

1 The Two-Bucket Camera

1.1 Notations

The coded two-bucket (C2B) camera is a pixel-wise coded exposure camera that outputs two images in a single exposure.[1] Each pixel in the sensor has two photo-collecting site, i.e. the two *buckets*, as well as a 1-bit writable memory controlling which bucket is actively collecting light. It was shown previously that C2B camera is capable of one-shot 3D reconstruction by solving a simpler image demosaicing and illumination demultiplexing problem instead of a difficult 3D reconstruction problem. We summarize the following notations relevant to discussion

	Notation	Meaning
	F	number of video frames
	P	number of pixels
	S	number of sub-frames
	h, w	dimension of image
$P \times F \times S$	\mathbf{C}	code tensor
$P \times 1 \times S$	$\tilde{\mathbf{C}}$	1-frame code tensor that spatially multiplex F frame tensor \mathbf{C}
$F \times S$	\mathbf{C}^p	activity of bucket 0 pixel p cross all frames and sub-frames
$F \times S$	$\overline{\mathbf{C}}^p$	activity of bucket 1 pixel p cross all frames and sub-frames
$1 \times S$	\mathbf{c}_f^p	active bucket of pixel p in the sub-frames of frame f
$1 \times L$	\mathbf{l}_s	scene's illumination condition in sub-frame s of every frame
$P \times S$	$\mathbf{C}_f = [\mathbf{c}_1^p; \dots; \mathbf{c}_F^p]$	activity of bucket activity of all pixels across all sub-frames of f
$S \times L$	$\mathbf{L} = [\mathbf{l}_1; \dots; \mathbf{l}_S]$	time-varying illumination condition (same for all frames)
$2F \times S$	\mathbf{W}	optimal bucket multiplexing matrix
	\mathbf{t}^p	transport vector at pixel p
$F \times 1$	$\mathbf{i}^p, \hat{\mathbf{i}}^p$	measured two-bucket intensity at pixel p in F frames
$F \times 1$	r, \hat{r}	illumination ratios at pixel p in F frames
$F \times P$	$\mathbf{I} = [\mathbf{i}^1 \dots \mathbf{i}^P], \hat{\mathbf{I}}$	two-bucket image sequence in F frames
$P \times 2F$	$\mathbf{I} = [\mathbf{I}^T \hat{\mathbf{I}}^T]$	two-bucket image sequence
$P \times 2$	\mathbf{Y}	two-bucket illumination mosaic
$S \times 1$	\mathbf{i}^p	pixel intensity under S illuminations at pixel p
$P \times S$	$\mathbf{X} = [\mathbf{i}^1 \dots \mathbf{i}^P]^T$	pixel intensity under S illuminations
$2P \times 1$	$\mathbf{y} = \text{vec}(\mathbf{Y})$	vectorized two-bucket illumination mosaic
$SP \times 1$	$\mathbf{x} = \text{vec}(\mathbf{X})$	vectorized pixel intensity under S illuminations
$2P \times 2PF$	\mathbf{B}	subsampling linear map
$2P \times SP$	$\mathbf{A} = \mathbf{B}(\mathbf{W} \otimes \mathbf{I}_P)$	illumination multiplexing and subsampling linear map

Illumination ratios are albedo *quasi-invariant*, a property which can be exploited for downstream processing

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

1.2 The Forward Model

Subsampling Mapping Let $\mathbf{S} \in \{1, 2, \dots, F\}^P$ be a vector specifying how the one-frame code tensor $\tilde{\mathbf{C}}$ is constructed, i.e. $\tilde{\mathbf{c}}_1^p := \mathbf{c}_{\mathbf{S}_p}^p$, for all pixels p . We can view \mathbf{S} as a mask to construct a **Subsampling** linear map that maps vectorized two-bucket image sequences \mathbf{I} to the vectorized illumination mosaics \mathbf{Y} .

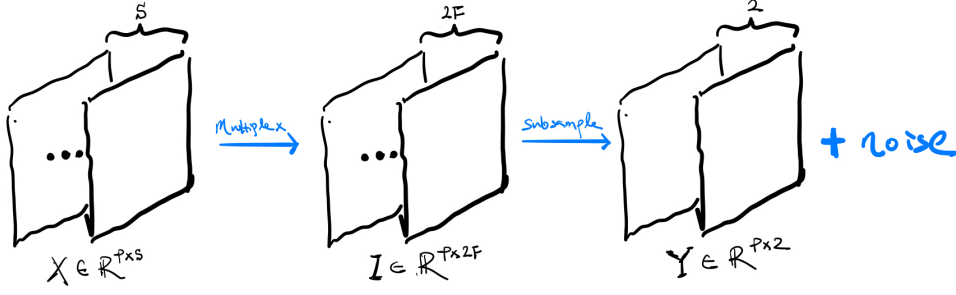


Figure 1: Image Formation Sketch

In particular, let $\mathbf{B}' \in \mathbb{R}^{P \times PF}$ and $\mathbf{B} \in \mathbb{R}^{2P \times 2PF}$ be defined as follows

$$\mathbf{B}' = [\text{diag} \mathbb{1}_{\{1\}}(\mathbf{S}) \quad \text{diag} \mathbb{1}_{\{2\}}(\mathbf{S}) \quad \cdots \quad \text{diag} \mathbb{1}_{\{F\}}(\mathbf{S})]$$

$$\mathbf{B} = \mathbf{I}_2 \otimes \mathbf{B}' = \begin{bmatrix} \mathbf{B}' & \mathbf{0} \\ \mathbf{0} & \mathbf{B}' \end{bmatrix}$$

Then we have the following relation between \mathbf{I} and \mathbf{Y} ,

$$\text{vec}(\mathbf{Y}) = \mathbf{B} \text{vec}(\mathbf{I}) \quad (1)$$

In essence, \mathbf{B} is a linear operator that trade spatial resolution (measures $\frac{1}{F}$ of the pixels for each frame) for temporal resolution (one two-bucket shot instead of acquiring F frames). We can think of a parallel in RGB color imaging, where bayer mosaic trade spatial resolution for spectral resolution. As an example when $F = 3$ and $P = 4$, the corresponding \mathbf{S} , when reshaped to dimension of a 2×2 image, and single image subsampling linear map \mathbf{B}' are given by

$$\mathbf{S} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \mathbf{B}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Image Formation Per-pixel image formation model is

$$\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix} = \begin{bmatrix} \mathbf{C}^p \\ \bar{\mathbf{C}}^p \end{bmatrix} \begin{bmatrix} \mathbf{1}_1 \mathbf{t}^p \\ \vdots \\ \mathbf{1}_S \mathbf{t}^p \end{bmatrix} = \begin{bmatrix} \mathbf{C}^p \\ \bar{\mathbf{C}}^p \end{bmatrix} \mathbf{i}^p$$

If bucket activity is same for all pixels and we use the optimal bucket multiplexing matrix \mathbf{W} , we can write the above linear relationship compactly for all pixels as

$$\mathbf{I} = \mathbf{X} \mathbf{W}^T \quad (2)$$

Linear Model As shown in Figure 1, illumination multiplexing and spatial subsampling can be combined to obtain a single linear function that maps images under S different illuminations \mathbf{X} to the two-bucket images \mathbf{Y} . From (1) and (2), there exists a linear relationship between \mathbf{x} and \mathbf{y} ,

$$\mathbf{y} = \mathbf{B} \text{vec}(\mathbf{I}) = \mathbf{B} \text{vec}(\mathbf{X} \mathbf{W}^T) = \mathbf{B}(\mathbf{W} \otimes \mathbf{I}_P) \text{vec}(\mathbf{X}) = \mathbf{A} \mathbf{x} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{2P \times SP}$ be a linear map that illumination multiplexes and subsamples \mathbf{X} ,

$$\mathbf{A} = \mathbf{B}(\mathbf{W} \otimes \mathbf{I}_P)$$

and $\mathbf{I}_P \in \mathbb{R}^{P \times P}$ is identity.

1.3 Image Processing Pipeline

The reconstruction pipeline is as follows

1. Use $\tilde{\mathbf{C}}$ for bucket activities and capture the two-bucket image \mathbf{Y}
2. upsample the images to full resolution images \mathbf{I}
3. demultiplex \mathbf{I} to obtain S full resolution images \mathbf{X} as a least squares solution to a (2)
4. use \mathbf{X} to solve for disparity and albedo

Step 2 and 3 are critical to downstream reconstructions. When $S = 3, S = 4$ and \mathbf{S} being analogous to bayer mask, we can upsample the images using standard demosaicing algorithms. However, it is not immediately obvious to extend demosaicing methods to support arbitrary \mathbf{S} , or more specifically, for scenarios where the spatial subsampling scheme is not bayer and when number of frames is not 3.