Coded Two-Bucket Cameras for Computer Vision

Mian Wei[1], Navid Sarhangnejad[2], Zhengfan Xia[2], Nikita Gusev[2], Nikola Katic[2],
Roman Genov[2], and Kiriakos N. Kutulakos[1]

[1] Department of Computer Science, University of Toronto, Canada
{mianwei,kyros}@cs.toronto.edu
[2] Department of Electrical Engineering, University of Toronto, Canada
{sarhangn,xia,nikita,roman}@ece.toronto.edu, katic.nik@gmail.com

Abstract. We introduce *coded two-bucket (C2B) imaging*, a new operating principle for computational sensors with applications in active 3D shape estimation and coded-exposure imaging. A C2B sensor modulates the light arriving at each pixel by controlling which of the pixel's two "buckets" should integrate it. C2B sensors output two images per video frame—one per bucket—and allow rapid, fully-programmable, per-pixel control of the active bucket. Using these properties as a starting point, we (1) develop an image formation model for these sensors, (2) couple them with programmable light sources to acquire *illumination mosaics*, *i.e.*, images of a scene under many different illumination conditions whose pixels have been multiplexed and acquired in one shot, and (3) show how to process illumination mosaics to acquire live disparity or normal maps of dynamic scenes at the sensor's native resolution. We present the first experimental demonstration of these capabilities, using a fully-functional C2B camera prototype. Key to this unique prototype is a novel programmable CMOS sensor that we designed from the ground up, fabricated and turned into a working system.

1 Introduction

New camera designs—and new types of imaging sensors—have been instrumental in driving the field of computer vision in exciting new directions. In the last decade alone, time-of-flight cameras [1,2] have been widely adopted for vision [3] and computational photography tasks [4–7]; event cameras [8] that support asynchronous imaging have led to new vision techniques for high-speed motion analysis [9] and 3D scanning [10]; high-resolution sensors with dual-pixel [11] and assorted-pixel [12] designs are defining the state of the art for smartphone cameras; and sensors with pixel-wise coded-exposure capabilities are starting to appear [13, 14] for compressed sensing applications [15].

Against this backdrop, we introduce a new type of computational video camera to the vision community—the *coded two-bucket (C2B) camera* (Fig. 1). The C2B camera is a pixel-wise coded-exposure camera that never blocks the incident light. Instead, each pixel in its sensor contains two charge-collection sites—two "buckets"—as well as a one-bit writeable memory that controls which bucket is active. The camera outputs two images per video frame—one per bucket—and performs exposure coding by rapidly controlling the active bucket of each pixel, via a programmable sequence of binary 2D patterns. Key to this unique functionality is a novel programmable CMOS sensor that

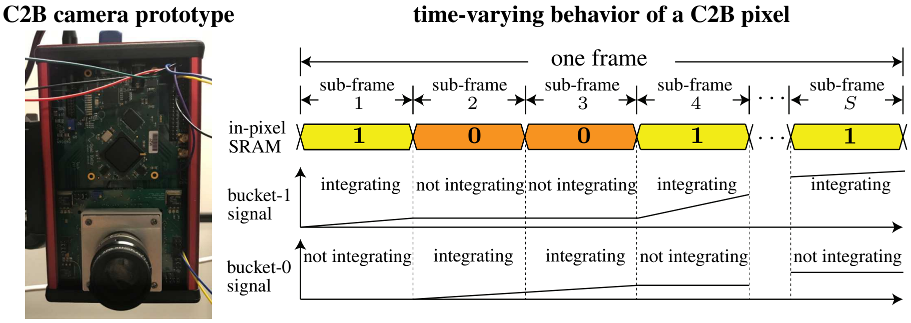


Fig. 1: *The C2B camera*. **Left:** Our prototype's sensor outputs video at 20 frames per second and consists of two arrays: a 244×160-pixel array that supports relatively slow bucket control (up to 4 sub-frames per frame) and a 35×48 array with much faster control (up to 120 sub-frames per frame). **Right:** Each frame is divided into S sub-frames during which the pixel's SRAM memory remains unchanged. A user-specified sequence of 2D binary patterns determines the SRAM's value at each pixel and sub-frame. Note that the two buckets of a pixel are never in the same state (*i.e.*, both active or both inactive) as this would degrade imaging performance—see [32] for a discussion of this and other related CMOS design issues. The light-generated charges of both buckets are read, digitized and cleared only once, at the end of each frame.

we designed from the ground up, fabricated in a CMOS image sensor (CIS) process technology [16] for the first time, and turned into a working camera system.

The light efficiency and electronic per-pixel coding capabilities of C2B cameras open up a range of applications that go well beyond what is possible today. This potentially includes compressive acquisition of high-speed video [17] with optimal light efficiency; simultaneous acquisition of both epipolar-only [18] and non-epipolar video streams; fully-electronic acquisition of high-dynamic-range AC-flicker videos [19]; conferring EpiScan3D-like functionality [20] to non-rectified imaging systems; and performing many other coded-exposure imaging tasks [15,21,22] with a compact camera platform.

Our focus in this first paper, however, is to highlight the novel capabilities of C2B cameras for live dense one-shot 3D reconstruction: we show that from just one grayscale C2B video frame of a dynamic scene under active illumination, it is possible to reconstruct the scene's 3D snapshot (*i.e.*, per-pixel disparity or normals, plus albedo) at a resolution comparable to the sensor's pixel array. We argue that C2B cameras allow us to reduce this very difficult 3D reconstruction problem [23–28] to the potentially much easier 2D problems of image demosaicing [29, 30] and illumination multiplexing [31].

In particular, we show that C2B cameras can acquire—in one frame—images of a scene under $S \geq 3$ linearly-independent illuminations, multiplexed across the buckets of $S-1$ neighboring pixels. We call such a frame a *two-bucket illumination mosaic*. In this setting, reconstruction at full sensor resolution involves four steps (Fig. 2): (1) control bucket activities and light sources to pack $2(S-1)$ distinct low-resolution images of the scene into one C2B frame (*i.e.*, $S-1$ images per bucket); (2) upsample these images

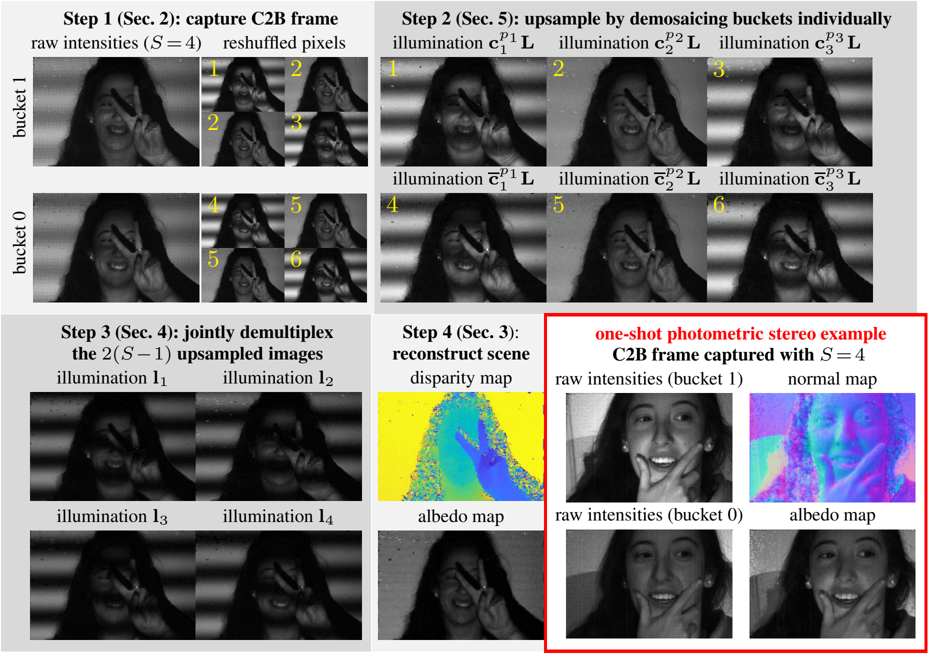

Step 1 (Sec. 2): capture C2B frame
raw intensities ($S=4$)
reshuffled pixels
bucket 1
bucket 0
1
2
2
3
4
5
5
6
Step 2 (Sec. 5): upsample by demosaicing buckets individually
illumination $\mathbf{c}_1^{p1}\mathbf{L}$
illumination $\mathbf{c}_2^{p2}\mathbf{L}$
illumination $\mathbf{c}_3^{p3}\mathbf{L}$
illumination $\overline{\mathbf{c}}_1^{p1}\mathbf{L}$
illumination $\overline{\mathbf{c}}_2^{p2}\mathbf{L}$
illumination $\overline{\mathbf{c}}_3^{p3}\mathbf{L}$
Step 3 (Sec. 4): jointly demultiplex
the $2(S-1)$ upsampled images
illumination $\mathbf{l}_1$
illumination $\mathbf{l}_2$
illumination $\mathbf{l}_3$
illumination $\mathbf{l}_4$
Step 4 (Sec. 3):
reconstruct scene
disparity map
albedo map
one-shot photometric stereo example
C2B frame captured with $S=4$
raw intensities (bucket 1)
normal map
raw intensities (bucket 0)
albedo map

Fig. 2: *Dense one-shot reconstruction with C2B cameras.* The procedure runs in real time and is illustrated for structured-light triangulation. Please zoom in to the electronic copy to see individual pixels of the C2B frame and refer to the listed sections for notation and details. Photometric stereo is performed in an analogous way, by replacing the structured-light projector with a set of S directional light sources (a reconstruction example is shown in the lower right).

to full resolution by demosaicing; (3) demultiplex all the upsampled images jointly, to obtain up to S linearly-independent full-resolution images; and (4) use these images to solve for shape and albedo at each pixel independently. We demonstrate the effectiveness of this procedure by recovering dense 3D shape and albedo from one shot with two of the oldest and simplest active 3D reconstruction algorithms available—multi-pattern cosine phase shifting [33, 34] and photometric stereo [35].

From a hardware perspective, we build on previous attempts to fabricate sensors with C2B-like functionality [36–38], which did not rely on a CMOS image sensor process technology. More broadly, our prototype can be thought of as generalizing three families of sensors. *Programmable coded-exposure sensors* [13] allow individual pixels to be "masked" for brief periods during the exposure of a video frame (Fig. 3, left). Just like the C2B sensor, they have a writeable one-bit memory inside each pixel to control masking, but their pixels lack a second bucket so light falling onto "masked" pixels is lost. *Continuous-wave time-of-flight sensors* [1, 2] can be thought of as having complementary functionality to coded-exposure sensors: their pixels have two buckets whose activity can be toggled programmatically (so no light is lost), but they have no in-pixel writeable memory. As such, the active bucket is constrained to be the same for all pix-

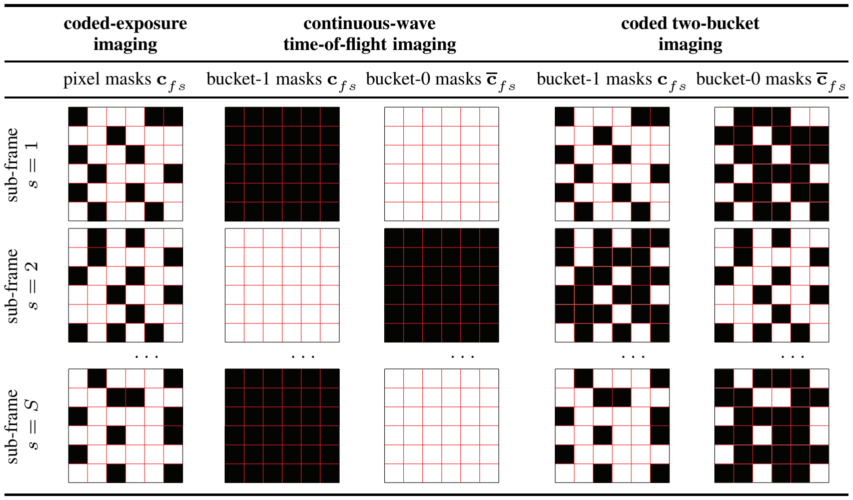
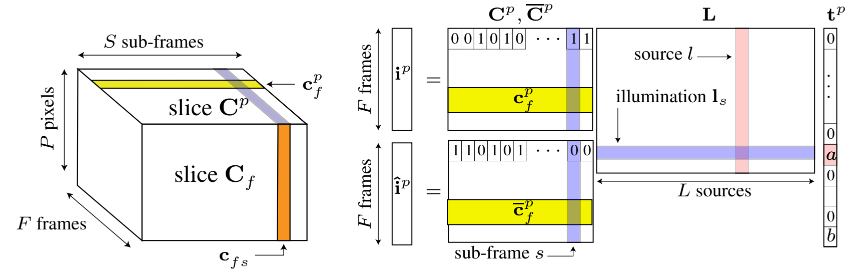

coded-exposure imaging
continuous-wave time-of-flight imaging
coded two-bucket imaging
pixel masks $\mathbf{c}_{fs}$
bucket-1 masks $\mathbf{c}_{fs}$
bucket-0 masks $\overline{\mathbf{c}}_{fs}$
bucket-1 masks $\mathbf{c}_{fs}$
bucket-0 masks $\overline{\mathbf{c}}_{fs}$
sub-frame $s = 1$
sub-frame $s = 2$
sub-frame $s = S$
...

Fig. 3: *Comparison of basic sensor abilities.* Coded-exposure sensors can rapidly mask individual pixels but cannot collect all the incident light; continuous-wave ToF sensors always collect all the incident light but they cannot mask pixels individually; C2B sensors can do both. The column vectors $\mathbf{c}_{fs}$ and $\overline{\mathbf{c}}_{fs}$ denote bucket-1 masks/activities and their binary complement, respectively.

els (Fig. 3, middle). This makes programmable per-pixel coding—and acquisition of illumination mosaics in particular—impossible without specialized optics (*e.g.*, [17]). *Multi-bucket (*a.k.a.*, "multi-tap") sensors* [39–42] have more than two buckets in each pixel but they have no writeable memory either, so per-pixel coding is not possible. In theory, an S-bucket sensor would be uniquely suited for dense one-shot reconstruction because it can acquire in each frame S full-resolution images corresponding to any set of S illuminations [43]. In practice, however, C2B sensors have several advantages: they are scalable because they can pack S linearly-independent images into one frame for any value of S—without hard-wiring this value into the pixel's CMOS design; they are much more light efficient because each extra bucket reduces the pixel's photo-sensitive region significantly for a given pixel size; and they have a broader range of applications because they enable per-pixel coding. To our knowledge, 2D sensors with more than four buckets have not been fabricated in a standard CMOS image process, and it is unclear if they could offer acceptable imaging performance.

On the conceptual side, our contributions are the following: (1) we put forth a general model for the C2B camera that opens up new directions for coded-exposure imaging with active sources; (2) we formulate its control as a novel multiplexing problem [31, 44–49] in the bucket and pixel domains; (3) we draw a connection between two-bucket imaging and algorithms that operate directly on intensity ratios [50]; and (4) we provide an algorithm-independent framework for dense one-shot reconstruction that is simpler than earlier attempts [18] and is compatible with standard image processing pipelines.

Last but not least, we demonstrate all the above experimentally, on the first fully-operational C2B camera prototype.



(a) (b)

Fig. 4: **(a)** Structure of the code tensor $\mathbf{C}$. **(b)** Image formation model for pixel p. We show the transport vector $\mathbf{t}^p$ for a structured-light setting, where the contribution of ambient illumination is $\mathbf{t}^p[L] = b$, the corresponding projector pixel is l, and its albedo is $\mathbf{t}^p[l] = a$.

2 Coded Two-Bucket Imaging

We begin by introducing an image formation model for C2B cameras. We consider the most general setting in this section, where a whole sequence of C2B frames may be acquired instead of just one.

C2B cameras output *two* images per video frame—one for each bucket (Fig. 2). We refer to these images as the *bucket*-1 *image* and *bucket*-0 *image*.

The code tensor. Programming a C2B camera amounts to specifying the time-varying contents of its pixels' memories at two different timescales: (1) at the scale of *sub-frames* within a video frame, which correspond to the updates of the in-pixel memories (Fig. 1, right), and (2) at the scale of frames within a video sequence. For a video sequence with F frames and a camera that has P pixels and supports S sub-frames, bucket activities can be represented as a three-dimensional binary tensor $\mathbf{C}$ of size $P \times F \times S$. We call $\mathbf{C}$ the *code tensor* (Fig. 4a).

We use two specific 2D "slices" of the code tensor in our analysis below, and have special notation for them. For a specific pixel p, slice $\mathbf{C}^p$ describes the activity of pixel p's buckets across all frames and sub-frames. Similarly, for a specific frame f, slice $\mathbf{C}_f$ describes the bucket activity of all pixels across all sub-frames of f:

$$\mathbf{C}^p = \underbrace{\begin{bmatrix} \mathbf{c}^p_1 \\ \mathbf{c}^p_2 \\ \vdots \\ \mathbf{c}^p_F \end{bmatrix}}_{F \times S} \qquad \mathbf{C}_f = \underbrace{\begin{bmatrix} \mathbf{c}_{f1} & \mathbf{c}_{f2} & \dots & \mathbf{c}_{fS} \end{bmatrix}}_{P \times S}, \tag{1}$$

where $\mathbf{c}^p_f$ is an S-dimensional row vector that specifies the active bucket of pixel p in the sub-frames of frame f; and $\mathbf{c}_{fs}$ is a P-dimensional column vector that specifies the active bucket of all pixels in sub-frame s of frame f.

The illumination matrix. Although C2B cameras can be used for passive imaging applications [15], we model the case where illumination is programmable at sub-frame timescales too. In particular, we represent the scene's time-varying illumination condition as an illumination matrix $\mathbf{L}$ that applies to all frames:

$$\mathbf{L} \; = \; \underbrace{\begin{bmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \\ \vdots \\ \mathbf{l}_S \end{bmatrix}}_{S \times L} \quad , \tag{2}$$

where row vector $\mathbf{l}_s$ denotes the scene's illumination condition in sub-frame s of every frame. We consider two types of scene illumination in this work: a set of L directional light sources whose intensity is given by vector $\mathbf{l}_s$; and a projector that projects a pattern specified by the first $L-1$ elements of $\mathbf{l}_s$ in the presence of ambient light, which we treat as an L-th source that is "always on" (*i.e.*, element $\mathbf{l}_s[L] = 1$ for all s).

Two-bucket image formation model for pixel p. Let $\mathbf{i}^p$ and $\hat{\mathbf{i}}^p$ be column vectors holding the intensities of pixel p's bucket 1 and bucket 0, respectively, in F frames. We model these intensities as the result of light transport from the L light sources to the pixel's two buckets (Fig. 4b):

$$\underbrace{\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix}}_{2F \times 1} \; = \; \underbrace{\begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix}}_{2F \times S} \underbrace{\;\mathbf{L}\;}_{S \times L} \underbrace{\;\mathbf{t}^p\;}_{L \times 1} \quad , \tag{3}$$

where $\overline{b}$ denotes the binary complement of matrix or vector b, $\mathbf{C}^p$ is the slice of the code tensor corresponding to p, and $\mathbf{t}^p$ is the pixel's *transport vector*. Element $\mathbf{t}^p[l]$ of this vector describes the transport of light from source l to pixel p in the timespan of one sub-frame, across all light paths.

To gain some intuition about Eq. (3), consider the buckets' intensity in frame f:

$$\mathbf{i}^p[f] = \underbrace{(\mathbf{c}^p_f \, \mathbf{L})}_{\substack{\text{illumination condition} \\ \text{of pixel } p, \text{ bucket 1, frame } f}} \; \mathbf{t}^p \qquad\qquad \hat{\mathbf{i}}^p[f] = \underbrace{(\overline{\mathbf{c}}^p_f \, \mathbf{L})}_{\substack{\text{illumination condition} \\ \text{of pixel } p, \text{ bucket 0, frame } f}} \; \mathbf{t}^p \quad . \tag{4}$$

In effect, the two buckets of pixel p can be thought of as "viewing" the scene under two potentially different illuminations given by vectors $\mathbf{c}^p_f \mathbf{L}$ and $\overline{\mathbf{c}}^p_f \mathbf{L}$, respectively. Moreover, if $\mathbf{c}^p_f$ varies from frame to frame, these illumination conditions may vary as well.

Bucket ratios as albedo "quasi-invariants." Since the two buckets of pixel p generally represent different illumination conditions, the two ratios

$$r \; = \; \frac{\mathbf{i}^p[f]}{\mathbf{i}^p[f] + \hat{\mathbf{i}}^p[f]} \quad , \quad \hat{r} \; = \; \frac{\hat{\mathbf{i}}^p[f]}{\mathbf{i}^p[f] + \hat{\mathbf{i}}^p[f]} \quad , \tag{5}$$

defined by p's buckets are *illumination ratios* [50–52]. Moreover, we show in [32] that under zero-mean Gaussian image noise, these ratios are well approximated by Gaussian random variables whose mean is the ideal (noiseless) ratio and whose standard deviation depends weakly on albedo. In effect, C2B cameras provide two "albedo-invariant" images per frame. We exploit this feature of C2B cameras for both shape recovery and demosaicing in Secs. 3 and 5, respectively.

2.1 Acquiring Two-Bucket Illumination Mosaics

A key feature of C2B cameras is that they offer an important alternative to multi-frame acquisition: instead of capturing F frames in sequence, they can capture a spatially-multiplexed version of them in a single C2B frame (Fig. 2). We call such a frame a *two-bucket illumination mosaic* in analogy to the RGB filter mosaics of color image sensors [12,53,54]. Unlike filter mosaics, however, which are attached to the sensor and cannot be changed, acquisition of illumination mosaics is programmable for any F.

The bucket-1 and bucket-0 image sequences. Collecting the two buckets' intensities in Eq. (4) across all frames and pixels, we define two matrices that hold all this data:

$$\mathbf{I} = \underbrace{\left[\mathbf{i}^1 \;\; \mathbf{i}^2 \;\; \ldots \;\; \mathbf{i}^P\right]}_{F \times P} \qquad \hat{\mathbf{I}} = \underbrace{\left[\hat{\mathbf{i}}^1 \;\; \hat{\mathbf{i}}^2 \;\; \ldots \;\; \hat{\mathbf{i}}^P\right]}_{F \times P} \ . \tag{6}$$

Code tensor for mosaic acquisition. Formally, a two-bucket illumination mosaic is a spatial sub-sampling of the sequences $\mathbf{I}$ and $\hat{\mathbf{I}}$ in Eq. (6). Acquiring it amounts to specifying a one-frame code tensor $\widetilde{\mathbf{C}}$ that spatially multiplexes the corresponding F-frame tensor $\mathbf{C}$ in Fig. 4(a). We do this by (1) defining a regular tiling of the sensor plane and (2) specifying a correspondence $(p_i \rightarrow f_i), 1 \leq i \leq K$, between the K pixels in a tile and frames. The rows of $\widetilde{\mathbf{C}}$ are then defined to be

$$\widetilde{\mathbf{c}}_1^{p_i} \overset{\text{def}}{=} \mathbf{c}_{f_i}^{p_i} \ . \tag{7}$$

Mosaic acquisition example. The C2B frames in Fig. 2 were captured using a 2×2-pixel tile to spatially multiplex a three-frame code tensor. The tensor assigned identical illumination conditions to all pixels within a frame and different conditions across frames. Pixels within each tile were assigned to individual frames using the correspondence $\{(1,1) \rightarrow 1,\ (1,2) \rightarrow 2,\ (2,1) \rightarrow 2,\ (2,2) \rightarrow 3\}$.

3 Per-Pixel Estimation of Normals and Depth

Let us now turn to the problem of normal and depth estimation using photometric stereo and structured-light triangulation, respectively. We consider the most basic formulation of these tasks, where all computations are done independently at each pixel and the relation between observations and unknowns is expressed as a system of linear equations. These formulations should be treated merely as examples that showcase the special characteristics of two-bucket imaging; as with conventional cameras, using advanced methods to handle more general settings [55,56] is certainly possible.

From bucket intensities to demultiplexed intensities. As a starting point, we expand Eq. (3) to get a relation that involves only intensities:

$$\underbrace{\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix}}_{\substack{\text{bucket measurements} \\ (2F \times 1)}} = \begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \mathbf{t}^p \\ \vdots \\ \mathbf{l}_S \mathbf{t}^p \end{bmatrix} \overset{\text{def}}{=} \underbrace{\begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix}}_{\substack{\text{bucket-multiplexing} \\ \text{matrix } \mathbf{W} \\ (2F \times S)}} \underbrace{\begin{bmatrix} i_1^p \\ \vdots \\ i_S^p \end{bmatrix}}_{\substack{\text{pixel intensity under} \\ \text{illuminations } \mathbf{l}_1, \ldots, \mathbf{l}_S \\ (S \times 1)}} \ . \tag{8}$$

	Lambertian photometric stereo	**Structured-light triangulation w/ cosine patterns**
Assumptions	Lambertian reflectance, non-uniform albedo; calibrated light sources; no ambient or indirect light	reflectance has non-negligible diffuse component; robustness to indirect light depends on frequency choice
Illumination vectors $\mathbf{l}_s$	each $\mathbf{l}_s$ corresponds to illumination with only source s turned on, *i.e.*, element $\mathbf{l}_s[l]$ is non-zero iff $s = l$	$\mathbf{l}_s[l] = \cos(\phi_s + \theta_l)$, where ϕ_s is phase shift of s-th pattern, θ_l is phase of projector pixel l
Vector $\mathbf{d}_s$	orientation and intensity of source s, expressed as a 3D row vector	$\mathbf{d}_s = \begin{bmatrix}\cos(\phi_s) & -\sin(\phi_s) & 1\end{bmatrix}$
Matrix $\mathbf{D}$	matrix whose rows are the vectors $\mathbf{d}_1, \ldots \mathbf{d}_S$	matrix whose rows are the vectors $\mathbf{d}_1, \ldots \mathbf{d}_S$
Transport vector $\mathbf{t}$	$\mathbf{t}[s] = a\mathbf{d}_s\mathbf{n}$ where a is the albedo and $\mathbf{n}$ is the unit surface normal	$\mathbf{t} = \begin{bmatrix}a\mathbf{m} & b\end{bmatrix}'$, where a is albedo, b is the contribution of ambient light, and binary row vector $\mathbf{m}$ indicates the matching projector pixel, *i.e.*, $\mathbf{m}[l] = 1$ iff that pixel is l (see Fig. 4b)
Vector $\mathbf{x}$	$\mathbf{x} = \mathbf{n}$	$\mathbf{x} = \begin{bmatrix}\cos(\theta) & \sin(\theta) & \frac{b}{a}\end{bmatrix}'$ if the same cosine frequency is used for all patterns; additional frequencies contribute two unknowns each; θ is the phase of the matching projector pixel

Table 1: *The two basic multi-image reconstruction techniques considered in this work.*

Each scalar i^p_s in the right-hand side of Eq. (8) is the intensity that a conventional camera pixel would record if the scene's illumination condition was $\mathbf{l}_s$. Therefore, Eq. (8) tells us that as far as a single pixel p is concerned, C2B cameras capture the same S measurements a conventional camera would capture for 3D reconstruction—except that those measurements are multiplexed over $2F$ bucket intensities. To retrieve them, these intensities must be demultiplexed by inverting Eq. (8):

$$\begin{bmatrix} i^p_1 \\ \vdots \\ i^p_S \end{bmatrix} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}' \begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix} \quad , \tag{9}$$

where $'$ denotes matrix transpose. This inversion is only possible if $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ is non-singular. Moreover, the signal-to-noise ratio (SNR) of the demultiplexed intensities depends heavily on $\mathbf{W}$ and $\mathbf{C}^p$ (Sec. 4). Setting aside this issue for now, we consider below the task of shape recovery from already-demultiplexed intensities. For notational simplicity we drop the pixel index p from the equations below.

Per-pixel constraints on 3D shape. The relation between demultiplexed intensities and the pixel's unknowns takes exactly the same form in both photometric stereo and structured-light triangulation with cosine patterns:

$$\begin{bmatrix} i_1 \\ \vdots \\ i_S \end{bmatrix} = a\mathbf{D}\mathbf{x} + \mathbf{e} \quad , \tag{10}$$

where $\mathbf{D}$ is known; $\mathbf{x}$ is a 3D vector that contains the pixel's shape unknowns; a is the unknown albedo; and $\mathbf{e}$ is observation noise. See Table 1 for a summary of each problem's assumptions and for the mapping of problem-specific quantities to Eq. (10).

There are (at least) three ways to turn Eq. (10) into a constraint on normals and depths under zero-mean Gaussian noise. The resulting constraints are *not* equivalent when combining measurements from small pixel neighborhoods—as we implicitly do—because they are not equally invariant to spatial albedo variations:

1. *Direct method (DM):* treat Eq. (10) as providing S independent constraints on vector $a\mathbf{x}$ and solve for both a and $\mathbf{x}$. The advantage of this approach is that errors are Gaussian by construction; its disadvantage is that Eq. (10) depends on albedo.
2. *Ratio constraint (R):* divide individual intensities by their total sum to obtain an illumination ratio, as in Eq. (5). This yields the following constraint on $\mathbf{x}$:

$$r_l \mathbf{1}\mathbf{D}\mathbf{x} \;=\; \mathbf{d}_l\mathbf{x} \quad , \tag{11}$$

   where $r_l = i_l / \sum_k i_k$ and $\mathbf{1}$ is a row vector of all ones. The advantage here is that both r_l and Eq. (11) are approximately invariant to albedo.
3. *Cross-product constraint (CP):* instead of computing an explicit ratio from Eq. (10), eliminate a to obtain

$$i_l \mathbf{d}_k \mathbf{x} \;=\; i_k \mathbf{d}_l \mathbf{x} \ . \tag{12}$$

   Since Eq. (12) has intensities i_l, i_k as factors, it does implicitly depend on albedo.

Solving for the unknowns. Both structured light and photometric stereo require at least $S \geq 3$ independent constraints for a unique solution. In the DM method we use least-squares to solve for $a\mathbf{x}$; when using the R or CP constraints, we apply singular-value decomposition to solve for $\mathbf{x}$.

4 Code Matrices for Bucket Multiplexing

The previous section gave ways to solve for 3D shape when we have enough independent constraints per pixel. Here we consider the problem of controlling a C2B camera to actually obtain them for a pixel p. In particular, we show how to choose (1) the number of frames F, (2) the number of sub-frames per frame S, and (3) the pixel-specific slice $\mathbf{C}^p$ of the code tensor, which defines the multiplexing matrix $\mathbf{W}$ in Eq. (8).

Determining these parameters can be thought of as an instance of the *optimal multiplexing* problem [31, 44–49]. This problem has been considered in numerous contexts before, as a one-to-one mapping from S desired measurements to S actual, noisy observations. In the case of coded two-bucket imaging, however, the problem is slightly different because each frame yields two measurements instead of just one.

The results below provide further insight into this particular multiplexing problem (see [32] for proofs). Observation 1 implies that even though a pixel's two buckets provide $2F$ measurements in total across F frames, at most $F + 1$ of them can be independent because the multiplexing matrix $\mathbf{W}$ is rank-deficient:

Observation 1 $\operatorname{rank} \mathbf{W} \leq \min(F + 1, S)$.

Intuitively, a C2B camera should not be thought of as being equivalent to two coded-exposure cameras that operate completely independently. This is because the activities of a pixel's two buckets are binary complements of each other, and thus not independent.

# sub-frames	$S=3$	$S=4$	$S=5$	$S=6$	$S=7$
Eq. (13) bound for $\sigma=1$	0.5556	0.4167	0.34	0.2889	0.2517
Optimal MSE for $\sigma=1$	0.8333	0.4167	0.3778	0.3467	0.3210
Optimal $\mathbf{C}^p$	1 0 0 0 1 0	1 1 0 0 1 0 1 0 1 0 0 1	1 1 0 0 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1	1 1 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 1 1 0 0 1 0 1	1 1 1 1 1 0 0 1 1 1 0 0 0 1 1 1 0 0 1 1 0 1 0 1 0 1 1 0 1 0 0 1 0 1 0 1 0 0 0 1 0 1

Table 2: *Optimal matrices* $\mathbf{C}^p$ *for small* S. Note that the lower bound given by Eq. (13) is attained only for $S = 4$, *i.e.*, for the smallest Hadamard-based construction of $\mathbf{C}^p$.

Corollary 1. Multiplexing S intensities requires $F \geq S - 1$ frames.

Corollary 2. The minimal configuration for fully-constrained reconstruction at a pixel p is $F = 2$ frames, $S = 3$ sub-frames per frame, and $S = 3$ linearly-independent illumination vectors of dimension $L \geq 3$. The next-highest configuration is 3 frames, 4 subframes/illumination vectors.

We now seek the optimal $(S - 1) \times S$ matrix $\mathbf{C}^p$, *i.e.*, the matrix that maximizes the SNR of the demultiplexed intensities in Eq. (9). Lemma 1 extends the lower-bound analysis of Ratner *et al.* [45] to obtain a lower bound on the mean-squared error (MSE) of two-bucket multiplexing [32]:

Lemma 1. For every multiplexing matrix $\mathbf{W}$, the MSE of the best unbiased linear estimator satisfies the lower bound

$$\text{MSE} \;=\; \frac{\sigma^2}{S}\text{trace}\left[\left(\mathbf{W}'\mathbf{W}\right)^{-1}\right] \geq 2\sigma^2\frac{(S-1)^2+1}{(S-1)S^2} \quad . \tag{13}$$

Although Lemma 1 does not provide an explicit construction, it does ensure the optimality of $\mathbf{W}$ matrices whose MSE is the lower bound. We used this observation to verify the optimality of matrices derived from the standard Hadamard construction [31]:

Proposition 1 Let $\mathbf{C}^p = \frac{1}{2}(\widetilde{\mathbf{H}}+1)$ where $\widetilde{\mathbf{H}}$ is derived from the $S \times S$ Hadamard matrix by removing its row of ones to create an $(S - 1) \times S$ matrix. The bucket-multiplexing matrix $\mathbf{W}$ defined by $\mathbf{C}^p$ is optimal.

The smallest S for which Proposition 1 applies are $S = 4$ and $S = 8$. Since our main goal is one-shot acquisition, optimal matrices for other small values of S are also of significant interest. To find them, we conducted a brute-force search over the space of small $(S-1) \times S$ binary matrices to find the ones with the lowest MSE. These matrices are shown in Table 2. See Fig. 6(a),(b) and [32] for an initial empirical SNR analysis.

5 One-Shot Shape from Two-Bucket Illumination Mosaics

We use three different ways of estimating shape from a two-bucket illumination mosaic:

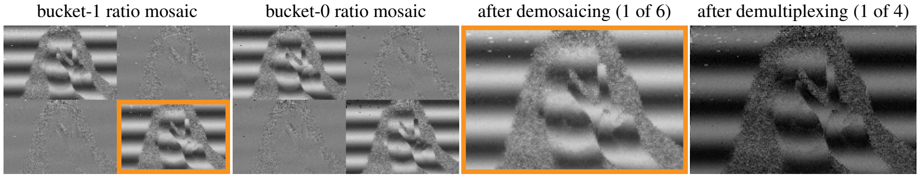


Fig. 5: *Processing ratio mosaics.* **Left to right:** Intermediate results of the BRD reconstruction procedure of Sec. 5, starting from the raw C2B frame shown in Fig. 2, Step 1. In contrast to the result of Steps 2 and 3 in Fig. 2, the images above are largely unaffected by albedo variations.

1. *Intensity demosaicing (ID):* treat the intensities in a mosaic tile as separate "imaging dimensions" for the purpose of demosaicing; upsample these intensities by applying either an RGB demosaicing algorithm to three of these dimensions at a time, or by using a more general assorted-pixel procedure [12, 54] that takes all of them into account; demultiplex the $2F$ upsampled images using Eq. (9); and apply any of the estimation methods in Sec. 3 to the result. Fig. 2 illustrates this approach.
2. *Bucket-ratio demosaicing (BRD):* apply Eq. (5) to each pixel in the mosaic to obtain two albedo-invariant "ratio mosaics"; demosaic and demultiplex them; and compute 3D shape using the ratio constraint of Sec. 3. See Fig. 5 for an example.
3. *No demosaicing (ND):* instead of upsampling, treat each mosaic tile as a "superpixel" whose unknowns (*i.e.*, normal, disparity, *etc.*) do not vary within the tile; compute one shape estimate per tile using any of the methods of Sec. 3.

Performance evaluation of one-shot photometric stereo on synthetic data. Figs. 6(c) and (d) analyze the effective resolution and albedo invariance of normal maps computed by several combinations of methods from Secs. 3 and 5, plus two more—*Baseline*, which applies basic photometric stereo to three full-resolution images; and *Color*, the one-shot color photometric stereo technique in [23]. To generate synthetic data, we (1) generated scenes with random spatially-varying normal maps and RGB albedo maps, (2) applied a spatial low-pass filter to albedo maps and the spherical coordinates of normal maps, (3) rendered them to create three sets of images—a grayscale C2B frame; three full-resolution grayscale images; and a Bayer color mosaic—and (4) added zero-mean Gaussian noise to each pixel, corresponding to a peak SNR of 30dB. Since all calculations except demosaicing are done per pixel, any frequency-dependent variations in performance must be due to this upsampling step. Our simulation results do match the intuition that performance should degrade for very high normal map frequencies regardless of the type of neighborhood processing. For spatial frequencies up to 0.3 the Nyquist limit, however, one-shot C2B imaging confers a substantial performance advantage. A similar evaluation for structured-light triangulation can be found in [32].

6 Live 3D Imaging with a C2B Camera

Experimental conditions. Both C2B frame acquisition and scene reconstruction run at 20Hz for all experiments, using $F = 3, S = 4$, the corresponding optimal $\mathbf{C}^p$ from

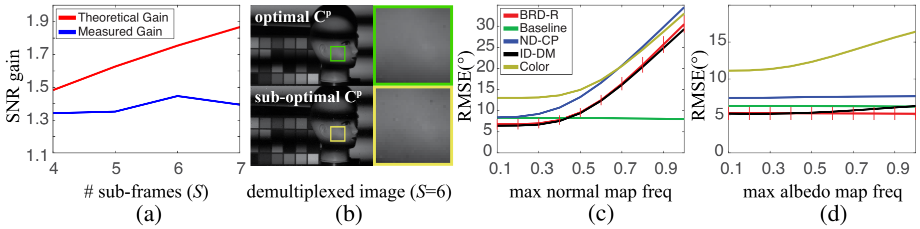
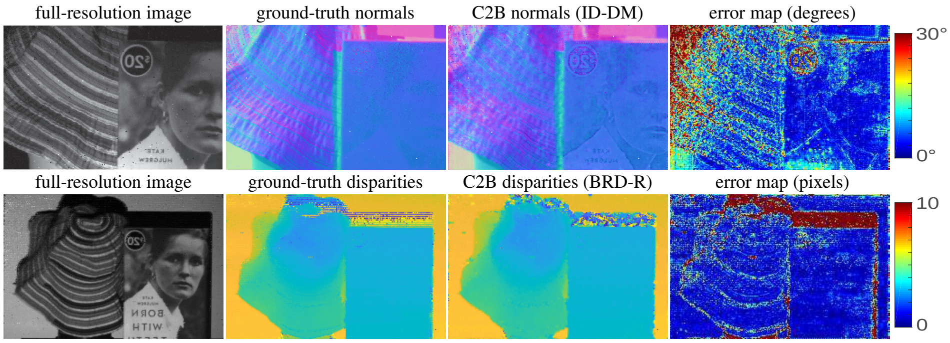


Fig. 6: (a) Optimal versus sub-optimal multiplexing. We applied bucket multiplexing to the scene shown in (b) and empirically measured the average SNR of demultiplexed images when (1) $\mathbf{C}^p$ is given by Table 2 and (2) $\mathbf{C}^p = [\mathbf{1}_{(S-1)\times(S-1)}\ \mathbf{0}]$, which is a non-degenerate and sub-optimal matrix according to Proposition 1 ($\mathbf{1}_{(S-1)\times(S-1)}$ is the identity matrix). The ratio of these SNRs is shown in blue, suggesting that SNR gains are possible. **(b)** One out of S demultiplexed images obtained with each $\mathbf{C}^p$. The optimal $\mathbf{C}^p$ yielded visibly less noisy images (please zoom in to the electronic copy). **(c)** Angular root-mean-squared error (RMSE) of normal estimates as a function of the normal map's highest spatial frequency. Frequency 1.0 corresponds to the Nyquist limit. The highest spatial frequency of albedos was set to 0.3 the Nyquist limit. **(d)** Angular error as a function of the spatial frequency of the albedo map, with the maximum spatial frequency of the normal map set to 0.3 the Nyquist limit. Line colors are as indicated in (c).

Table 2, and the 2×2 mosaic tile defined in Sec. 2.1. C2B frames are always processed by the same sequence of steps—demosaicing, demultiplexing and per-pixel reconstruction. For structured light, we fit an 8mm Schneider Cinegon $f/1.4$ lens to our camera with its aperture set to $f/2$, and use a TI LightCrafter for projecting 684×608-pixel, 24-gray-level patterns at a rate of $S\times$20Hz in sync with sub-frames. The stereo baseline was approximately 20cm, the scene was $1.1\sim1.5$m away, and the cosine frequency was 5 for all patterns and experiments. For photometric stereo we switch to a 23mm Schneider APO-Xenoplan $f/1.4$ lens to approximate orthographic imaging conditions, and illuminate a scene $2\sim3$m away with four sub-frame synchronized Luxdrive 7040 Endor Star LEDs, fitted with 26.5mm Carclo Technical Plastics lenses.

Quantitative experiments. Our goal was to compare the 3D accuracy of one-shot C2B imaging against that of full-resolution sequential imaging—using the exact same system and algorithms. Fig. 7 shows the static scenes used for these experiments, along with example reconstructions for photometric stereo and structured light, respectively. The "ground truth," which served as our reference, was computed by averaging 1000 sequentially-captured, bucket-1 images per illumination condition and applying the same reconstruction algorithm to the lower-noise, averaged images. To further distinguish the impact of demosaicing from that of sensor-specific non-idealities, we also compute shape from a *simulated* C2B frame; to create it we spatially multiplex the S averaged images computationally in a way that simulates the operation of our C2B sensor. Row 3 of Fig. 7 shows some of these comparisons for structured light. The BRD-R method, coupled with OpenCV's demosaicing algorithm, yields the best performance in this case, corresponding to a disparity error of 4%. See [32] for more details and additional results. **Reconstructing dynamic scenes.** Fig. 8 shows several examples.



demosaicing algorithm	**input images**	**ID-DM (RMSE)**	**BRD-R (RMSE)**
MATLAB demosaic()	one C2B frame simulated C2B	*6.442* 5.688	*5.223* 4.588
OPENCV demosaicing (BayerBG2RGB_EA)	one C2B frame simulated C2B	*5.778* 4.910	***4.855*** 4.462
Demosaic Net [29]	one C2B frame simulated C2B	*5.319* 4.719	*4.874* 4.434

Fig. 7: *Quantitative experiments for photometric stereo (Row 1) and structured light (Rows 2, 3).* Per-pixel unit normals $\mathbf{n}$ are visualized by assigning them the RGB color vector $0.5\mathbf{n} + 0.5$.

7 Concluding Remarks

Our experiments relied on some of the very first images from a C2B sensor. Issues such as fixed-pattern noise; slight variations in gain across buckets and across pixels; and other minor non-idealities do still exist. Nevertheless, we believe that our preliminary results support the claim that 3D data are acquired at near-sensor resolution.

We intentionally used raw, unprocessed intensities and the simplest possible approaches for demosaicing and reconstruction. There is no doubt that denoised images and more advanced reconstruction algorithms could improve reconstruction performance considerably. Our use of generic RGB demosaicing software is also clearly sub-optimal, as their algorithms do not take into account the actual correlations that exist across C2B pixels. A prudent approach would be to train an assorted-pixel algorithm on precisely such data.

Last but certainly not least, we are particularly excited about C2B cameras sparking new vision techniques that take full advantage of their advanced imaging capabilities.

Acknowledgements. We gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada under the RGPIN, RTI and SGP programs, and of DARPA under the REVEAL program. We also wish to thank Hui Feng Ke and Gilead Wolf Posluns for FPGA programming related to the C2B sensor, Sarah Anne Kushner for help with live imaging experiments, and Michael Brown, Harel Haim and the anonymous reviewers for their many helpful comments and suggestions on earlier versions of this manuscript.

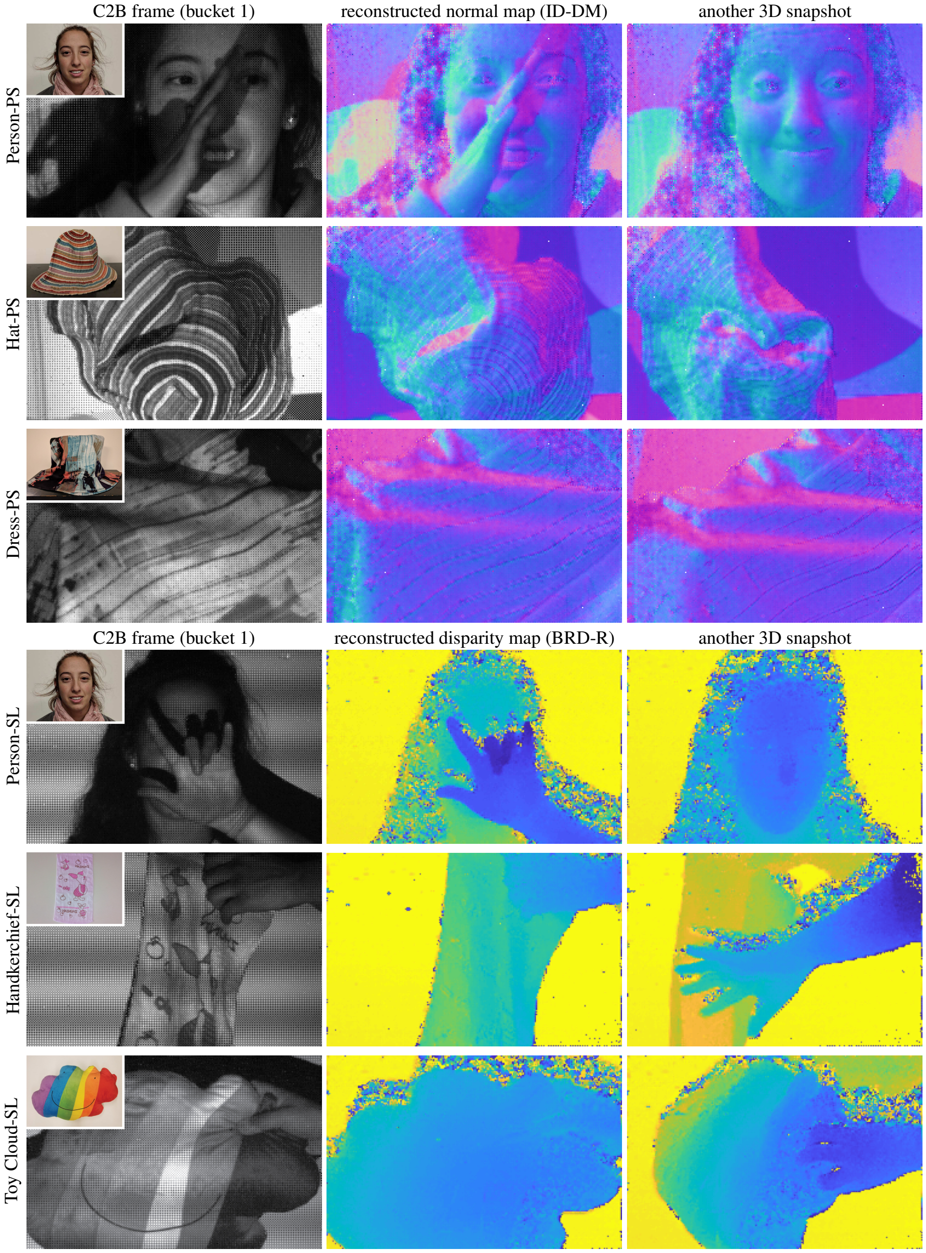


Fig. 8: *Live 3D acquisition experiments for photometric stereo (top) and structured light (bottom).* Scenes were chosen to exhibit significant albedo, color, normal and/or depth variations, as well as discontinuities. For reference, color photos of these scenes are shown as insets in Column 1. Qualitatively, reconstructions appear to be consistent with the scenes' actual 3D geometry except in regions of low albedo (*e.g.*, hair) or cast shadows.

References

1. Lange, R., Seitz, P.: Solid-state time-of-flight range camera. IEEE J. Quantum Electron. **37**(3), 390–397 (2001)
2. Bamji, C.S., O'Connor, P., Elkhatib, T., Mehta, S., Thompson, B., Prather, L.A., Snow, D., Akkaya, O.C., Daniel, A., Payne, A.D., Perry, T., Fenton, M., Chan, V.H.: A 0.13 μm CMOS System-on-Chip for a 512x424 Time-of-Flight Image Sensor With Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC. IEEE J. Solid-State Circuits **50**(1), 303–319 (2015)
3. Newcombe, R.A., Fox, D., Seitz, S.: DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In: Proc. IEEE CVPR (2015)
4. Heide, F., Hullin, M.B., Gregson, J., Heidrich, W.: Low-budget Transient Imaging Using Photonic Mixer Devices. In: Proc. ACM SIGGRAPH (2013)
5. Kadambi, A., Bhandari, A., Whyte, R., Dorrington, A., Raskar, R.: Demultiplexing illumination via low cost sensing and nanosecond coding. In: Proc. IEEE ICCP (2014)
6. Shrestha, S., Heide, F., Heidrich, W., Wetzstein, G.: Computational imaging with multi-camera time-of-flight systems. In: Proc. ACM SIGGRAPH (2016)
7. Callenberg, C., Heide, F., Wetzstein, G., Hullin, M.B.: Snapshot difference imaging using correlation time-of-flight sensors. In: Proc. ACM SIGGRAPH Asia (2017)
8. Lichtsteiner, P., Posch, C., Delbruck, T.: A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. IEEE J. Solid-State Circuits **43**(2), 566–576 (2008)
9. Kim, H., Leutenegger, S., Davison, A.J.: Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera. In: Proc. ECCV (2016)
10. Matsuda, N., Cossairt, O., Gupta, M.: MC3D: Motion Contrast 3D Scanning. In: Proc. IEEE ICCP (2015)
11. Jang, J., Yoo, Y., Kim, J., Paik, J.: Sensor-Based Auto-Focusing System Using Multi-Scale Feature Extraction and Phase Correlation Matching. Sensors **15**(3) (2015)
12. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.K.: Generalized Assorted Pixel Camera: Post-capture Control of Resolution, Dynamic Range, and Spectrum. IEEE-TIP **19**(9), 2241–2253 (2010)
13. Zhang, J., Etienne-Cummings, R., Chin, S., Xiong, T., Tran, T.: Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure. Opt. Express **24**(8), 9013–9024 (2016)
14. Sonoda, T., Nagahara, H., Endo, K., Sugiyama, Y., Taniguchi, R.: High-speed imaging using CMOS image sensor with quasi pixel-wise exposure. In: Proc. IEEE ICCP (2016)
15. Baraniuk, R.G., Goldstein, T., Sankaranarayanan, A.C., Studer, C., Veeraraghavan, A., Wakin, M.B.: Compressive Video Sensing: Algorithms, architectures, and applications. IEEE Signal Proc. Mag. **34**(1), 52–66 (2017)
16. Fossum, E.R., Hondongwa, D.B.: A Review of the Pinned Photodiode for CCD and CMOS Image Sensors. IEEE J. Electron Devices **2**(3), 33–43 (2014)
17. Hitomi, Y., Gu, J., Gupta, M., Mitsunaga, T., Nayar, S.K.: Video from a single coded exposure photograph using a learned over-complete dictionary. In: Proc. IEEE ICCV (2011)
18. O'Toole, M., Mather, J., Kutulakos, K.N.: 3D Shape and Indirect Appearance by Structured Light Transport. IEEE T-PAMI **38**(7), 1298–1312 (2016)
19. Sheinin, M., Schechner, Y., Kutulakos, K.N.: Computational Imaging on the Electric Grid. In: Proc. IEEE CVPR (2017)
20. O'Toole, M., Achar, S., Narasimhan, S.G., Kutulakos, K.N.: Homogeneous codes for energy-efficient illumination and imaging. In: Proc. ACM SIGGRAPH (2015)
21. Heintzmann, R., Hanley, Q.S., Arndt-Jovin, D., Jovin, T.M.: A dual path programmable array microscope (PAM): simultaneous acquisition of conjugate and non-conjugate images. J. Microscopy **204**(2), 119–135 (2001)

22. Raskar, R., Agrawal, A., Tumblin, J.: Coded exposure photography: motion deblurring using fluttered shutter. In: Proc. ACM SIGGRAPH (2006)
23. Hernandez, C., Vogiatzis, G., Brostow, G.J., Stenger, B., Cipolla, R.: Non-rigid Photometric Stereo with Colored Lights. In: Proc. IEEE ICCV (2007)
24. Kim, H., Wilburn, B., Ben-Ezra, M.: Photometric stereo for dynamic surface orientations. In: Proc. ECCV (2010)
25. Fyffe, G., Yu, X., Debevec, P.: Single-shot photometric stereo by spectral multiplexing. In: Proc. IEEE ICCP (2011)
26. Van der Jeught, S., Dirckx, J.J.J.: Real-time structured light profilometry: a review. Optics and Lasers in Engineering **87**, 18–31 (2016)
27. Sagawa, R., Furukawa, R., Kawasaki, H.: Dense 3D Reconstruction from High Frame-Rate Video Using a Static Grid Pattern. IEEE T-PAMI **36**(9), 1733–1747 (2014)
28. Narasimhan, S.G., Koppal, S.J., Yamazaki, S.: Temporal Dithering of Illumination for Fast Active Vision. In: Proc. ECCV (2008)
29. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. In: Proc. ACM SIGGRAPH Asia (2016)
30. Heide, F., Steinberger, M., Tsai, Y.T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., Kautz, J., Pulli, K.: FlexISP: a flexible camera image processing framework. In: Proc. ACM SIGGRAPH Asia (2014)
31. Schechner, Y.Y., Nayar, S.K., Belhumeur, P.N.: Multiplexing for optimal lighting. IEEE T-PAMI **29**(8), 1339–1354 (2007)
32. Wei, M., Sarhangnejad, N., Xia, Z., Katic, N., Genov, R., Kutulakos, K.N.: Coded Two-Bucket Cameras for Computer Vision: Supplemental Document. In: Proc. ECCV (2018), also available at http://www.dgp.toronto.edu/C2B
33. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recognition **43**(8), 2666–2680 (2010)
34. Salvi, J., Pages, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition **37**(4), 827–849 (2004)
35. Woodham, R.J.: Photometric Method For Determining Surface Orientation From Multiple Images. Opt. Eng. **19**(1) (1980)
36. Sarhangnejad, N., Lee, H., Katic, N., O'Toole, M., Kutulakos, K.N., Genov, R.: CMOS Image Sensor Architecture for Primal-Dual Coding. In: Int. Image Sensor Workshop (2017)
37. Luo, Y., Mirabbasi, S.: Always-on CMOS image sensor pixel design for pixel-wise binary coded exposure. In: IEEE Int. Symp. on Circuits & Systems (2017)
38. Luo, Y., Ho, D., Mirabbasi, S.: Exposure-Programmable CMOS Pixel With Selective Charge Storage and Code Memory for Computational Imaging. IEEE Trans. Circuits Syst. **65**(5), 1555–1566 (2018)
39. Wan, G., Li, X., Agranov, G., Levoy, M., Horowitz, M.: CMOS Image Sensors With Multi-Bucket Pixels for Computational Photography. IEEE J. Solid-State Circuits **47**(4), 1031–1042 (2012)
40. Wilburn, B.S., Ben-Ezra, M.: Time Interleaved Exposures And Multiplexed Illumination. US Patent 9,100,581 (2015)
41. Wan, G., Horowitz, M., Levoy, M.: Applications of Multi-Bucket Sensors to Computational Photography. Tech. rep., Stanford Computer Graphics Lab (2012)
42. Seo, M.W., Shirakawa, Y., Masuda, Y., Kawata, Y., Kagawa, K., Yasutomi, K., Kawahito, S.: 4.3 A programmable sub-nanosecond time-gated 4-tap lock-in pixel CMOS image sensor for real-time fluorescence lifetime imaging microscopy. In: Proc. IEEE ISSCC (2017)
43. Yoda, T., Nagahara, H., Taniguchi, R.i., Kagawa, K., Yasutomi, K., Kawahito, S.: The Dynamic Photometric Stereo Method Using a Multi-Tap CMOS Image Sensor. Sensors **18**(3) (2018)

44. Wetzstein, G., Ihrke, I., Heidrich, W.: On Plenoptic Multiplexing and Reconstruction. Int. J. Computer Vision **101**(2), 384–400 (2013)
45. Ratner, N., Schechner, Y.Y., Goldberg, F.: Optimal multiplexed sensing: bounds, conditions and a graph theory link. Opt. Express **15**(25), 17072–17092 (2007)
46. Brown, C.M.: Multiplex Imaging and Random Arrays. Ph.D. thesis, University of Chicago (1972)
47. Ratner, N., Schechner, Y.Y.: Illumination Multiplexing within Fundamental Limits. In: Proc. IEEE CVPR (2007)
48. Nonoyama, M., Sakaue, F., Sato, J.: Multiplex Image Projection Using Multi-band Projectors. In: IEEE Workshop on Color and Photometry in Computer Vision (2013)
49. Mitra, K., Cossairt, O.S., Veeraraghavan, A.: A Framework for Analysis of Computational Imaging Systems: Role of Signal Prior, Sensor Noise and Multiplexing. IEEE T-PAMI **36**(10), 1909–1921 (2014)
50. Liu, Z., Shan, Y., Zhang, Z.: Expressive expression mapping with ratio images. In: Proc. ACM SIGGRAPH (2001)
51. Wang, L., Yang, R., Davis, J.: BRDF invariant stereo using light transport constancy. IEEE T-PAMI **29**(9), 1616–1626 (2007)
52. Pilet, J., Strecha, C., Fua, P.: Making Background Subtraction Robust to Sudden Illumination Changes. In: Proc. ECCV (2008)
53. Bayer, B.E.: Color imaging array. US Patent 3,971,065 (1976)
54. Narasimhan, S.G., Nayar, S.: Enhancing resolution along multiple imaging dimensions using assorted pixels. IEEE T-PAMI **27**(4), 518–530 (2005)
55. Queau, Y., Mecca, R., Durou, J.D., Descombes, X.: Photometric stereo with only two images: A theoretical study and numerical resolution. Image and Vision Computing **57**, 175–191 (2017)
56. Gupta, M., Nayar, S.K.: Micro Phase Shifting. In: Proc. IEEE CVPR (2012)