Supplemental Document: Joint Calibration of Cameras and Projectors for Multiview Phase Measuring Profilometry

Hyeongjun Cho[1][a] and Min H. Kim[1][b]

[1]*School of Computing, KAIST, Daejeon, South Korea*

{hjcho,minhkim}@vclab.kaist.ac.kr

This supplemental document provides more technical details about camera/projector models and phase extraction in our multiview PMP calibration.

1 Geometry Models

1.1 Camera Model

Intrinsic and extrinsic parameters of camera/projector associating a 3-dimensional point to an image coordinate consist of a camera/projector matrix $\mathbf{K}$ including focal length along x and y axis and an image center u_0 and v_0, and distortion coefficients $\mathbf{d}_c = (k_1, k_2, k_3)$, rotation and translation (Zhang, 2000). Given a point $\mathbf{X}_w = \begin{bmatrix} x_w & y_w & z_w \end{bmatrix}^\intercal$ in the world coordinates can be transformed into 3-dimensional camera coordinates as follows: $\mathbf{X}_c = \begin{bmatrix} x_c & y_c & z_c \end{bmatrix}^\intercal = \mathbf{R}_c \mathbf{X}_w + \mathbf{t}_c$. Then the point is projected to the normalized image plane by

$$\mathbf{x}_c = \begin{bmatrix} x_n & y_n \end{bmatrix}^\intercal = \begin{bmatrix} x_c/z_c & y_c/z_c \end{bmatrix}^\intercal. \quad (1)$$

Finally, for a camera c (or a projector p), the image coordinates are achieved by

$$\hat{\mathbf{p}}_c = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}_c \hat{\mathbf{x}}_c = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \hat{\mathbf{x}}_c, \quad (2)$$

where the hat symbol denotes that a point is expressed in a homogeneous coordinate notation.

However, it cannot be compared to a point in the image directly because the raw image is distorted. The radial distortion removal is performed on the normalized coordinate in our model. A distorted pixel index $\hat{\mathbf{p}}'_c$ is transformed into the normalized coordinate by

$$\hat{\mathbf{x}}'_c = \begin{bmatrix} x'_n & y'_n & 1 \end{bmatrix}^\intercal = \mathbf{K}_c^{-1} \hat{\mathbf{p}}'_c. \quad (3)$$

Then the distortion is removed by

$$\begin{aligned} \hat{\mathbf{p}}_c^{\text{undist}} &= \mathbf{K}_c \begin{bmatrix} x_c & y_c & 1 \end{bmatrix}^\intercal = \mathbf{K}_c \times \text{undist}(\hat{\mathbf{x}}'_c, \mathbf{d}_c) \\ &= \mathbf{K}_c \begin{bmatrix} x'_n(1 + k_1 r'^2 + k_2 r'^4 + k_3 r'^6) \\ y'_n(1 + k_1 r'^2 + k_2 r'^4 + k_3 r'^6) \\ 1 \end{bmatrix}, \end{aligned} \quad (4)$$

where $r' = \sqrt{x'^2_n + y'^2_n}$ (Ricolfe-Viala and Sanchez-Salmeron, 2010).

A telecentric camera uses an orthographic projection model, while a pinhole camera uses a perspective projection model. In other words, the z-component of the camera coordinate system is ignored (Chen et al., 2014). To model orthographic projection into our model, we need to alter $\mathbf{x}_c$ in Eq. (1) to

$$\mathbf{x}_c = \begin{bmatrix} x_n & y_n \end{bmatrix}^\intercal = \begin{bmatrix} x_c & y_c \end{bmatrix}^\intercal. \quad (5)$$

1.2 Projector Model

The projector model is an inverse pinhole camera model in our method, the only difference is that the projector uses phases in horizontal and vertical axes rather than uv-image coordinates. However, the phases have linear relation with uv-coordinates, $\psi_u = w_u u + b_u$ and $\psi_v = w_v v + b_v$.

$$\begin{bmatrix} \psi_u \\ \psi_v \\ 1 \end{bmatrix} = \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \mathbf{K}_p \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix}. \quad (6)$$

Since the relationship between the pixel index of the projector and the phase is unnecessary to our model, we can define a new projector matrix $\mathbf{K}'_p$,

[a] https://orcid.org/0000-0001-9399-4232

[b] https://orcid.org/0000-0002-5078-4005

which maps normalized coordinates to phases.

$$\begin{aligned}\mathbf{K}'_p &= \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \mathbf{K}_p \\ &= \begin{bmatrix} w_u & 0 & b_u \\ 0 & w_v & b_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} w_u f_x & 0 & w_x u_0 + b_u \\ 0 & w_v f_y & w_y v_0 + b_v \\ 0 & 0 & 1 \end{bmatrix}.\end{aligned} \tag{7}$$

2 Phase Extraction

A sequence of N sinusoidal structured light images with a frequency f encodes a phase, which is linearly related to the pixel index of a projector in vertical or horizontal direction (Feng et al., 2021). For $k \in [0, N-1]$, when the n-th structured light image with frequency f is projected onto an object, the reflected light is captured by a camera with the intensity of $\mathbf{I}_{fk}(u,v)$ for each pixel u,v.

$$\mathbf{I}_{fk}(u,v) = \mathbf{I}_a(u,v) + \mathbf{I}_b(u,v)\cos(\phi_f(u,v) + 2\pi k/N), \tag{8}$$

where $\phi_f(u,v) \in [-f\pi, f\pi]$. In this case, we can easily reconstruct the phase ϕ by

$$\phi_f(u,v) = \tan^{-1} \frac{\sum_{k=0}^{N-1} I_{fk}(u,v)\sin(2\pi k/N)}{\sum_{k=0}^{N-1} I_{fk}(u,v)\cos(2\pi k/N)}. \tag{9}$$

However, due to the limitation of the dynamic range of the projector, we need to vary the frequency of structured light to precisely measure the phase. In other words, a sequence of wrapped phase $\phi_f(u,v)$ should be unwrapped to obtain $\psi(u,v)$. We build the structured light images by recursively multiplying a scale factor $\alpha > 1$ to the frequency f to produce n structured lights.

Given wrapped phases ϕ_f where $f = \alpha^0, \cdots, \alpha^{n-1}$, where n is the number of frequencies used for phase unwrapping, we recursively unwrap the phases to produce $\psi_{\alpha f}$ with ψ_f and $\phi_{\alpha f}$. At the first, $\psi_1(u,v) = \phi_1(u,v)$. The final unwrapped phase $\psi(u,v)$ is equal to $\psi_f(u,v)$ where $f = \alpha^{n-1}$. Unwrapped phase $\psi_f(u,v)$ and wrapped phase $\phi_f(u,v)$ has the relationship as follows:

$$\psi_f(u,v) = \phi_f(u,v) + 2\pi h_f(u,v). \tag{10}$$

where h is a fringe order, which is an integer at frequency f. Given ψ_f and $\phi_{\alpha f}$, the fringe order at frequency αf can be estimated by approximating $\psi_{\alpha f} \simeq \alpha \psi_f$,

$$h_{\alpha f}(u,v) = \mathrm{round}\left(\frac{\alpha \psi_f(u,v) - \phi_{\alpha f}(u,v)}{2\pi} \right). \tag{11}$$

Then the unwrapped phase with frequency αf is obtained via Eq. (10) (Juarez-Salazar et al., 2019). In practice, we choose the scale factor $\alpha = 4$ and $n = 4$, i.e., we capture images with structured lights of frequencies 1, 4, 16, and 64.

2.1 Phase Outlier Removal

Since a phase map produced by a series of images is susceptible to the noise of the images, a minor noise occurring to a pixel of a single image can bring about serious errors in 3D points. Therefore, we must filter out the noisy phases from the phase map to reconstruct the correct shape.

First of all, a dark reflectance surface may exist in an object, which means the phase is not correctly extracted since light emitted by the projector does not reach enough of some part of the object. For instance, a shadowed or tilted part of the projector does not have a proper phase. In this case, we can filter out those parts by computing the standard deviation of a series of images. A small standard deviation value denotes that a phase is not properly extracted. In Eq. (8), $\mathbf{I}_b(u,v)$ is related to the brightness of an object lighted up by a projector. Therefore, it is obvious that the light emitted by the projector is not enough to arrive at the object if the value of $\mathbf{I}_b(u,v)$ is too small. Since $N \times n$ images are given, the mean pixel intensity at (u,v) is

$$\mu(u,v) = \frac{1}{nN} \sum_f \sum_{k=0}^{N-1} \mathbf{I}_{fk}(u,v) = \mathbf{I}_a(u,v), \tag{12}$$

where $f = \alpha^0, \cdots, \alpha^{n-1}$. Then the variance per pixel of $N \times n$ images is

$$\begin{aligned}\mathrm{Var}(u,v) &= \frac{1}{nN} \sum_f \sum_{k=0}^{N-1} \{\mathbf{I}_{fk}(u,v) - \mu(u,v)\}^2 \\ &= \frac{1}{nN} \sum_f \sum_{k=0}^{N-1} \{\mathbf{I}_b(u,v)\cos(\phi_f(u,v) + 2\pi k/N)\}^2.\end{aligned} \tag{13}$$

Since we choose $N = 4$, we can convert $\cos(\phi_f(u,v) + 2\pi k/N + \pi/2) = -\sin(\phi_f(u,v) + 2\pi k/N)$ for $k = 0, 2$. Then the standard deviation per pixel of nN images are,

$$\begin{aligned}\sigma(u,v) &= \sqrt{\frac{1}{4n} \sum_f \{2 \times \mathbf{I}_b^2(u,v)\}} \\ &= \sqrt{\frac{\mathbf{I}_b^2(u,v)}{2}} = \frac{\mathbf{I}_b(u,v)}{\sqrt{2}}.\end{aligned} \tag{14}$$

We filter out the pixels that $\sigma(u,v) < \tau_{\mathrm{std}}$.

Moreover, we additionally remove the phases that break local consistency. Since the phases are geometrically related to the projector, each pixel in the phase map has a value close to its neighborhoods. Therefore, a pixel with a phase that is very far from its neighborhood is supposed to be an outlier due to the sensor noise. Let $\mu_{\text{local}}(u,v)$ and $\sigma_{\text{local}}(u,v)$ denote a local mean and standard deviation of pixel (u,v) and its neighborhoods, each pixel should satisfy

$$|\psi(u,v) - \mu_{\text{local}}(u,v)| < \tau_{\text{local}}\sigma_{\text{local}}(u,v). \tag{15}$$

The value of τ_{local} must be generous not to filter out geometric non-continuities of an object.

REFERENCES

Chen, Z., Liao, H., and Zhang, X. (2014). Telecentric stereo micro-vision system: Calibration method and experiments. *Optics and Lasers in Engineering*, 57:82–92.

Feng, S., Zuo, C., Zhang, L., Tao, T., Hu, Y., Yin, W., Qian, J., and Chen, Q. (2021). Calibration of fringe projection profilometry: A comparative review. *Optics and Lasers in Engineering*, 143:106622.

Juarez-Salazar, R., Giron, A., Zheng, J., and Diaz-Ramirez, V. H. (2019). Key concepts for phase-to-coordinate conversion in fringe projection systems. *Applied optics*, 58(18):4828–4834.

Ricolfe-Viala, C. and Sanchez-Salmeron, A.-J. (2010). Lens distortion models evaluation. *Applied optics*, 49(30):5914–5928.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.