Constraints for Time-Multiplexed Structured Light with a Hand-held Camera

Sepehr Ghavam — Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada
Matthew Post — Christie Digital Systems Canada Inc., Kitchener, 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: {s2ghavam, mohamed.naiel, pfieguth}@uwaterloo.ca, {matthew.post, mark.lamm}@christiedigital.com

Abstract

Multi-frame structured light in projector-camera systems affords high-density and non-contact methods of 3D surface reconstruction. However, they have strict setup constraints which can become expensive and time-consuming. Here, we investigate the conditions under which a projective homography can be used to compensate for small perturbations in pose caused by a hand-held camera. We synthesize data using a pinhole camera model and use it to determine the average 2D reprojection error per point correspondence. This error map is grouped into regions with specified upper-bounds to classify which regions produce sufficiently minimal error to be considered feasible for a structured-light projector-camera system with a hand-held camera. Empirical results demonstrate that a sub-pixel reprojection accuracy is achievable with a feasible geometric constraints.

1 Introduction

Structured light (SL) profilometry has been used in projector-camera systems as it offers high-density, high accuracy and non-contact 3D mapping. It has been used extensively in applications of projection mapping and to correctly display media content on 3D surfaces [1–5]. However, the integrity of this method is predicated on static equipment and, to a lesser extent, a static scene.

Introducing a hand-held camera to the projector-camera system offers a great deal of flexibility, however a compensation mechanism is required to correctly register each image so that the time-multiplexed SL pattern can be correctly decoded. This is a difficult requirement, as banded SL patterns do not have the necessary spatial features to correctly register to one-another between frames, and are subject to the aperture problem.

A simple way to tackle the camera motion is to use a single-shot spatially-encoded pattern to generate point correspondences [6–8]. These methods offer robustness in the face of a dynamic environment [9], however such approaches significantly reduce the number of correspondences that the system is able to attain, with a commensurate reduction in spatial resolution.

On the other hand, using time-multiplexed SL as described in [5, 10], the maximum number of achievable correspondences is equivalent to the number of projector pixels. In the application of projection mapping, and the projection of content on irregular geometry, a greater number of correspondences is almost always better.

The desired outcome of this paper is to study the robustness of spatially-encoded SL patterns that allow the use of a hand-held camera, while retaining the correspondence density of time-multiplexed SL methods.

2 Background

2.1 Gray Code

The emphasis of this paper is to assess the ability of a projective homography (Section 2.2) to compensate for the motion of an unfixed projector-camera system, while preserving the point cloud density afforded by capturing multiple frames of structured light. Given that the application requires pixel-level accuracy for projector calibration, a temporally encoded address using structured light (Gray code, in this case) is more appropriate. An example of Gray code SL pattern can be seen in Figures 1a-1c. Note that due to the banding of the Gray code patterns, there is an absence of features with which to register frames to one-another.

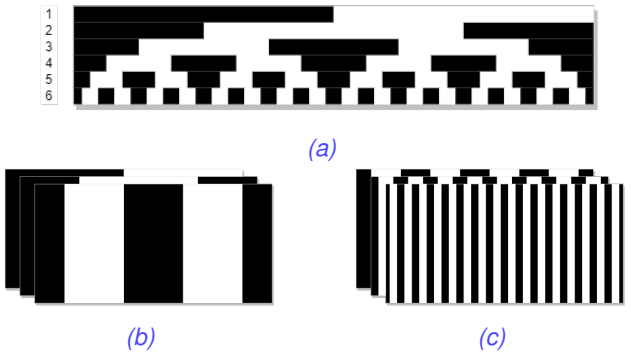


Fig. 1: (a) Overview of multi-frame SL pattern banding for 6-bit Gray code. Visualization for gray code based SL frames number 1 to 3 and 4 to 6 are shown in (b) and (c), respectively.

2.2 Projection Model

A camera projection model [11] is defined as the projection of 3D world points to the camera image plane

$$\vec{x} = \mathbf{K}\mathbf{P}\vec{X} \tag{1}$$

where the vector $\vec{X}$ represents some homogeneous 3D world coordinate, vector $\vec{x}$ represents the projection of world coordinate $\vec{X}$ on to the camera image plane π_c, $\mathbf{K}$ represents the camera intrinsic parameters, and $\mathbf{P}$ represents the augmented camera pose all of which are defined below:

$$\vec{X} = \begin{bmatrix} X_1 & X_2 & X_3 & 1 \end{bmatrix}^\top \quad \vec{X} \in \mathbb{R}^4 \tag{2}$$

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & 1 \end{bmatrix}^\top \quad \vec{x}_i \in \pi_c \tag{3}$$

$$K = \begin{bmatrix} f_x & \phi & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \tag{4}$$

where f_x and f_y are the camera focal lengths of the in the x and y directions, respectively, ϕ is the skew, and c_x and c_y are the coordinates of the camera principal point. To improve the constraints of the calibration, the focal lengths in x and y will be assumed equal, and the pixel skew will be assumed negligible, which leaves us with

$$K = \begin{bmatrix} f & & c_x \\ & f & c_y \\ & & 1 \end{bmatrix} \tag{5}$$

Finally, the camera pose $\mathbf{P}$, used in (1), is defined as an augmented $\begin{bmatrix} 3 \times 4 \end{bmatrix}$ matrix, containing the rotation matrix $\mathbf{R}$ and translation vector $\vec{t}$ as:

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} \mid \vec{t} \end{bmatrix}, \quad \mathbf{R} \in \mathbf{SO(3)}, \ \vec{t} \in \mathbb{R}^3 \tag{6}$$

$$\mathbf{P} = \left[\begin{array}{ccc|c} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{array}\right] \tag{7}$$

2.3 Projective Transform

For any set of image pairs, there exists a 3×3 projective homography $\mathbf{H}$ (with 8 degrees of freedom) that maps all points between any 2 camera views defined by $\mathbf{P}_1$ and $\mathbf{P}_2$ if the change in pose is rotational only, i.e. $\vec{t} = \vec{0}$ [11]:

$$\mathbf{P}_1 = \begin{bmatrix} \mathbf{I} \mid \vec{0} \end{bmatrix} \quad \mathbf{P}_2 = \begin{bmatrix} \mathbf{R} \mid \vec{0} \end{bmatrix} \tag{8}$$

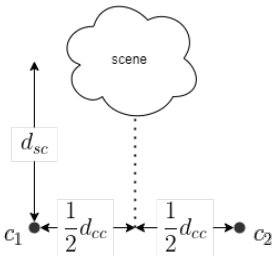


Fig. 2: The setup to generate synthetic data. The distance between c_1 and c_2 is applied on the xz plane, symmetric about the z-axis

$$\mathbf{H} : \vec{x}_1 \rightarrow \vec{x}_2 \tag{9}$$

Then the homography $\mathbf{H}$ offers a perfect mapping between $\mathbf{P}_1$ and $\mathbf{P}_2$

$$\vec{x}_1 = \mathbf{H}\vec{x}_2 \tag{10}$$

Since the homography transform is of size 3×3, $\vec{x}_i$ must be homogeneous coordinates.

3 Problem Formulation

If the camera motion could be locked simply to rotational motion, then eq. (10) would solve our dilemma. However, since such a steady hand is not a feasible assumption, we have to take into account some non-trivial translation between each pose of a series of captured images. This changes (10) to an approximation

$$\vec{\hat{x}}_1 = \mathbf{H}\vec{x}_2 \tag{11}$$

3.1 Reprojection Error

Since the 3D coordinates of the scene are known, the average 2D reprojection error, e, for each point in a scene S

$$S = \{\vec{X}_i\} \tag{12}$$

can be generated for each pose

$$e = \frac{1}{N}\sum_{i=1}^{N} \|\vec{\hat{x}}_i - \vec{x}_i\|, \quad N = |S| \tag{13}$$

where $\|\cdot\|$ denotes the euclidean distance, and $\vec{\hat{x}}_i$ is the estimated position of the point correspondence at index i given by

$$\vec{\hat{x}}_i = \mathbf{H}\vec{x}'_i \tag{14}$$

The error in (13) is expected to be directly proportional to d_{cc} and inversely proportional to d_{sc}

$$e \propto d_{cc}, \frac{1}{d_{sc}} \tag{15}$$

4 Methodology

To map the reprojection error of the projective transform, a pair of synthetic cameras are generated, with the same intrinsics, a focal length $f_x = f_y$ and the same principal point coordinates $c_{x_1} = c_{x_2}$ and $c_{y_1} = c_{y_2}$. These cameras are assigned poses defined by a distance to the scene, d_{sc}, and the distance between the cameras, d_{cc}. The locations of c_1 and c_2 are defined as

$$\tau_1 = \begin{bmatrix} -\frac{1}{2}d_{cc} & 0 & -d_{sc} \end{bmatrix}^\top \tag{16}$$

$$\tau_2 = \begin{bmatrix} \frac{1}{2}d_{cc} & 0 & -d_{sc} \end{bmatrix}^\top \tag{17}$$

An illustration of the system geometry can be seen in Figure 2.

A point cloud, denoted by a set of points $\{\vec{X}_i\}$, sampled from an existing mesh, is used to produce point correspondences between the camera image planes. The point cloud is normalized such that the centroid lies on the origin

$$\sum_{i=1}^{N} \vec{X}_i = \vec{0} \tag{18}$$

and scaled uniformly using a scale factor, α, such that the range of $\vec{X}$ is limited along the z-axis

$$X_{3_i} - X_{3_j} \leq 1, \ \forall \vec{X} \tag{19}$$

The above condition allows for the determination of the reprojection error as a function of the depth of the point correspondences. This is important, as the error of the projective transform increases for points that lie further away from the plane formed by the control points defining the projective transform.

The rotation $\mathbf{R}_i$ for each camera is configured such that the principal axis passes through the origin.

The matrix $\mathbf{H}$, with 8 degrees of freedom, is inferred using the Direct Linear Transform (DLT) algorithm described in [11] using 4 selected point correspondences. We opt for 4 points as that minimizes the number of pixels dedicated to compensation, so that the remaining camera pixels can be used to generate dense correspondences.

Each pair of point correspondences (u_i, v_i) and (u'_i, v'_i) adds a set of constraints defined by

$$A_i = \begin{bmatrix} -u_i & -v_i & -1 & 0 & 0 & 0 & u_i u'_i & v_i u'_i & u'_i \\ 0 & 0 & 0 & -u_i & -v_i & -1 & u_i u'_i & v_i u'_i & u'_i \end{bmatrix} \tag{20}$$

Using the minimum number of control points necessary to solve for the system explicitly, therefore

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & A_3 & A_4 \end{bmatrix}^T \tag{21}$$

where $\mathbf{A}$ is the constrains matrix of size 8×9 that can be used to solve for the homography matrix as:

$$\mathbf{A}\vec{h} = \vec{0}, \qquad \|\vec{h}\| = 1 \tag{22}$$

where $\vec{h} \in \mathbb{R}^9$ is a unit vector form of the homography matrix $\mathbf{H}$.

As in [11], the system now is solvable using the Single Value Decomposition.

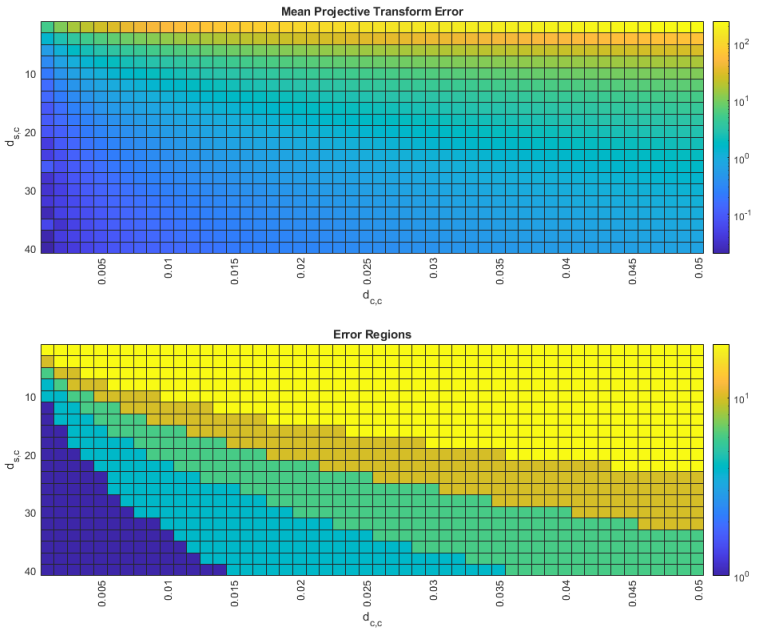
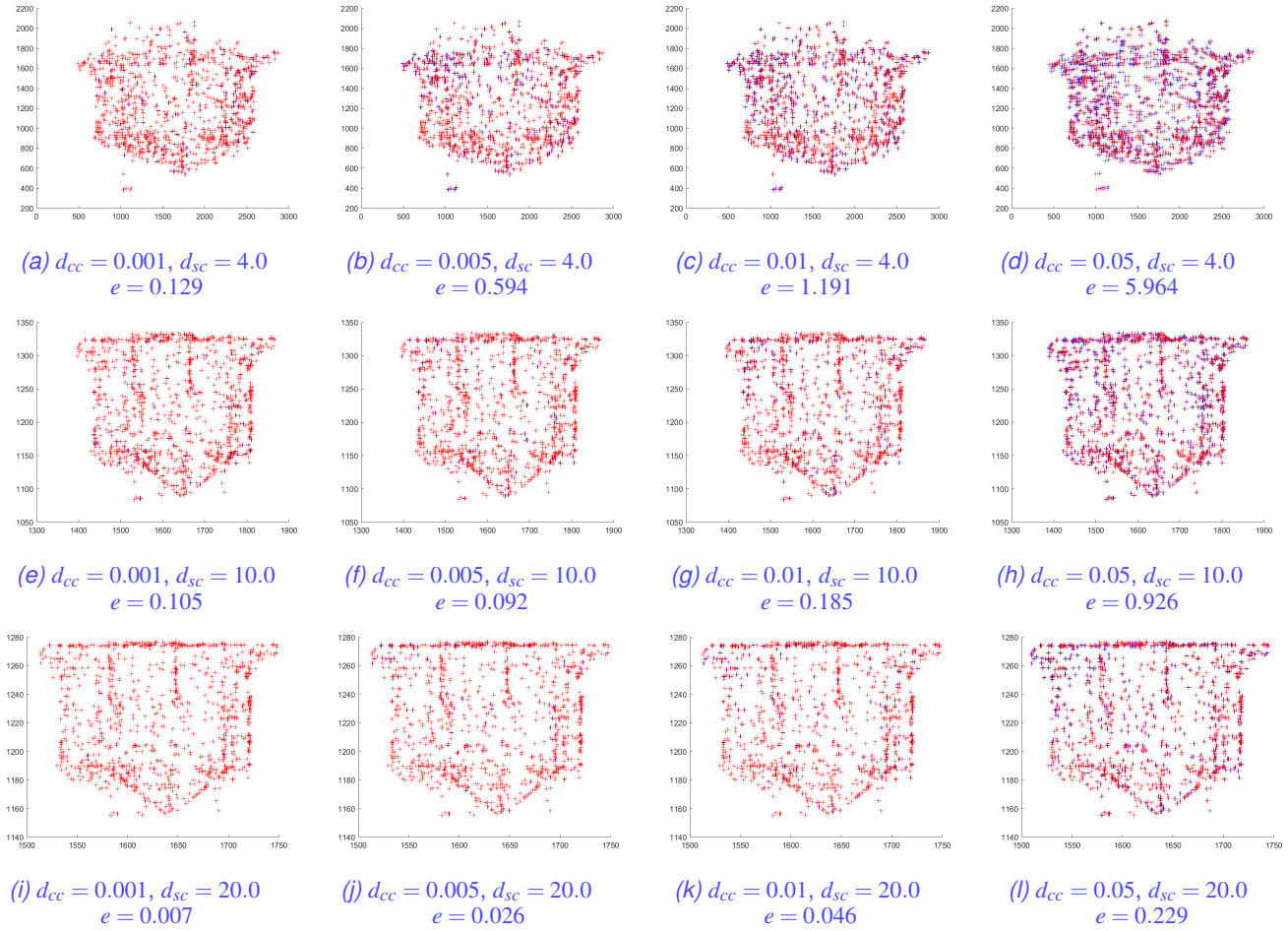


Fig. 3: Top: The average 2D reprojection error in pixels defined by (13) as result of combinations of $d_{s,c}$ and d_{c_1,c_2}. Bottom: error regions. Navy: $0 < e \leq 0.2$; Blue: $0.2 < e \leq 0.5$; Teal: $0.5 < e \leq 1.0$; Gold: $1.0 < e \leq 2.0$; Yellow: $e > 2.0$

5 Results

To evaluate the reprojection error of the projective transform, a pair of synthetic cameras of the same intrinsics are generated, such that



(a) $d_{cc} = 0.001$, $d_{sc} = 4.0$ $e = 0.129$ (b) $d_{cc} = 0.005$, $d_{sc} = 4.0$ $e = 0.594$ (c) $d_{cc} = 0.01$, $d_{sc} = 4.0$ $e = 1.191$ (d) $d_{cc} = 0.05$, $d_{sc} = 4.0$ $e = 5.964$

(e) $d_{cc} = 0.001$, $d_{sc} = 10.0$ $e = 0.105$ (f) $d_{cc} = 0.005$, $d_{sc} = 10.0$ $e = 0.092$ (g) $d_{cc} = 0.01$, $d_{sc} = 10.0$ $e = 0.185$ (h) $d_{cc} = 0.05$, $d_{sc} = 10.0$ $e = 0.926$

(i) $d_{cc} = 0.001$, $d_{sc} = 20.0$ $e = 0.007$ (j) $d_{cc} = 0.005$, $d_{sc} = 20.0$ $e = 0.026$ (k) $d_{cc} = 0.01$, $d_{sc} = 20.0$ $e = 0.046$ (l) $d_{cc} = 0.05$, $d_{sc} = 20.0$ $e = 0.229$

Fig. 4: Examples of projective transforms for various values of d_{cc} and d_{sc}, and their corresponding 2D reprojection error e, where red markers represent $\vec{x}$ and blue represents $\vec{\hat{x}}$.

the focal length $f = 3000$ and the principal point coordinates are $c_x, c_y = (1632, 1224)$. In general, regions shown in Figure 3 indicate that less than 1px reprojection error is possible when constraints are applied to the system geometry. It is important to note that since the point cloud is normalized, the units of d_{cc} and d_{sc} are relative to the size of the scene. Assuming that the observed point cloud has a depth of $1m$, the results shown in Figure 3 imply that at 16m from the scene, the projective transform can compensate camera motions with a magnitude of 5mm with a reprojection error of less than 0.5px.

The error values of the projective transform corrections shown in (13) are shown in Figure 4. For a constant value of d_{cc}, an increase in d_{sc} results in an exponential decrease in reprojection error. It is important to point out that while the reprojection error decreases with an increase in d_{sc}, such an increase has a practical upper bound, particularly for a camera with a fixed focal length. This can be seen in Figure 4, where the footprint of the point correspondences at $d_{sc} = 20$ is barely 250×120. In practical scenarios, a much longer focal length is required to use sufficient camera pixels to develop dense point correspondences in a projector-camera system.

6 Conclusion

The results indicate that there is a set of geometric constraints for a projector camera system that affords a simple corrective process for camera motion compensation. Average reprojection errors per point correspondence of sub-pixel accuracy are achievable for feasible geometric constraints that would not require special equipment or stabilizing hardware. This opens the door to time-multiplexed structured light systems using hand-held cameras, and eventually even mobile phones. Future work in this regard could concentrate on identifying transforms that can improve the system robustness to motion effects.

Acknowledgments

M.A. Naiel and P. Fieguth are supported in part by the Natural Sciences and Engineering Research Council of Canada - Collaborative Research and Development (NSERC-CRD) grant and Christie Digital Systems Inc.

References

[1] M. Brown, A. Majumder, and R. Yang, "Camera-based calibration techniques for seamless multiprojector displays," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 11, no. 2, pp. 193–206, 2005.

[2] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhoefer, "The visual computing of projector-camera systems," *Computer Graphics Forum*, vol. 27, no. 8, pp. 2219–2245, 2008.

[3] B. Sajadi and A. Majumder, "Scalable multi-view registration for multi-projector displays on vertically extruded surfaces," *Computer Graphics Forum*, vol. 29, pp. 1063–1072, 2010.

[4] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers and Graphics*, vol. 37, no. 8, pp. 1012 – 1038, 2013.

[5] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong, "Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 74–83, March 2017.

[6] J. Salvi, J. Batlle, and E. M. Mouaddib, "A robust-coded pattern projection for dynamic 3d scene measurement," *Pattern Recognition Letters*, vol. 19, pp. 1055–1065, 1998.

[7] A. Ben-Hamadou, C. Soussen, C. Daul, W. Blondel, and D. Wolf, “Flexible calibration of structured-light systems projecting point patterns,” *Computer Vision and Image Understanding*, vol. 117, pp. 1468–1481, 2013.

[8] B. Huang, S. Ozdemir, Y. Tang, and H. Ling, “A single-shot camera-projector calibration system for imperfect planar targets,” *ArXiv*, vol. abs/1803.09058, 2018.

[9] F. Zhong, R. Kumar, and C. Quan, “A cost-effective single-shot structured light system for 3d shape measurement,” *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7335–7346, Sep. 2019.

[10] J. Salvi, J. Pagès, and J. Batlle, “Pattern codification strategies in structured light systems,” *Pattern Recognition*, vol. 37, no. 4, pp. 827 – 849, 2004, Agent Based Computer Vision.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. USA: Cambridge University Press, 2003.