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
$P \times 1 \times S$	$\widetilde{\mathbf{C}}$	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
$P \times S$	$\mathbf{C}_f = [\mathbf{c}_1^p; \cdots ; \mathbf{c}_F^p]$	activity of bucket activity of all pixels across all sub-frames of f
$S \times L$	$\mathbf{L} = [\mathbf{l}_1; \cdots ; \mathbf{l}_S]$	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
$F \times P$	$\mathbf{I} = [\mathbf{i}^1 \cdots \mathbf{i}^P], \hat{\mathbf{I}}$	two-bucket image sequence in F frames
$P \times 2F$	$\boldsymbol{I} = [\mathbf{I}^T \; \hat{\mathbf{I}}^T]$	two-bucket image sequence
$P \times 2$	$\mathbf{Y}$	two-bucket illumination mosaic
$S \times 1$	$\boldsymbol{i}^p$	pixel intensity under S illuminations at pixel p
$P \times S$	$\mathbf{X} = [\boldsymbol{i}^1 \cdots \boldsymbol{i}^P]^T$	pixel intensity under S illuminations
$2P \times 1$	$\mathbf{y} = vec\,(\mathbf{Y})$	vectorized two-bucket illumination mosaic
$SP \times 1$	$\mathbf{x} = vec\,(\mathbf{X})$	vectorized pixel intensity under S illumiantions
$2P \times 2PF$	$\mathbf{B}$	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

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

1.2 The Forward Model

Subsampling Mapping Let $\mathbf{S} \in \{1, 2, \cdots, F\}^P$ be a vector specifying how the one-frame code tensor $\widetilde{\mathbf{C}}$ 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 **S**ubsampling

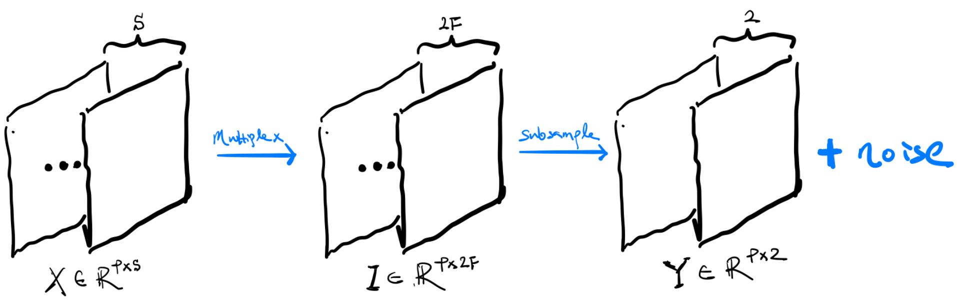

Figure 1: Image Formation Sketch

linear map that maps vectorized two-bucket image sequences $\boldsymbol{I}$ to the vectorized illumination mosaics $\mathbf{Y}$. In particular, let $\mathbf{B}' \in \mathbb{R}^{P\times PF}$ and $\mathbf{B} \in \mathbb{R}^{2P\times 2PF}$ be defined as follows

$$\mathbf{B}' = \begin{bmatrix} \mathbf{diag}\mathbb{1}_{\{1\}}(\mathbf{S}) & \mathbf{diag}\mathbb{1}_{\{2\}}(\mathbf{S}) & \cdots & \mathbf{diag}\mathbb{1}_{\{F\}}(\mathbf{S}) \end{bmatrix}$$
$$\mathbf{B} = \mathbf{I}_2 \otimes \mathbf{B}' = \begin{bmatrix} \mathbf{B}' & \mathbf{0} \\ \mathbf{0} & \mathbf{B}' \end{bmatrix}$$

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

$$vec\,(\mathbf{Y}) = \mathbf{B}vec\,(\boldsymbol{I}) \tag{1}$$

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

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

Image Formation Per-pixel image formation model is

$$\begin{bmatrix} \mathbf{i}^p \\ \hat{\mathbf{i}}^p \end{bmatrix} = \begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix} \begin{bmatrix} \mathbf{l}_1 \mathbf{t}^p \\ \vdots \\ \mathbf{l}_S \mathbf{t}^p \end{bmatrix} = \begin{bmatrix} \mathbf{C}^p \\ \overline{\mathbf{C}}^p \end{bmatrix} \boldsymbol{i}^p$$

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

$$\boldsymbol{I} = \mathbf{X}\mathbf{W}^T \tag{2}$$

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 $\mathbf{y}$,

$$\mathbf{y} = \mathbf{B}vec\,(\boldsymbol{I}) = \mathbf{B}vec\left(\mathbf{X}\mathbf{W}^T\right) = \mathbf{B}(\mathbf{W} \otimes \mathbf{I}_P)vec\,(\mathbf{X}) = \mathbf{A}\mathbf{x} \tag{3}$$

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 The Inverse Problem

The reconstruction pipeline is as follows

1. Use $\widetilde{\mathbf{C}}$ for bucket activities and capture the two-bucket image $\mathbf{Y}$
2. upsample the images to full resolution images $\boldsymbol{I}$
3. demultiplex $\boldsymbol{I}$ 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 reconstuctions. 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. One approach which we consider later involves the following steps

1. Use $\widetilde{\mathbf{C}}$ for bucket activities and capture the two-bucket image $\mathbf{Y}$
2. Recover full resolution images $\mathbf{X}$ under S illuminations from $\mathbf{Y}$ by solving a linear inverse problem
3. use $\mathbf{X}$ to solve for disparity and albedo

Jointly upsample and demultiplex enforces a prior knowledge of image formation. Instead of treating upsampling (recover $2F$ images $\boldsymbol{I}$ from 2 images $\mathbf{Y}$) and demultiplexing (recover S images $\mathbf{X}$ from $2F$ images $\boldsymbol{I}$) as distinct steps, we aim to recover $\mathbf{X}$ directy from $\mathbf{Y}$, in a single step, by solving a linear inverse problem.