

Project Report

Peiqi Wang

July 20, 2019

Contents

| | | |
|----------|-------------------------------------|----------|
| 1 | Abstract | 3 |
| 2 | The Two-Bucket Camera | 3 |
| 2.1 | Notations | 3 |
| 2.2 | Subsampling Mapping | 4 |
| 2.3 | Image Formation | 4 |
| 2.4 | Image Processing Pipeline | 5 |
| 3 | Problem Formulation | 5 |
| 3.1 | A Linear Inverse Problem | 5 |
| 3.2 | Image Priors | 6 |

1 Abstract

We aim to improve upon low level image processing pipeline for the coded two-bucket camera. Specifically, we aim to jointly upsample, demultiplex, and denoise two-bucket images to produce full resolution images under different illumination conditions for downstream reconstruction tasks.

2 The Two-Bucket Camera

2.1 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 |
| | 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

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

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

$$\begin{aligned} \mathbf{B}' &= [\mathbf{diag}\mathbb{1}_{\{1\}}(\mathbf{S}) \quad \mathbf{diag}\mathbb{1}_{\{2\}}(\mathbf{S}) \quad \dots \quad \mathbf{diag}\mathbb{1}_{\{F\}}(\mathbf{S})] \\ \mathbf{B} &= \mathbf{B}' \oplus \mathbf{B}' = \begin{bmatrix} \mathbf{B}' & \mathbf{0} \\ \mathbf{0} & \mathbf{B}' \end{bmatrix} \end{aligned}$$

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

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

We are motivated to think of an analogue where it is common place to perform spatial subsampling. In RGB color imaging, bayer mosaics trade spatial resolution for spectral resolution (R,G,B colors). We can find an analogous one-frame code tensor which generate illumination mosaics that trade spatial resolution for temporal resolution ($1, 2, \dots, F$ frames). As an example in case of $F = 3$ 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 simply

$$\begin{aligned} \mathbf{S} &= \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \\ \mathbf{B}' &= [\mathbf{diag}\mathbb{1}_{\{1\}}(\mathbf{S}) \quad \mathbf{diag}\mathbb{1}_{\{2\}}(\mathbf{S}) \quad \mathbf{diag}\mathbb{1}_{\{3\}}(\mathbf{S})] = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

2.3 Image Formation

Per-pixel image formation model is

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

If bucket activity is same for all pixels and we use the optimal bucket multiplexing matrix \mathbf{W} , then

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

2.4 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.

3 Problem Formulation

3.1 A Linear Inverse Problem

We consider the problem of recovering full resolution images \mathbf{X} under S illuminations from a two-bucket image \mathbf{Y} as an linear inverse problem. Let $\mathbf{A} \in \mathbb{R}^{2P \times SP}$ represent a linear map that illumination multiplexes and subsamples \mathbf{X} ,

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

where $\mathbf{I}_P \in \mathbb{R}^{P \times P}$ is identity. From (1) and (2), there exists a linear relationship between \mathbf{x} and \mathbf{y} ,

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

Note (3) is an underdetermined system. Given 2 images, we want to recover S images - the larger the number of subframes, the harder the recovery becomes. This asks for stronger prior knowledge of the underlying distribution of \mathbf{x} to restrict search space for solutions as S increases. Jointly upsample and demultiplex enforces a prior knowledge of image formation. Instead of treating upsampling (recover $2F$ images \mathbf{I} from 2 images \mathbf{Y}) and demultiplexing (recover S images \mathbf{X} from $2F$ images \mathbf{I}) as distinct steps, we aim to recover \mathbf{X} directly from \mathbf{Y} , in a single step, by solving the following unconstrained optimization problem,

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\rho(\mathbf{x}) \quad (4)$$

where $\rho : \mathbb{R}^{SP} \rightarrow \mathbb{R}$ is a regularizer for \mathbf{x} , the optimization variable. The problem (4) has a bayesian interpretation. Specifically, the ℓ_2 norm can be interpreted as a log-likelihood *data term* that captures the following probabilistic relationship between recovered image \mathbf{x} and observation \mathbf{y} ,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

where \mathbf{e} is the noise random variable, usually assumed to be Gaussian. The regularizer can be then interpreted as *prior* knowledge on the distribution of \mathbf{x} . The data term is continuous and fully differentiable in all of \mathbb{R}^n . Therefore, tractability of (4) usually depends on how well behaved ρ is. If ρ is convex but possibly non-smooth, e.g. $\rho(\mathbf{x}) = \|\mathbf{x}\|_1$, the problem can be efficiently solved with standard convex optimization methods like proximal gradients with guaranteed convergence and global optimality. [2] More realistic priors, i.e. regularizers that more precisely capture the prior knowledge of image distributions, might potentially make (4) a much harder problem that compromises convergence and optimality properties. [3]

3.2 Image Priors

The choice of regularization has been an important research topic in image processing. Handcrafted priors have been successful in a number of different image recovery tasks. For example, we can choose to enforce (1) the sparsity of \mathbf{x} with ℓ_1 norm in image deblurring [2] (2) total variation in image denoising [4] (3) cross-channel correlation in color image demosaicing [5] (4) dark channel prior in image dehazing [6], etc.

Recently, there has been interest and development in algorithm induced priors.

References

- [1] Mian Wei et al. “Coded Two-Bucket Cameras for Computer Vision”. en. In: *Computer Vision – ECCV 2018*. Ed. by Vittorio Ferrari et al. Vol. 11207. Cham: Springer International Publishing, 2018, pp. 55–73. ISBN: 978-3-030-01218-2 978-3-030-01219-9. DOI: [10.1007/978-3-030-01219-9_4](#).
- [2] A. Beck and M. Teboulle. “A Fast Iterative Shrinkage-Thresholding Algorithm with Application to Wavelet-Based Image Deblurring”. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Apr. 2009, pp. 693–696. DOI: [10.1109/ICASSP.2009.4959678](#).
- [3] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. “Deep Image Prior”. In: *arXiv:1711.10925 [cs, stat]* (Nov. 2017). arXiv: [1711.10925 \[cs, stat\]](#).
- [4] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. “Nonlocal Image and Movie Denoising”. en. In: *International Journal of Computer Vision* 76.2 (Feb. 2008), pp. 123–139. ISSN: 1573-1405. DOI: [10.1007/s11263-007-0052-1](#).
- [5] H. S. Malvar, Li-wei He, and R. Cutler. “High-Quality Linear Interpolation for Demosaicing of Bayer-Patterned Color Images”. In: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 3. May 2004, pp. iii–485. DOI: [10.1109/ICASSP.2004.1326587](#).
- [6] Raanan Fattal. “Single Image Dehazing”. In: *ACM SIGGRAPH 2008 Papers*. SIGGRAPH ’08. New York, NY, USA: ACM, 2008, 72:1–72:9. ISBN: 978-1-4503-0112-1. DOI: [10.1145/1399504.1360671](#).