Supplementary Technical Report for the Article: Blocks-World Cameras

Jongho Lee Mohit Gupta
University of Wisconsin-Madison
{jongho, mohitg}@cs.wisc.edu

1. Overview

This document provides derivations, explanations, and more results supporting the content of the paper submission titled, "Blocks-World Cameras".

2. Geometric Relationship between Pattern Feature and Image Feature

In this section, we algebraically derive geometric relationship between a pattern feature and a corresponding image feature, for given scene plane parameters. Based on this relationship, we can estimate plane parameters given a pair of image feature and pattern feature. Toward that end, we first review mathematical preliminaries in Section 3 of the main manuscript for completeness.

Two-view geometry of structured-light: The Blocks-World Camera consists of a projector and a camera, as shown in Fig. 1 (a). We define the camera and projector coordinate systems (CCS and PCS) centered at $\mathbf{c}_c$ and $\mathbf{c}_p$, the optical centers of the camera and the projector, respectively. $\mathbf{c}_c$ and $\mathbf{c}_p$ are separated by the projector-camera baseline b along the x axis. The world coordinate system (WCS) is assumed to be the same as the CCS centered at $\mathbf{c}_c$, i.e., $\mathbf{c}_c = [0,0,0]^T$ and $\mathbf{c}_p = [b,0,0]^T$ in the WCS. Without loss of generality, both the camera and the projector are assumed to have the same focal length f, i.e., the image planes of both are located at a distance f from their optical centers along the z-axis. For simplicity, we further assume a rectified system such that the epipolar lines are along the rows of the camera image and the projector pattern.

Plane parameterization: A 3D scene plane is characterized by $\Pi = \{D, \theta, \varphi\}$, where $D \in [0, \infty)$ is the perpendicular (shortest) distance from the origin ($\mathbf{c}_c$) to Π, $\theta \in [0, \pi]$ is the polar angle between the plane normal and the $-z$ axis, and $\varphi \in [0, 2\pi)$ is the azimuthal angle between the plane normal and the x axis (measured clockwise), as shown in Fig. 1 (a). The pattern consists of a sparse set of cross-shaped features, which get mapped to cross-shaped features in the camera image via homographies induced by scene planes, as shown in Fig. 1 (b).

Pattern feature and image feature: Consider a pattern feature P described by $\mathrm{P} = \{\mathbf{u}_p, \mathbf{v}_p, \mathbf{p}_p\}$, where $\mathbf{v}_p$ and $\mathbf{u}_p$ are two line vectors and $\mathbf{p}_p$ is the intersection point of $\mathbf{v}_p$ and $\mathbf{u}_p$ as shown in Fig. 1 (b). Let the corresponding image feature I be described by $\mathrm{I} = \{\mathbf{u}_c, \mathbf{v}_c, \mathbf{p}_c\}$, where $\mathbf{v}_c$ and $\mathbf{u}_c$ are line vectors corresponding to $\mathbf{v}_p$ and $\mathbf{u}_p$, respectively, and $\mathbf{p}_c$ is the intersection point of $\mathbf{v}_c$ and $\mathbf{u}_c$. The elements in P and I are described in their own coordinate systems (PCS and CCS, respectively), i.e., for the pattern feature $\mathrm{P} = \{\mathbf{u}_p, \mathbf{v}_p, \mathbf{p}_p\}$,

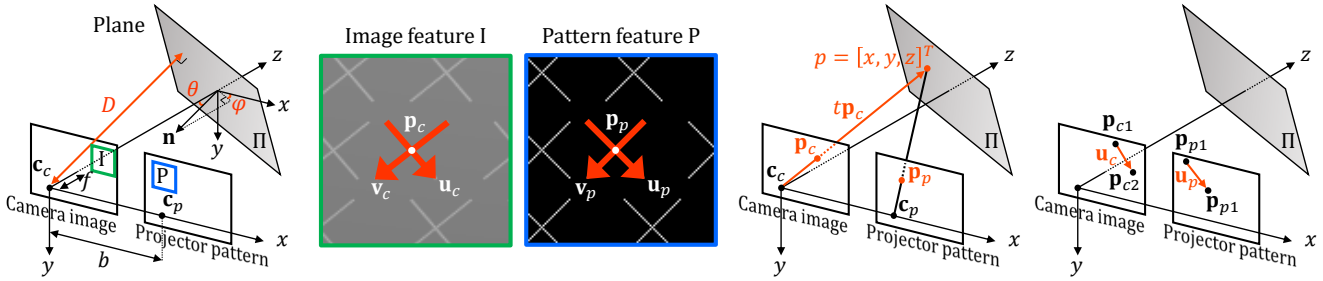
$$\mathbf{u}_p = [u_{px}, u_y, 0]^T, \mathbf{v}_p = [v_{px}, v_y, 0]^T, \mathbf{p}_p = [p_{px}, p_y, f]^T. \tag{1}$$

Similarly, for the corresponding image feature $\mathrm{I} = \{\mathbf{u}_c, \mathbf{v}_c, \mathbf{p}_c\}$,

$$\mathbf{u}_c = [u_{cx}, u_y, 0]^T, \mathbf{v}_c = [v_{cx}, v_y, 0]^T, \mathbf{p}_c = [p_{cx}, p_y, f]^T. \tag{2}$$

Depth estimation: To derive the relationship between the pattern feature and the image feature in terms of the plane parameters, we first derive the scene depth at the intersection point $\mathbf{p}_c$ of the image feature in terms of the plane parameters. Let the ray passing $\mathbf{c}_p$ and $\mathbf{p}_p$ intersect Π at $p = [x, y, z]^T$, and p is imaged at $\mathbf{p}_c$ as shown in Fig 1 (c). Let the equation of the line passing $\mathbf{c}_c$ and $\mathbf{p}_c$ be $t\mathbf{p}_c = t\,[p_{cx}, p_y, f]$ $(-\infty < t < \infty)$ as shown in Fig 1 (c). Depth z of $\mathbf{p}_c$ can be obtained by intersecting this line with the plane $\Pi = \{D, \theta, \varphi\}$ and taking the third element of $t\mathbf{p}_c$. By substituting $t\mathbf{p}_c$ to the plane equation, we get $\mathrm{n}^T(t\mathbf{p}_c) + D = 0$ and $t = -\frac{D}{\mathrm{n}^T\mathbf{p}_c}$. Thus,

$$z = tf = -f\frac{D}{\mathrm{n}^T\mathbf{p}_c}. \tag{3}$$



(a) Imaging system (b) A pair of pattern and image features (c) Point correspondence (d) Line correspondence

Figure 1. **Feature correspondences in the Blocks-World Cameras.** (a) The Blocks-World Cameras are based on a structured light system consisting of a projector which projects a single pattern on the scenes and a camera to capture the images. (b) The pattern consists of a sparse set of cross-shaped features, which get mapped to cross-shaped features in the camera image via homographies induced by scene planes. The plane parameters can be estimated by measuring the deformation between these features. (c) To derive the relationship between the pattern feature and the image feature, point correspondence can be established from a triangle defined by $\mathbf{c}_p$, p, and $\mathbf{c}_c$. (d) Similarly, line vector correspondence can be defined from two pairs of point correspondences.

The depth z of $\mathbf{p}_c$ can be also represented in terms of the corresponding pattern intersection point $\mathbf{p}_p$. By intersecting the line passing $\mathbf{c}_p$ and $\mathbf{p}_p$ with the plane Π (Fig 1 (c)):

$$z = -f\frac{\mathrm{n}^T\mathbf{c}_p + D}{\mathrm{n}^T\mathbf{p}_p}. \tag{4}$$

In the following, we describe the image feature $\mathrm{I} = \{\mathbf{u}_c, \mathbf{v}_c, \mathbf{p}_c\}$ in terms of the pattern feature $\mathrm{P} = \{\mathbf{u}_p, \mathbf{v}_p, \mathbf{p}_p\}$ given the plane parameters $\Pi = \{D, \theta, \varphi\}$.

Point correspondence: From a triangle defined by $\mathbf{c}_p$, p, and $\mathbf{p}_p$ (Fig 1 (c)), the imaged point $\mathbf{p}_c$ corresponding to $\mathbf{p}_p$ is:

$$\mathbf{p}_c = \begin{bmatrix} 1 & 0 & \frac{b}{z} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}_p, \tag{5}$$

where z is the depth of $\mathbf{p}_c$. By substituting Eq. 3 or Eq. 4 to Eq. 5, we get the point correspondence in terms of the plane parameters:

$$\mathbf{p}_c = \begin{bmatrix} \frac{D}{D+bn_x} & -\frac{bn_y}{D+bn_x} & -\frac{bn_z}{D+bn_x} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}_p, \tag{6}$$

where $\mathbf{n} = [n_x, n_y, n_z]^T$.

Line vector correspondence: Let $\mathbf{p}_{p1}$ and $\mathbf{p}_{p2}$ be two points defining $\mathbf{u}_p$ as shown in Fig 1 (d). Using Eq. 5, we can define two imaged points $\mathbf{p}_{c1}$ and $\mathbf{p}_{c2}$ corresponding to $\mathbf{p}_{p1}$ and $\mathbf{p}_{p2}$, respectively as shown in Fig 1 (d):

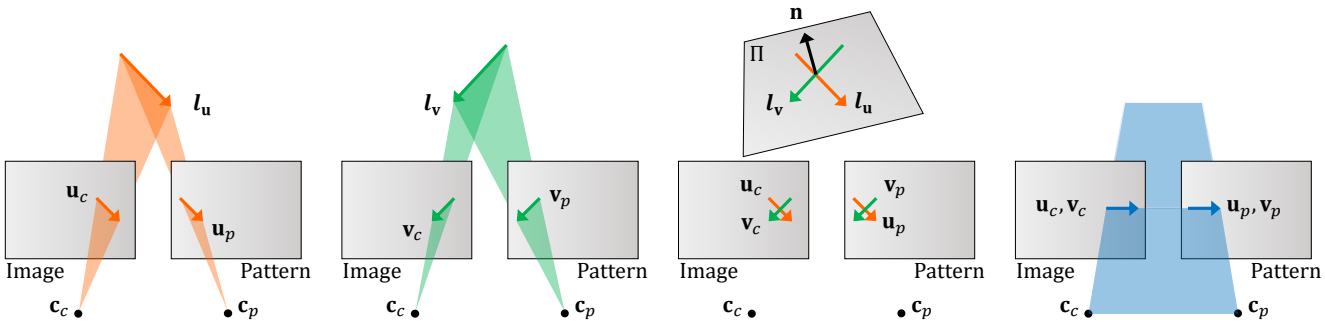
$$\mathbf{p}_{c1} = \begin{bmatrix} 1 & 0 & \frac{b}{z_1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}_{p1} \text{ and } \mathbf{p}_{c2} = \begin{bmatrix} 1 & 0 & \frac{b}{z_2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}_{p2}, \tag{7}$$

where z_1 and z_2 are the depths of $\mathbf{p}_{c1}$ and $\mathbf{p}_{c2}$, respectively. By defining $\mathbf{u}_c = \mathbf{p}_{c2} - \mathbf{p}_{c1}$ and $\mathbf{u}_p = \mathbf{p}_{p2} - \mathbf{p}_{p1}$,

$$\mathbf{u}_c = \mathbf{u}_p + \begin{bmatrix} fb\left(\frac{1}{z_2} - \frac{1}{z_1}\right) \\ 0 \\ 0 \end{bmatrix}. \tag{8}$$

From Eq. 3, we get:

$$\frac{1}{z_2} - \frac{1}{z_1} = -\frac{\mathbf{n}^T\mathbf{u}_c}{fD}. \tag{9}$$



(a) $\boldsymbol{l}_\mathbf{u}$ formed by $\mathbf{u}_c$ and $\mathbf{u}_p$ (b) $\boldsymbol{l}_\mathbf{v}$ formed by $\mathbf{v}_c$ and $\mathbf{v}_p$ (c) Π defined by $\boldsymbol{l}_\mathbf{u}$ and $\boldsymbol{l}_\mathbf{v}$ (d) Degenerate case

Figure 2. **Plane estimation from a known feature correspondence.** (a) Line segments $\mathbf{u}_p$ and $\mathbf{u}_c$ from image and pattern features create a pair of planes which meet at a 3D line vector $\mathbf{l}_u$. (b) Another pair of image and pattern feature segments create a pair of planes meeting at another 3D line. (c) The two 3D lines define a 3D plane. (d) Pattern features aligned with epipolar lines create a degenerate case, which is avoided by designing pattern features such that line segments do not lie along epipolar lines.

By substituting Eq. 9 to Eq. 8, we get:

$$\mathbf{u}_c = \begin{bmatrix} \frac{D}{D+bn_x} & -\frac{bn_y}{D+bn_x} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{u}_p. \tag{10}$$

Similarly,

$$\mathbf{v}_c = \begin{bmatrix} \frac{D}{D+bn_x} & -\frac{bn_y}{D+bn_x} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{v}_p. \tag{11}$$

From these relationships, we can derive the equations for the plane parameters given the pattern feature and the corresponding image feature as explained in the next Section.

3. Plane Parameter Estimation from a Known Correspondence

In this section, we derive the expression for scene plane parameters when the correspondence between pattern feature and the corresponding image feature is *known*. Consider the plane including the projector center $\mathbf{c}_p$ and the line vector $\mathbf{u}_p$ of the pattern feature as shown in Fig 2 (a). Similarly, consider another plane including the camera center $\mathbf{c}_c$ and the line vector $\mathbf{u}_c$ of the corresponding image feature. These two planes meet at a 3D line vector $\mathbf{l}_u$ as shown in Fig 2 (a) if $\mathbf{u}_p$ and $\mathbf{u}_c$ are *not on the same epipolar line* as shown in Fig 2 (d). $\mathbf{l}_u$ can be computed by the cross product of the surface normals of these two planes. The surface normals of the plane including $\mathbf{c}_p$ and $\mathbf{u}_p$ and the plane including $\mathbf{c}_c$ and $\mathbf{u}_c$ are:

$$\mathbf{n}_{up} = \frac{\mathbf{p}_p \times \mathbf{u}_p}{\|\mathbf{p}_p \times \mathbf{u}_p\|} \text{ and } \mathbf{n}_{uc} = \frac{\mathbf{p}_c \times \mathbf{u}_c}{\|\mathbf{p}_c \times \mathbf{u}_c\|}, \tag{12}$$

respectively, where $\times$ is a cross product of the vectors and $\|\cdot\|$ is a norm of the vector. Thus, $\mathbf{l}_u$ can be obtained by:

$$\mathbf{l}_u = \mathbf{n}_{up} \times \mathbf{n}_{uc} = \frac{(\mathbf{p}_p \times \mathbf{u}_p) \times (\mathbf{p}_c \times \mathbf{u}_c)}{\| (\mathbf{p}_p \times \mathbf{u}_p) \times (\mathbf{p}_c \times \mathbf{u}_c) \|}. \tag{13}$$

Similarly, another pair of two planes created by $\mathbf{v}_p$ and $\mathbf{v}_c$ meet at a 3D line vector $\mathbf{l}_v$ as shown in Fig 2 (b), and $\mathbf{l}_v$ can be obtained by:

$$\mathbf{l}_v = \frac{(\mathbf{p}_p \times \mathbf{v}_p) \times (\mathbf{p}_c \times \mathbf{v}_c)}{\| (\mathbf{p}_p \times \mathbf{v}_p) \times (\mathbf{p}_c \times \mathbf{v}_c) \|}. \tag{14}$$

The surface normal of the plane can be obtained by the cross product of $\mathbf{l}_v$ and $\mathbf{l}_u$ as shown in Fig 2 (c):

$$\underbrace{\mathbf{n} = \mathbf{l}_v \times \mathbf{l}_u = \frac{((\mathbf{p}_p \times \mathbf{v}_p) \times (\mathbf{p}_c \times \mathbf{v}_c)) \times ((\mathbf{p}_p \times \mathbf{u}_p) \times (\mathbf{p}_c \times \mathbf{u}_c))}{\| ((\mathbf{p}_p \times \mathbf{v}_p) \times (\mathbf{p}_c \times \mathbf{v}_c)) \times ((\mathbf{p}_p \times \mathbf{u}_p) \times (\mathbf{p}_c \times \mathbf{u}_c)) \|}}_{\text{Eq. 3 of the main manuscript}}. \tag{15}$$

The polar angle θ and the azimuthal angle φ of the plane normal can be obtained from $\mathbf{n} = [\sin\theta\cos\varphi, \sin\theta\sin\varphi, -\cos\theta]^T$. From Eq. 4 and Eq. 5, the shortest distance D from the origin ($\mathbf{c}_c$) to Π is:

$$\underbrace{D = \frac{b\mathbf{n}^T\mathbf{p}_p}{p_{px} - p_{cx}} - \mathbf{n}^T\mathbf{c}_p}_{\text{Eq. 4 of the main manuscript}}. \tag{16}$$

4. Plane Parameter Locus

In this section, we derive the algebraic equations describing the plane parameter locus explained in Section 5 of the main manuscript. Let $\mathrm{I} = \{\mathbf{u}_c, \mathbf{v}_c, \mathbf{p}_c\}$ be the image feature corresponding to a pattern feature $\mathrm{P} = \{\mathbf{u}_p, \mathbf{v}_p, \mathbf{p}_p\}$. By pairing I and P, we can obtain the true plane parameter set $\Pi = [D, \theta, \varphi]^T$. For a given I, consider a set of pattern features $\mathrm{P}' = \{\mathbf{u}_p, \mathbf{v}_p, \mathbf{p'}_p\}$ on the same epipolar line, where $\mathbf{p'}_p = [p'_{px}, p_y, f]^T = [p_{px} + \omega, p_y, f]^T$ ($\omega \in \mathbb{R}$). By pairing I and P′, we can derive the equations of the plane parameter locus from Eq. 15 and Eq. 16 .

Let $\mathbf{n}'$ be the surface normal of the plane created by pairing I and P′. Using Eq. 15,

$$\mathbf{n}' = \left(\left(\mathbf{p'}_p \times \mathbf{v}_p\right) \times \left(\mathbf{p}_c \times \mathbf{v}_c\right)\right) \times \left(\left(\mathbf{p'}_p \times \mathbf{u}_p\right) \times \left(\mathbf{p}_c \times \mathbf{u}_c\right)\right). \tag{17}$$

We drop the normalization factor without loss of generality. By applying $\mathbf{p'}_p = [p_{px} + \omega, p_y, f]^T$, Eq. 1, and Eq. 2 to Eq. 17, we get:

$$\mathbf{n}' = \begin{bmatrix} f\left(-u_{cx}v_y + u_{px}v_y + u_y v_{cx} - u_y v_{px}\right) \\ f\left(-u_{px}v_{cx} + u_{cx}v_{px}\right) \\ p_{cx}u_y v_{px} - p_y u_{cx} v_{px} - p'_{px}u_y v_{cx} - p_{cx}u_{px}v_y + p_y u_{px} v_{cx} + p'_{px}u_{cx}v_y \end{bmatrix}. \tag{18}$$

Please note that we dropped the common factors in x-,y-, and z-components of $\mathbf{n}'$ without loss of generality. By applying Eq. 6, Eq. 10, and Eq. 11 to Eq. 18 and arranging terms, we get:

$$\mathbf{n}' = \begin{bmatrix} n'_x \\ n'_y \\ n'_z \end{bmatrix} = \frac{1}{a}\begin{bmatrix} n_x \\ n_y \\ n_z + \frac{\omega D}{bf} \end{bmatrix}, \tag{19}$$

where $a = \sqrt{n_x^2 + n_y^2 + \left(n_z + \frac{\omega D}{bf}\right)^2}$ is the normalization factor.

Let D' be the perpendicular distance from the camera origin to the plane created by pairing I and P′. Using Eq. 16,

$$D' = \frac{b\mathbf{n'}^T\mathbf{p'}_p}{p'_{px} - p_{cx}} - \mathbf{n'}^T\mathbf{c}_p = \frac{b\mathbf{n'}^T\mathbf{p}_c}{p'_{px} - p_{cx}} = \frac{Db\mathbf{n'}^T\mathbf{p}_c}{b\mathbf{n}^T\mathbf{p}_c + \omega D} = \frac{D}{a}. \tag{20}$$

Thus, the plane parameter set Π' of the plane created by pairing I and P′ is:

$$\Pi' = \begin{bmatrix} D' \\ \theta' \\ \varphi' \end{bmatrix} = \begin{bmatrix} D' \\ \tan^{-1}\left(\frac{\sqrt{{n'_x}^2 + {n'_y}^2}}{-n'_z}\right) \\ \tan^{-1}\left(\frac{n'_y}{n'_x}\right) \end{bmatrix} = \begin{bmatrix} \frac{D}{a} \\ \tan^{-1}\left(\frac{\sqrt{n_x^2 + n_y^2}}{-\left(n_z + \frac{\omega D}{bf}\right)}\right) \\ \varphi \end{bmatrix}. \tag{21}$$

Property 1. The parameter locus $\Lambda_m = \{\Pi_{m1}, \ldots, \Pi_{mN}\}$ created by pairing an image feature I_m and a set of pattern features $\{\mathrm{P}_1, \ldots, \mathrm{P}_N\}$ on the same epipolar line always lies on a plane parallel to the $\varphi = 0$ plane in the Π-space.

Proof: From Eq. 21, the azimuth angle φ' of Π' is always a constant φ.

Property 2. Let $\Lambda_m = \{\Pi_{m1}, \ldots, \Pi_{mN}\}$ be the parameter locus created in the same way as Property 1. Let P_μ ($\mu \in \{1, \ldots, N\}$) be the true corresponding pattern feature of I_m. Let $d_{\mu n}$ be the distance between pattern features P_μ and P_n on the epipolar line. Then, the locations of the elements of Λ_m are a function *only* of the *set* $D_\mu = \{d_{\mu n} \,|\, n \in \{1, \ldots, N\}\}$ of relative distances between the true and candidate pattern features.

Proof: From Eq. 21, the locations of the elements of Λ_m are a function only of $\omega = p'_{px} - p_{px}$ (not a function of $\mathbf{p}_p$ or $\mathbf{p}_c$), which is the relative distance between the true and candidate pattern features.

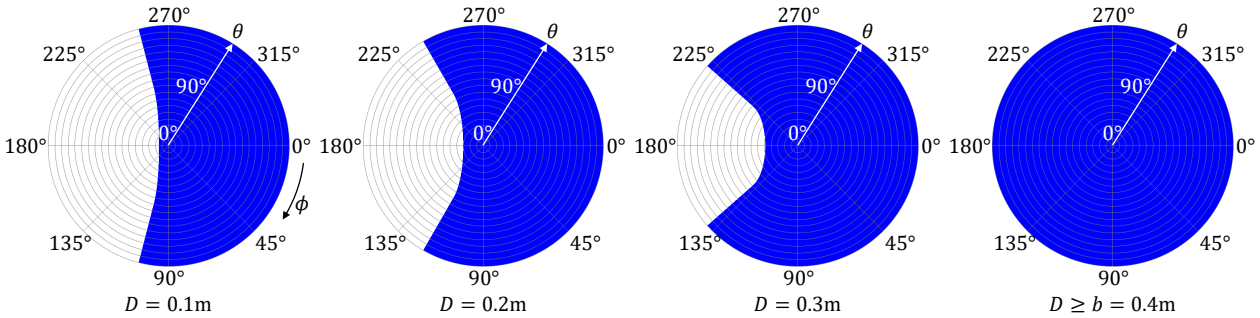


Figure 3. **Measurable plane space.** The measurable plane normal space increases with D, and if $D \geq b$ (baseline between the camera center and the projector center, $0.4\,\mathrm{m}$ assumed here), planes with any surface normals can be measured theoretically.

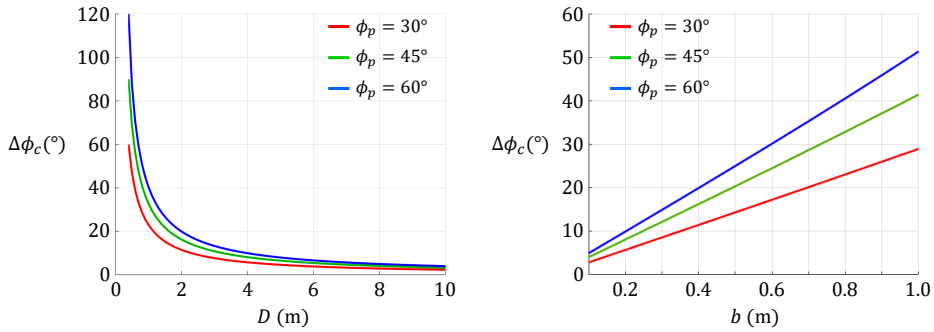


Figure 4. **Image feature angle variation over D (perpendicular distance to plane) and b (baseline).** Image feature angle variation decreases with D or increases with b. Therefore, it becomes difficult to measure precise plane parameters for in practice if D is very large or b is very small.

5. Measurable Plane Space

One of the important practical questions regarding the performance of Blocks-World Cameras is "what is the plane parameter space measurable with the Blocks-World Cameras?" The measurable plane space is determined by (a) fundamental limitations of the two-viewing imaging systems, and (b) measurement accuracy of image features.

Scene planes passing between the projector's and camera's optical centers or any segments on these planes are not measurable. This is because for such planes, the projected pattern cannot be observed by the camera. These non-measurable planes can be described as:

$$n_x = 0,\ D = 0 \quad \text{(planes including } x\text{-axis)} \tag{22}$$

and

$$n_x \neq 0,\ 0 \leq -\frac{D}{n_x} \leq b \quad \text{(planes with the } x\text{-intercept between 0 and } b\text{)}. \tag{23}$$

The measurable plane space is the complement of the set of the planes described by Eq. 22 and Eq. 23. Fig. 3 shows examples of the measurable plane space when the baseline $b = 0.4\,\mathrm{m}$. The measurable plane normals (represented by θ and φ) at different D values are shown in blue with the polar plot, where the radial direction and the clockwise direction represent θ and φ directions, respectively. The measurable plane normal space increases with D. If $D \geq b$, planes with any surface normals can be measured theoretically.

In practice, however, it is challenging to estimate plane parameters precisely when D is very large or b is very small. Let the pattern feature angle (i.e., the angle between one of the line segments of the pattern feature and the epipolar line) be ϕ_p, and the corresponding image feature angle (e.g., the angle between the corresponding image feature's line segment and the epipolar line) be ϕ_c. For a given ϕ_p, ϕ_c is a function of the plane parameters. As shown in Fig. 4, when the plane normal changes, the corresponding image feature angle variation $\Delta\phi_c$ gets smaller as D increases (b is fixed as $0.4\,\mathrm{m}$) or b decreases (D is fixed as $2\,\mathrm{m}$). If $\Delta\phi_c$ is too small, it is difficult to distinguish between different plane parameters.

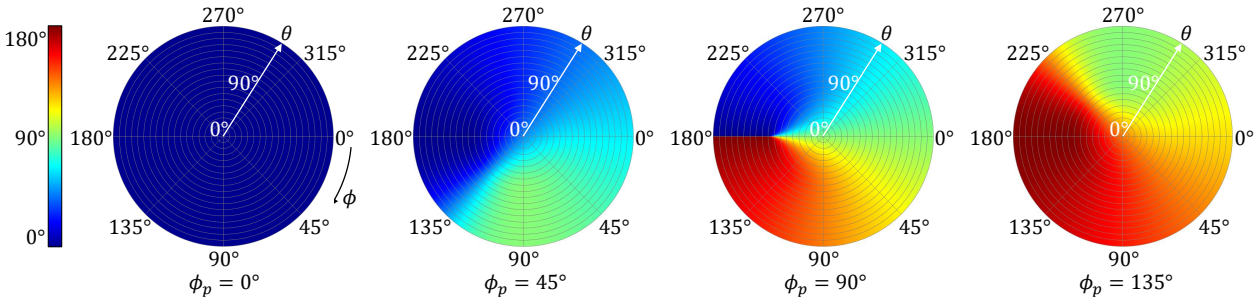


Figure 5. **Image feature angles according to plane normal direction over different pattern feature angles.** Image feature angle ϕ_c does not change when $\phi_p = 0°$, thus plane parameters cannot be estimated from feature deformation. On the other hand, ϕ_c changes sensitively according to the plane normal direction when $\phi_p = 90°$, leading to more accurate plane parameter estimation. $D = b = 0.4\,\mathrm{m}$ was assumed.

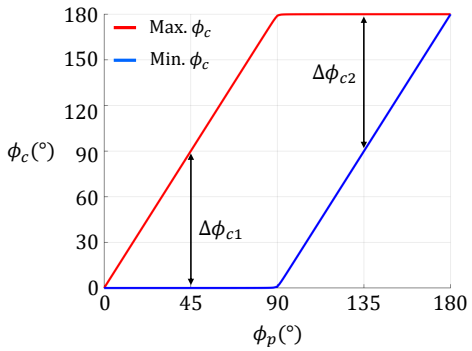


Figure 6. **Bound of image feature angle when plane normal changes over pattern feature angle.** We use two pattern feature angles $\phi_{p1} = 45°$ and $\phi_{p2} = 135°$ since they give the maximum image feature angle variation without the overlap.

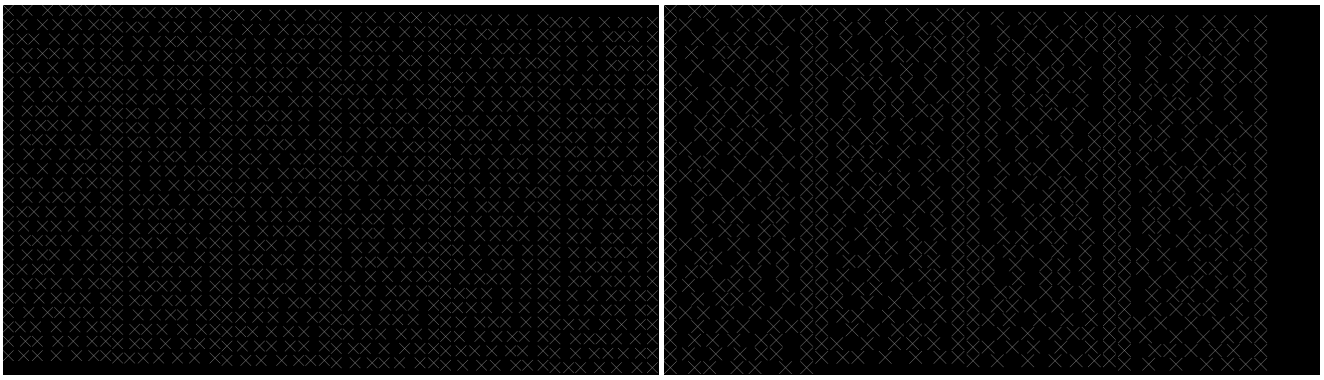
6. Pattern Design

In this section, we discuss various parameters for pattern design and their conditions for optimal performance of the Blocks-World Cameras.

Angles of pattern feature (why are $45°$ and $135°$ of pattern feature angles used?): The pattern feature angle (e.g., angle between the line segment of the pattern feature and the epipolar line) is an important parameter for pattern design since it influences the accuracy of plane parameter estimation. For a given pattern feature angle ϕ_p, the corresponding image feature angle ϕ_c (e.g., angle between the line segment of the corresponding image feature and the epipolar line) varies as the plane parameters Π change. The range of image feature angles $\Delta\phi_c$ (over all possible plane parameters) is determined by the pattern feature angle ϕ_p. For precise plane parameter estimation, $\Delta\phi_c$ should be sufficiently large. Fig. 5 shows ϕ_c as a function of the scene plane normal direction (D is fixed as $D = b = 0.4\,\mathrm{m}$) for different ϕ_ps. If $\phi_p = 0°$, ϕ_c is always $0°$ regardless of the plane normal direction (this corresponds to the degerate case in Fig. 2 (d)), which makes it impossible to estimate the plane parameters. On the other hand, if $\phi_p = 90°$, ϕ_c changes from $0°$ to $180°$, which makes it much easier to estimate the plane parameters accurately. Then, what is the optimal ϕ_p for precise plane parameter estimation?

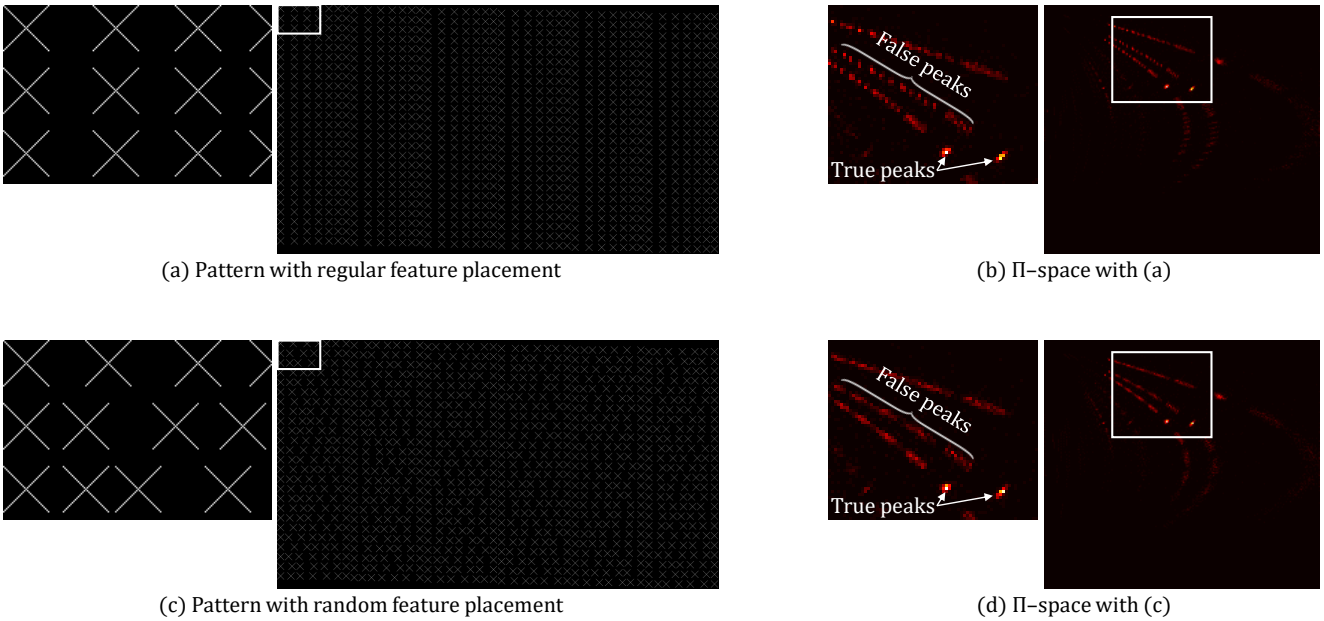
Fig. 6 shows the maximum and minimum ϕ_c values over ϕ_p when the plane normal changes. We achieve the maximum $\Delta\phi_c$ of $180°$ when $\phi_p = 90°$, and the minimum $\Delta\phi_c$ of $0°$ when $\phi_p = 0°$ or $\phi_p = 180°$. Therefore, $\phi_p = 90°$ should be selected for precise plane estimation. However, we need two ϕ_p values for plane estimation, and the range of two ϕ_cs corresponding to two ϕ_ps should not overlap for distinction between two image line segments when a *single* pattern is used. For this purpose, we chose $\phi_{p1} = 45°$ and $\phi_{p2} = 135°$ since $\Delta\phi_{c1}$ and $\Delta\phi_{c2}$ are maximized while achieving no overlap in the corresponding ϕ_c values as shown in Fig. 6.

Other pattern parameters: In addition to the pattern feature angle, there are more parameters for pattern design: radius of the line segment r, number of pattern features on a single epipolar line n, distance between adjacent epipolar lines with pattern features k, decrement or increment of distance between adjacent pattern features on each epipolar line h. These parameters



Pattern for simulations Pattern for experiments

Figure 7. **Example patterns for simulations and experiments.**



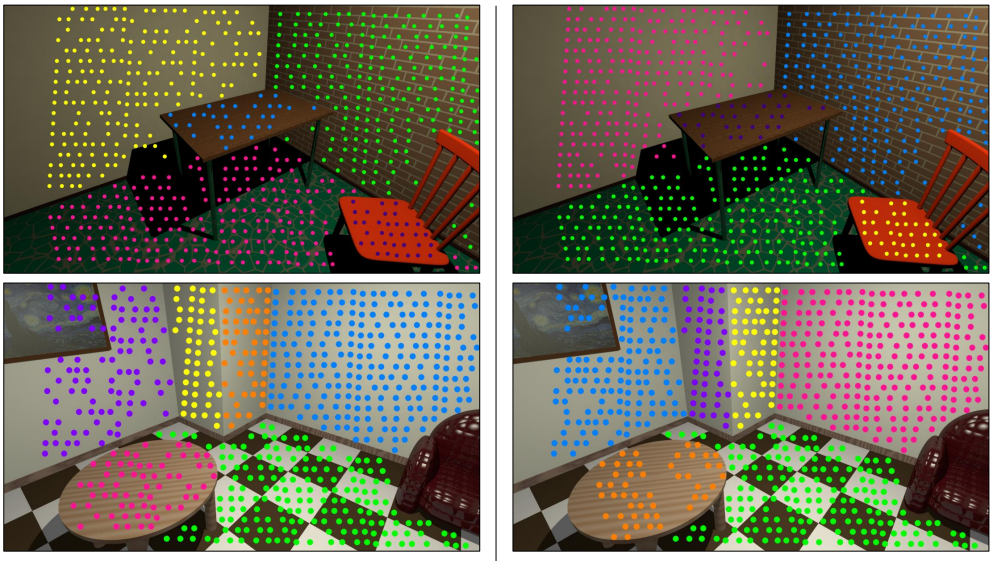
(a) Pattern with regular feature placement (b) Π–space with (a)

(c) Pattern with random feature placement (d) Π–space with (c)

Figure 8. **Pattern feature displacement for different epipolar lines and corresponding Π-spaces.** (a, b) Same pattern feature displacement for different epipolar lines and the resulting Π-space, respectively. (c, d) Random pattern feature displacement for different epipolar lines and the resulting Π-space, respectively. If the pattern feature displacement for different epipolar lines is different, false local peaks in Π-space are spread over the locus, leading to more robust true peak estimation.

can be chosen appropriately according to the scene or imaging conditions. Fig. 7 shows example patterns for our simulations and experiments. We use $r = 15$, $n = 7$, $k = 7$, $h = 5$ (in pixels) for simulations and $r = 18$, $n = 7$, $k = 10$, $h = 9$ (in pixels) for experiments. The total number of pattern features are 1050 and 735 for simulations and experiments, respectively. Larger size and sparser distribution of pattern features for the experiments are to be robust to various imperfections in the system. The resolution of the pattern is 1920×1080.

Pattern feature displacement for different epipolar lines: If pattern features are non-uniformly distributed (more specifically, all distances between pattern features (multi-hops as well as single-hop between pattern features) are different) on each epipolar line, true plane parameters can be estimated. Then what about the pattern feature displacement for different epipolar lines? Is the pattern feature displacement for different epipolar lines the same or different? As long as the non-uniform feature distribution on each epipolar line is satisfied, we can estimate the true plane parameters. However, if the pattern feature distribution for different epipolar lines is all different, we can find the local peaks for the true plane parameters more robustly. Fig. 8 shows Π-spaces for the scene of Fig. 5 in the main manuscript when the pattern feature displacement for



(a) Blocks-World Cameras with voting in Π- space

(b) Blocks-World Cameras with RANSAC-like procedure (no voting in Π- space)

Figure 9. **Blocks-World Cameras in two ways.** (a) Blocks-World Cameras with voing in Π-space. (b) Blocks-World Cameras with a RANSAC-like procedure, which are more memory-efficient because they do not require voting in Π-space.

different epipolar lines is the same (upper row) and random (lower row). True peaks representing the true plane parameters are not affected by the pattern feature displacement. However, false peaks are spread over the locus when the pattern feature displacement for different epipolar lines is random, which enables more robust true peak finding as shown in Fig. 8.

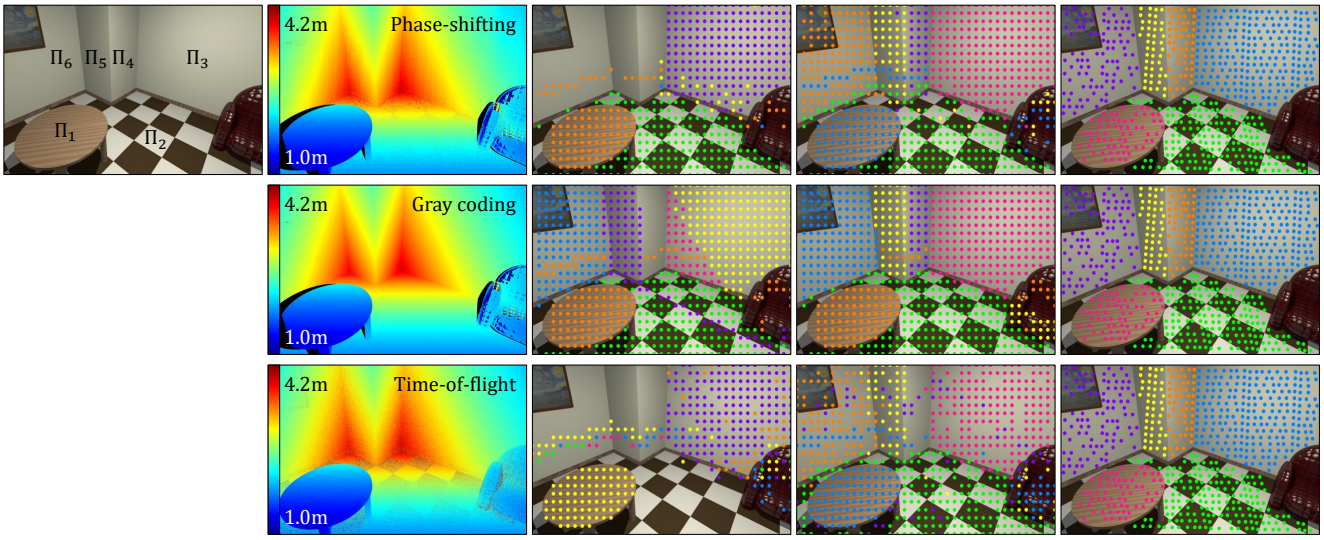
7. More Memory-Efficient Blocks-World Cameras

In this section, we discuss more memory-efficient Blocks-World Cameras which do not require a plane parameter Π-space for voting. Because the Blocks-World Cameras provide a pool of plane candidates with different confidence (e.g., larger number of plane candidates for real dominant scene planes), parameters for dominant planes can be estimated by finding inliers via a RANSAC-like procedure, instead of voting in the Π-space. The algorithm is as follows.

Algorithm 1: More Memory-Efficient Blocks-World Cameras

Input: $\mathbf{n}_{tol}$: error tolerance for plane normal, D_{tol}: error tolerance for D, T: number of iterations,
I_m $(m \in \{1, \dots M\})$: M number of image features,
P_n $(n \in \{1, \dots N\})$: N number of pattern features on the same epipolar line,
Π_{mn}: all plane candidates created by pairing I_m and P_n
Output: Π_q $(q \in \{1, \dots, Q\})$: Q number of dominant scene planes
for *q=1 to Q* **do**
  for *t=1 to T* **do**
    Randomly choose Π from $\{\Pi_{mn}\}$;
    Inliers $\leftarrow$ Π_{mn} within $\mathbf{n}_{tol}$ and D_{tol};
  end
  $\Pi_q \leftarrow$ model with the largest number of inliers (can be averaged inliers for more accurate model);
  Remove current inliers from $\{\Pi_{mn}\}$;
  Find I_m which participated in creating current inliers;
  Remove plane candidates created by these I_m from $\{\Pi_{mn}\}$;
end

Fig 9 (a) and (b) show the Blocks-World Camera results with voting in the Π-space and with a RANSAC-like procedure, respectively. All dominant planes are extracted well with the RANSAC-based Blocks-World Cameras as well. $\mathbf{n}_{tol} = 4°$,



(a) Scene (b) Simulated depth maps (c) Randomized 3D Hough (d) RANSAC (e) Blocks-World Cameras

Figure 10. **Comparison with plane-fitting.** (a) Scene. (b) Depth maps simulated by structured-light systems (sinusoid phase-shifting, binary Gray coding) and continuous wave time-of-flight (C-ToF) imaging. (c, d, e) Plane segmentation results by randomized 3D Hough transform (RHT), RANSAC, and Blocks-World Cameras.

$D_{tol} = 0.1$, and $T = 100$ are used to get the results.

8. Comparisons with Conventional Plane-Fitting Approaches

We compare plane estimation performance of the Blocks-World Cameras to the conventional approaches fitting the planes to 3D point clouds. For conventional approaches, we use 3D Hough transform and RANSAC, which are two popular plane-fitting methods. We use the randomized version of the 3D Hough transform (RHT) due to its computational efficiency. To create 3D point clouds for conventional approaches, we simulate structured-light (SL) systems and continuous-wave time-of-flight (C-ToF) imaging systems. We use sinusoid phase-shifting [3] and binary Gray coding [1, 2] to test two different sets of patterns for the SL systems. To simulate the C-ToF systems, multiple variables are required. Assuming sinusoid coding for amplitude modulation, the average numbers of signal photons (at the minimum depth) and ambient photons are assumed to be 1×10^6 and 5×10^5, respectively. $10\,\mathrm{MHz}$ of modulation frequency and $10\,\mathrm{ms}$ of integration time are used. Fig. 10 (b) shows the depth maps created by phase-shifting, Gray coding, and C-ToF, respectively. In case of the SL systems, Gray coding where 24 patterns are used shows a higher quality depth map than phase-shifting where 12 patterns are used. Although the performance of C-ToF depends on various parameter values, in general, C-ToF enables faster depth acquisition, longer depth range, but lower depth resolution compared to the SL systems. After creating 3D point clouds from the depth maps, we down-sample the point clouds such that the number of 3D points is the same as the number of image features captured by the Blocks-World Cameras to ensure fair comparison especially in terms of run-time.

Several parameter values are required for the MATLAB implementations of the conventional plane-fitting approaches. For RANSAC, we set the maximum number of iterations as 10^3 and the maximum distance from an inlier 3D point to the plane as $0.1\,\mathrm{m}$. For RHT, the bin sizes for θ and ϕ ranges are the same as 3° and the bin size for D range is $0.04\,\mathrm{m}$. These relatively large bin sizes are to handle noise existing in the 3D point clouds. The number of iterations is 10^6 for RHT. All these values are determined empirically to generate the most reasonable results. Fig. 10 (c), (d), and (e) show the plane segmentation results by RHT, RANSAC, and Blocks-World Cameras, respectively. For RHT (Fig. 10 (c)), it is challenging to extract small, distant or noisy planes because the votes for these planes are not reliably accumulated by random selection of points. Although RANSAC (Fig. 10 (d)) achieves better plane extraction, both RHT and RANSAC result in erroneous plane segmentation results (e.g., erroneous points on the imaginary planes created when the round table is expanded). This is a common issue with point cloud-based approaches since each 3D point does not have local plane information. In comparison, Blocks-World Cameras achieve accurate plane segmentation since each cross-shaped image feature contains partial information on the plane it belongs to, and does not need global reasoning.

Plane estimation results by the conventional approaches and the Blocks-World Cameras are compared to the ground truth

Plane	Π_1	Π_2	Π_3	Π_4	Π_5	Π_6
RHT	66, 270, 1.08	66, 270, 1.68	48, 153, 2.97	NA	NA	NA
RANSAC	64, 273, 1.15	59, 270, 1.84	46, 154, 3.00	48, 22, 2.31	37, 144, 2.41	55, 18, 3.00
Blocks-World Cameras	65, 270, 1.10	64, 269, 1.70	46, 155, 3.00	53, 20, 2.06	46, 152, 2.02	52, 18, 3.04
Ground truth	65, 270, 1.10	65, 270, 1.70	46, 153, 3.00	54, 20, 2.00	46, 153, 2.00	54, 20, 3.00

Table 1. Plane estimation results when the point cloud is created by a structured-light system with *phase-shifting* for conventional approaches.

Plane	Π_1	Π_2	Π_3	Π_4	Π_5	Π_6
RHT	66, 270, 1.08	66, 270, 1.68	51, 153, 2.96	NA	45, 153, 2.00	54, 21, 3.00
RANSAC	65, 271, 1.10	65, 270, 1.69	46, 153, 2.99	54, 19, 2.02	44, 151, 2.12	54, 20, 2.99
Blocks-World Cameras	65, 270, 1.10	64, 269, 1.70	46, 155, 3.00	53, 20, 2.06	46, 152, 2.02	52, 18, 3.04
Ground truth	65, 270, 1.10	65, 270, 1.70	46, 153, 3.00	54, 20, 2.00	46, 153, 2.00	54, 20, 3.00

Table 2. Plane estimation results when the point cloud is created by a structured-light system with *Gray coding* for conventional approaches.

Plane	Π_1	Π_2	Π_3	Π_4	Π_5	Π_6
RHT	63, 270, 1.12	NA	48, 159, 3.00	NA	NA	NA
RANSAC	67, 278, 1.13	63, 271, 1.76	47, 147, 3.00	NA	45, 154, 2.03	53, 21, 3.05
Blocks-World Cameras	65, 270, 1.10	64, 269, 1.70	46, 155, 3.00	53, 20, 2.06	46, 152, 2.02	52, 18, 3.04
Ground truth	65, 270, 1.10	65, 270, 1.70	46, 153, 3.00	54, 20, 2.00	46, 153, 2.00	54, 20, 3.00

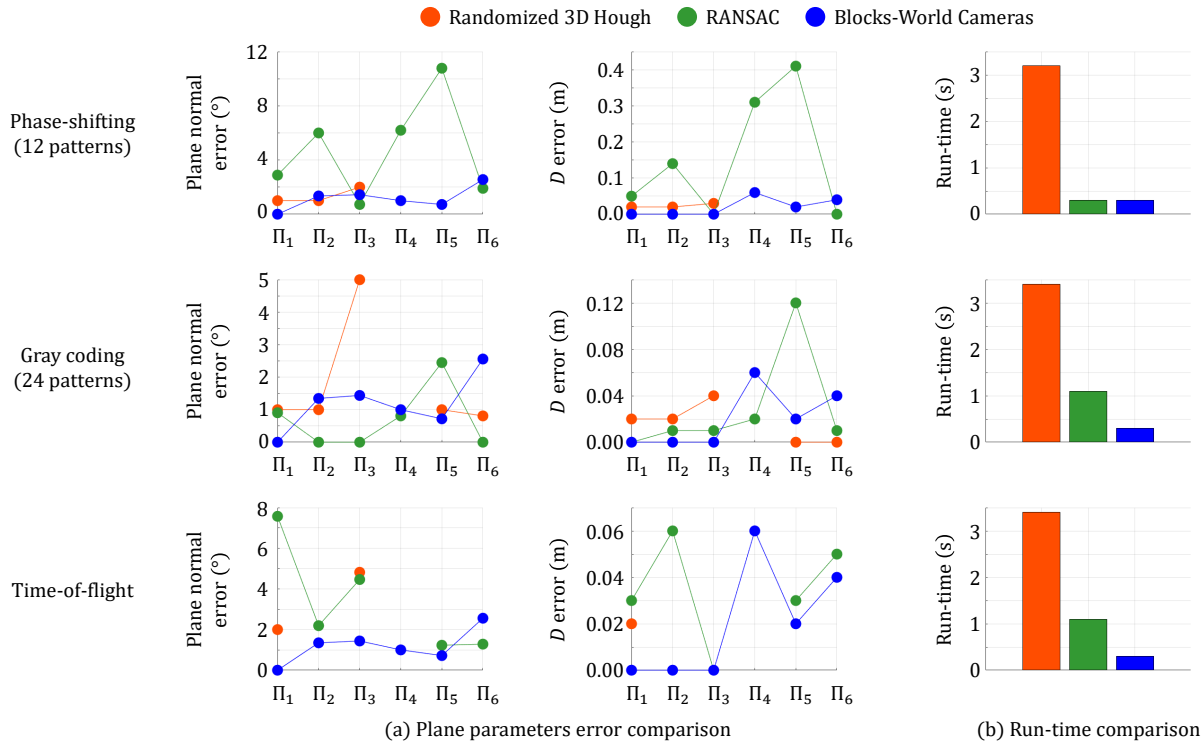
Table 3. Plane estimation results when the point cloud is created by a *continuous-wave time-of-flight imaging* for conventional approaches.

Plane	Π_1	Π_2	Π_3	Π_4	Π_5	Π_6
RHT	66, 270, 1.08	66, 270, 1.68	45, 159, 3.00	NA	NA	NA
RANSAC	63, 270, 1.14	62, 269, 1.75	46, 151, 2.99	56, 19, 1.93	43, 152, 2.13	56, 21, 2.97
Blocks-World Cameras	65, 270, 1.10	64, 269, 1.70	46, 155, 3.00	53, 20, 2.06	46, 152, 2.02	52, 18, 3.04
Ground truth	65, 270, 1.10	65, 270, 1.70	46, 153, 3.00	54, 20, 2.00	46, 153, 2.00	54, 20, 3.00

Table 4. Plane estimation results when the point cloud is down-sampled by 0.5 sampling rate for conventional approaches.

in Table 1, 2, and 3. Phase-shifting, Gray coding, and C-ToF are used to create the 3D point clouds for conventional approaches in Table 1, 2, and 3, respectively. Each cell in the table represents $[\theta(^\circ), \phi(^\circ), D(\mathrm{m})]$. The planes which cannot be segmented by the conventional approaches are represented by NA. The estimation errors and the run-time are shown in Fig. 11 (a) and (b), respectively. For the SL systems, conventional approaches show better performance in plane parameters error with Gray coding than phase-shifting since Gray coding uses more patterns leading to more accurate correspondence matching. The Blocks-World Cameras shows comparable performance to Gray coding even with a single pattern. For the C-ToF systems, the conventional approaches fail to find all dominant planes while the Blocks-World Cameras can. The conventional approaches are slower than the Blocks-World Cameras in run-time regardless of the imaging modalities.

We also discuss the trade-off between run-time and plane estimation accuracy while varying the sampling rate of the 3D point clouds. Fig. 12 (a) and (b) show the plane segmentation results by RHT and RANSAC, respectively after down-sampling the point clouds with different rates. The sampling rate of 0.02 is to ensure that the number of 3D points is the same as the number of image features of the Blocks-World Cameras. The sampling rate of 1.0 means no down-sampling of the 3D point clouds. The plane segmentation error by the conventional approaches is not reduced by increasing the sampling rates. Tables 4 and 5 summarize the ground truth plane parameters and the plane estimation results by the Blocks-World Cameras and the conventional approaches with different sampling rates. The plane estimation errors and the run-time comparisons with different sampling rates are shown in Fig. 13 (a) and (b), respectively. When the sampling rate increases, RANSAC becomes more accurate in plane parameter estimation, but it becomes slower in run-time. The Blocks-World Cameras show better performance than conventional approaches in both plane parameters error and run-time regardless of the sampling rate.



(a) Plane parameters error comparison (b) Run-time comparison

Figure 11. **Quantitative performance comparison with different imaging modalities.** Quantitative performance of the conventional approaches with different imaging modalities and the Blocks-World Cameras are compared in terms of (a) plane parameters error and (b) run-time (point cloud acquisition time is not included for conventional approaches). Structured-light systems (upper and middle rows) and continuous wave time-of-flight imaging system (lower row) are simulated. For the structured-light systems, sinusoid phase-shifting with 12 patterns (upper row) and binary Gray coding with 24 patterns are used. The Blocks-World Cameras with a single pattern show the performance comparable to the Gray coding with 24 patterns in plane parameters error while achieving very low computational complexity.

Plane	Π_1	Π_2	Π_3	Π_4	Π_5	Π_6
RHT	66, 270, 1.08	66, 270, 1.68	51, 153, 2.92	NA	NA	NA
RANSAC	65, 274, 1.12	66, 270, 1.69	43, 149, 3.03	55, 19, 1.99	48, 150, 1.96	54, 17, 3.01
Blocks-World Cameras	65, 270, 1.10	64, 269, 1.70	46, 155, 3.00	53, 20, 2.06	46, 152, 2.02	52, 18, 3.04
Ground truth	65, 270, 1.10	65, 270, 1.70	46, 153, 3.00	54, 20, 2.00	46, 153, 2.00	54, 20, 3.00

Table 5. Plane estimation results when the point cloud is down-sampled by 1.0 sampling rate for conventional approaches.

9. Mechanics of Plane Estimation in Plane Parameter Space

Bin sizes of plane parameter space: The optimal bin sizes of the plane parameter space (Π-space) depends on various factors such as scene conditions and imaging conditions. Roughly speaking, relatively larger bin sizes are used for low SNR conditions (e.g., noisy imaging conditions) and for non-planar scenes (e.g., Fig. 11 of the main manuscript). We use 1° for θ and φ ranges and $0.02\,\mathrm{m}$ for D range. To approximate the non-planar scene with piece-wise planar scene in the third result of Fig. 11 of the main manuscript, we use 7° for θ range, 10° for φ range and $0.05\,\mathrm{m}$ for D range.

Finding local peaks for true plane parameters: Multiple loci created by multiple image features on the same scene plane build several local peaks in Π-space. Only one peak represents true scene plane parameters, and others are false peaks created by possible candidate voting. Since a true local peak for a small scene plane can be lower than false local peaks for a huge scene plane, true local peaks should be selected carefully. We describe how to find the local peaks for true plane parameters with the Π-space of the scene in Fig. 5 of the main manuscript. 1) Find the maximum peak (e.g., peak pointed by Π_3 in the first sub-figure of Fig. 14) in the Π-space. This peak represents the true plane parameters for the plane Π_3. 2) Identify all image features which voted for this peak and remove all votes by these image features from the Π-space. Then all false peaks

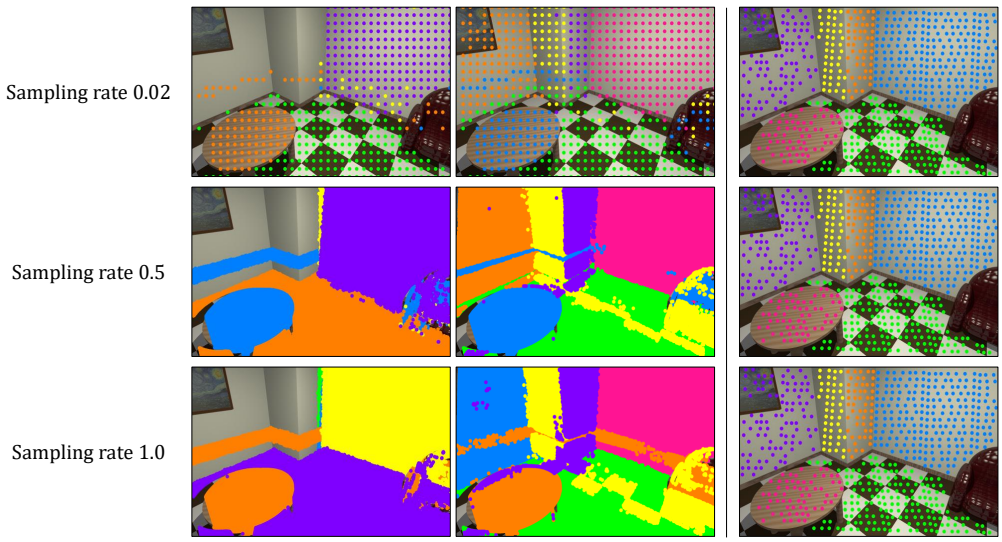


Figure 12. **Plane segmentation comparisons with different sampling rates of 3D point clouds.** After 3D point clouds are down-sampled with different sampling rates, planes are segmented by (a) randomized 3D Hough transform and (b) RANSAC. The segmentation results are compared to the result by (c) Blocks-World Cameras. The randomized 3D Hough transform fails to find all dominant planes even with the increased sampling rates. The plane segmentation error by conventional approaches is not reduced by increasing the sampling rate.

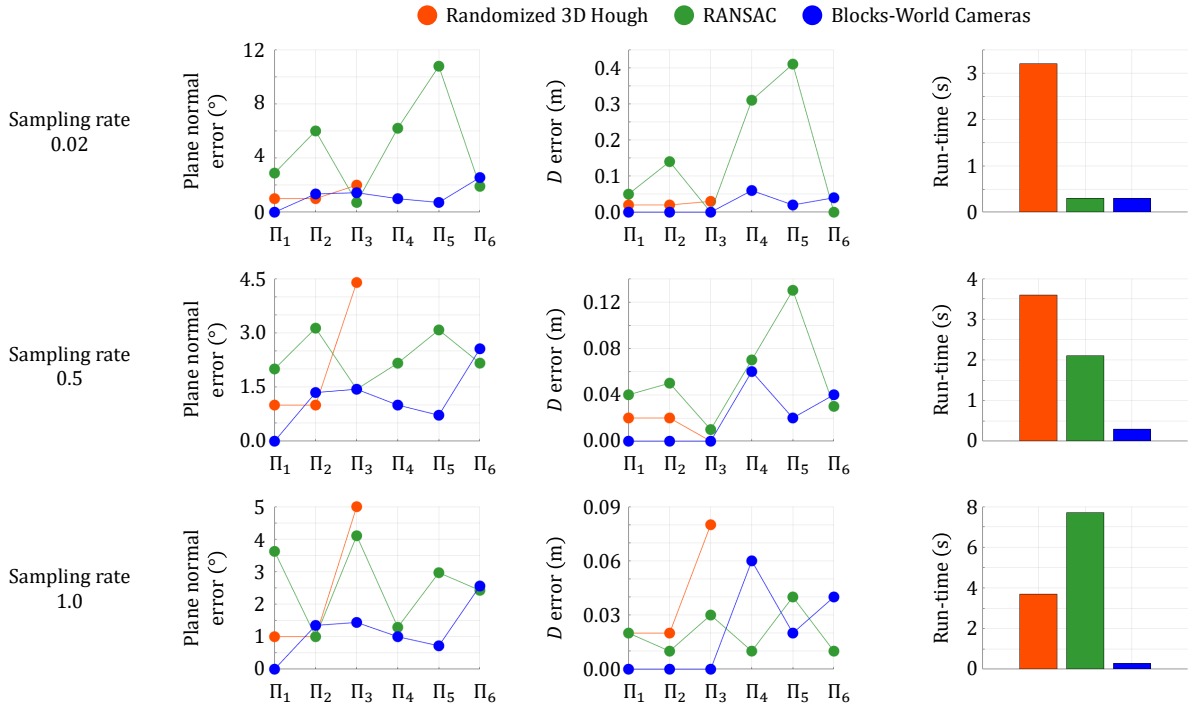
by these image features will disappear as shown in the second sub-figure of Fig. 14. Repeat 1) and 2) to find all true local peaks (true plane parameters) (Fig. 14).

References

[1] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. In *International Conference Pattern Recognition*, pages 806–808, 1984. 9

[2] K. Sato and S. Inokuchi. 3d surface measurement by space encoding range imaging. *Journal of Robotic Systems*, 2(1):27–39, 1985. 9

[3] V Srinivasan, Hsin-Chu Liu, and Maurice Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Applied optics*, 24(2):185–188, 1985. 9



(a) Plane parameters error comparison (b) Run-time comparison

Figure 13. **Quantitative performance comparison with different sampling rates of 3D point clouds.** Quantitative performance of the conventional approaches with different sampling rates of 3D point clouds and the Blocks-World Cameras are compared in terms of (a) plane parameters error and (b) run-time (point cloud acquisition time is not included for conventional approaches). When the sampling rate increases, RANSAC becomes more accurate in plane parameter estimation, but it becomes slower in run-time. Blocks-World Cameras show better performance than conventional approaches in both plane parameters error and run-time regardless of the sampling rate.

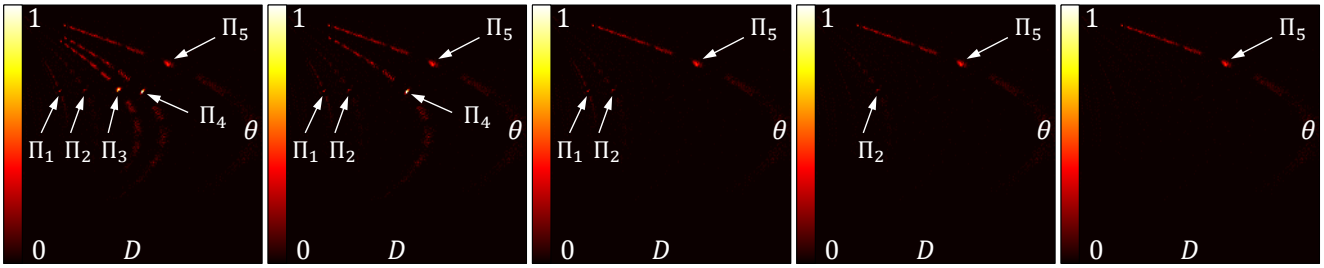


Figure 14. **Finding true plane parameters in Π-space.** Find the maximum peak and identify the image features voted for the maximum peak. Remove all votes by these image features from the Π-space. Repeat this to find all true plane parameters.