$$\mathbf{x}_{j,k} \leftrightarrow \mathbf{v}_{j,k} \tag{1}$$

where $r$ is the number of detected vertices per tag, and $\mathbf{x}_{j,k} = (x_{j,k}, y_{j,k})$ and $\mathbf{v}_{j,k} = (u_{j,k}, v_{j,k})$ are the coordinates in the camera and projector spaces, respectively, and $k = 1, 2, ..., N_c$ and $j = 1, 2, ..., r$. Obtaining the correspondences in (1) requires (a) estimating the locations of $r$ corner points per detected tag in both the projector and camera spaces, and (b) spatially rectify and classify the detected tags in the camera space and use the decoding scheme to localize the corresponding tags in the projector space, we will focus in this section on the former and in the next section on the latter.

**Corner Points Estimation:** As we are using square tags in SL image construction, therefore, the number of corners per tag is $r = 4$. The corner points in the projector image $\mathbf{v}_{j,k}$ are known from the SL image construction step. However, for each connected component (tag) in the masked camera image, the corner points, $\mathbf{x}_{j,k}$, are unknown and should be obtained. Also, let $a_k$ be the boundary region of the $k$th tag. The goal is to find a quadrilateral of four vertices $q_k = (\mathbf{x}_{1,k}, \mathbf{x}_{2,k}, \mathbf{x}_{3,k}, \mathbf{x}_{4,k})$ using an iterative optimization approach with $N_T$ iterations, which has the highest overlap with the tag. Starting with a rectangular bounding box $q_{0,k}$, we find $q_k$ that
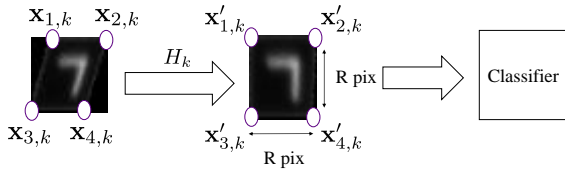
Fig. 2: Tag rectification and its usage for tag classification and classifier training

minimizes one minus the intersection over union [10] between $q_k$ and $a_k$:

$$\min_{q_k} \quad 1 - \frac{\text{Area}(a_k \cap q_k)}{\text{Area}(a_k \cup q_k)} \quad (2)$$

where $q_k = \{\mathbf{x}_{j,k} \in [Q_U \cap Q_L], j = 1, 2, ..., 4\}$, and $Q_U$ and $Q_L$ are the convex sets generated by the region of the initial rectangular box, $q_{0,k}$, and the $k$th boundary region $a_k$, respectively. In order to optimize the cost function in (2), we use the interior-point optimizer [11]. By performing this optimization for each connected component, the corner points in the camera image will be obtained.

**Tag Rectification and Classification:** After obtaining the optimal corner points in the camera image for the $k$th tag, $q_k$, the tags should be rectified and then a classifier can be constructed to infer the class label of the rectified tag. Assuming the input image patches for the classifier should have a size $R \times R$ pixels, we use the obtained corner points in Section 2 to rectify the tags as follows. Given the optimal corner points of the $k$th tag in the camera image, $q_k = (\mathbf{x}_{1,k}, \mathbf{x}_{2,k}, \mathbf{x}_{3,k}, \mathbf{x}_{4,k})$, we find a projective transformation that can rectify each tag into an image patch of size $R \times R$ pixels. Assuming that each tag is projected on a relatively planar surface, the relationship between corner points of the original and rectified tags vertices can be written as:

$$\begin{bmatrix} \mathbf{x}'_{1,k} & \mathbf{x}'_{2,k} & \mathbf{x}'_{3,k} & \mathbf{x}'_{4,k} \\ 1 & 1 & 1 & 1 \end{bmatrix} = H_k \begin{bmatrix} \mathbf{x}_{1,k} & \mathbf{x}_{2,k} & \mathbf{x}_{3,k} & \mathbf{x}_{4,k} \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (3)$$

where $H_k$ is the transformation matrix for the $k$th tag and $\mathbf{x}'_{1,k}, \mathbf{x}'_{2,k}, \mathbf{x}'_{3,k}$ and $\mathbf{x}'_{4,k}$ are the corner points of the rectified tag that can be represented as $\mathbf{x}'_{1,k} = [0,0]^\top$, $\mathbf{x}'_{2,k} = [R,0]^\top$, $\mathbf{x}'_{3,k} = [R,R]^\top$ and $\mathbf{x}'_{4,k} = [0,R]^\top$. By applying $H_k$ to the $k$th tag, the rectified tag is obtained. During the training phase, different tags are projected on the target surface and camera images are captured. Next, training image patches, each of size $(R \times R)$, are obtained by detecting the tags, finding the corner points and the transformations for each tag and rectifying them. Then, a support-vector machine (SVM) classifier with a linear kernel is trained using gradient features [12] extracted from the rectified tags. At the test phase, given a single-shot SL, tags are detected and corner points are found using (2), the tags are rectified (3) and the trained classifier is used for tag classification as shown in Figure 2.

## 3 Experimental Results

In this section, an experimental evaluation for using the proposed method for 3D reconstruction is presented. In our experiments, we have used a Point Grey Flea camera with a resolution of $2448 \times 2048$ pixels and a Christie DWX600-G projector with a resolution of $1280 \times 800$ pixels. The experiments are performed by utilizing a pseudo-random SL image with an alphabet size of 8 and the codewords are decoded as mentioned in [4]. We perform our experiments using two surfaces, namely, the *Curve* [13] and *Zigzag* surfaces. We compare the accuracy of the proposed method that uses quadrilateral tag representation with $N_T = 1000$ iterations to that of a baseline method which utilizes rectangle vertices around tags as pixel correspondences.

**Quantitative Results:** Considering the CAD model of a target surface as a ground-truth, the root mean square error (RMSE) value between the obtained 3D point cloud using the proposed or baseline method and that of the CAD model is used for quantitative evaluation. To study the effect of the pixel correspondences quality of the 3D reconstruction process, we calculate the 3D point-clouds for the methods in consideration using pre-calibrated parameters obtained by the state-of-the-art method in [13], and another time by directly solving for the projector-camera calibration parameters using the obtained pixel correspondences and the parameter estimation

Table 1: Comparison of the RMSE (*mm*) between the proposed and the baseline methods on the *Curve* and *Zigzag* shapes, where boldface denotes the best results.

| Surface Name | Method Name | RMSE (*mm*) | |
|---|---|---|---|
| | | Pre-calibrated | Self-calibrated |
| *Curve* | Baseline | 6.91 | 22.84 |
| | *Proposed* | **6.80** | **12.15** |
| *ZigZag* | Baseline | 6.21 | 33.18 |
| | *Proposed* | **6.11** | **18.80** |

method in [13], we call the latter as a self-calibrated approach. As it is seen in Table 1, the proposed method offers much lower RMSE than that of the baseline technique, specially in the self-calibration case. This is due to the proposed method, which leverages quadrilateral representations for the detected tags, is able to offer more accurate pixel-correspondences between the projector and camera images that is one of the main factors that affects the quality of a 3D reconstructed shape [14].

The current computational complexity of the proposed scheme for obtaining the pixel correspondences using the proposed method is $\mathcal{O}(N_c N_T)$ and for the baseline method is $\mathcal{O}(1)$. Therefore, in future work we would like to investigate more in how to reduce the computational cost of the proposed method to be more suitable for real-time applications.

**Qualitative Results:** Figure 3 shows a sample captured camera image, $I_C$ and its corresponding masked image, $I_M$ for the two considered surfaces. In addition, it shows the obtained rectangular and quadrilateral representations generated by the baseline and proposed methods, respectively. It can be seen that the proposed method is able to extract the tag corners for quadrilateral shapes that represent better the detected tags than that of the baseline rectangular one.

The 3D reconstructed point clouds of the two target surfaces using the proposed method as well as the baseline method are illustrated in Figure 4. It can be seen that the resulting point clouds obtained using the proposed method are more similar to the CAD model. Unlike the methods in [4, 6] that extract one corner point per tag, the proposed method is able to extract accurately 4 corner points per tag resulting in pixel-correspondences and 3D point clouds that are 4 times denser than those offered by the methods in [4, 6].

## 4 Conclusion and Future Work

In this paper, we have presented a new method to find accurate pixel correspondences between the projector and the camera in a single-shot SL system using a simple cost function and with automatic parameter tuning. This is an important factor since the accuracies in calibration and 3D reconstruction are affected by the quality of these correspondences. Furthermore, the extracted pixel correspondences have been used for tag rectification and classification, where the tags are deformed due to the surface geometry. The qualitative and quantitative results show that the proposed method results in an accurate and dense set of correspondences and 3D point clouds in comparison with the baseline, showing the importance of the proposed method.

In future, we would like to test the current spatial rectification method on more challenging surfaces where tags are more deformed due to the surface geometry. In addition, although the proposed scheme results in more accurate pixel correspondences, there is still a room for improving the accuracy of the 3D reconstructed point cloud. This is due to the inaccuracy coming from the projector-camera self-calibration process. Therefore, as a future work, we would like to develop more accurate single-shot self-calibration methods.
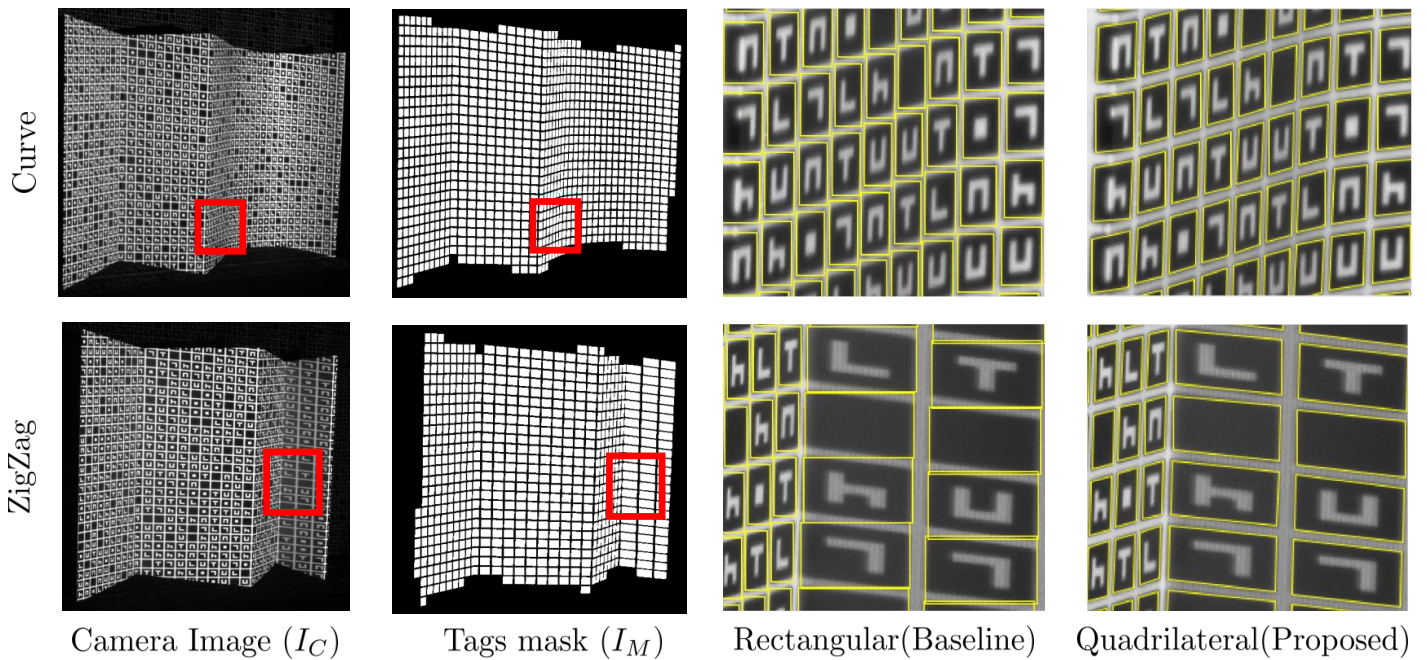
## Acknowledgments

**Fig. 3:** The captured camera images, $I_C$, and the tags mask, $I_M$, from the projected structured light on two different surfaces, *Curve* and *Zigzag*, alongside with the two representations for detected tags, namely, the baseline and proposed methods that use rectangular and quadrilateral representations, respectively.
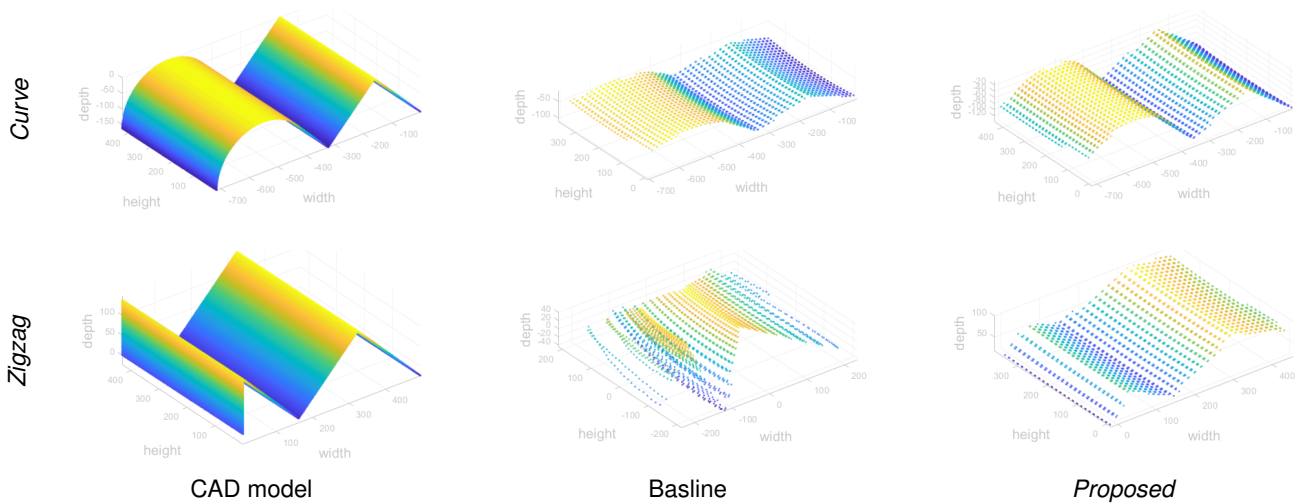


**Fig. 4:** Qualitative comparison for the point-cloud generated by the proposed scheme and the baseline, where *Curve* and *Zigzag* surfaces are used. It is clear that the proposed scheme is able to follow the shape information better than the rectangular representation.

# References

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

[2] J. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," *Comput. Graphics and Image Process.*, vol. 18, no. 1, pp. 1 – 17, 1982.

[3] S. Inokuchi, "Range-imaging system for 3D object recognition," in *Proc. 7th Int. Conf. Pattern Recognition*, pp. 806–808, 1984.

[4] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissano, "Structured light using pseudorandom codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 322–327, 1998.

[5] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki, "Dense 3D reconstruction method using a single pattern for fast moving object," in *Proc. IEEE Int. Conf. on Comput. Vision (ICCV)*, pp. 1779–1786, 2009.

[6] S. Tang, X. Zhang, Z. Song, L. Song, and H. Zeng, "Robust pattern decoding in shape-coded structured light," *Optics and Lasers in Eng.*, vol. 96, pp. 50 – 62, 2017.

[7] K. Yang, Z. Ling, J. Li, X. Gao, L. Xie, and Z. Bai, "Color M-array shape reconstruction of using grid points and center points," in *Proc. Int. Conf. on Inform. Optics and Photonics (CIOP)*, vol. 11209, pp. 659 – 668, 2019.

[8] J. Geng, "Structured-light 3D surface imaging: A tutorial," *Adv. Opt. Photon.*, vol. 3, no. 2, pp. 128–160, 2011.

[9] D. H. Douglas and T. K. Peicker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.

[10] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *ISVC*, 2016.

[11] R. Byrd, J. Gilbert, and J. Nocedal *Mathematical Programming*, vol. 89, pp. 149–185, Jan. 2000.

[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995.

[13] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. A. Clausi, J. S. Zelek, and A. Wong, "Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping," *IEEE Trans. on Comput. Imaging*, vol. 3, pp. 74–83, 2017.

[14] P. Rachakonda, B. Muralikrishnan, and D. Sawyer, "Sources of errors in structured light 3D scanners," in *Dimensional Optical Metrology and Inspection for Practical Applications VIII*, vol. 10991, pp. 25 – 37, SPIE, 2019.