1 Methods

1.1 Joint Demultiplex and Upsample

One approach which we can consider 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 phase and albedo using ZNCC decoding [1]

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. Jointly upsample and demultiplex enforces a prior knowledge of image formation and reduces accumulation of error at each image processing step [2]. This approach allows encoding arbitrary subsampling schemes in the forward operator and can be adapted to any frames F.

MAP Inference with Implicit Regularizer Assuming an isotropic Gaussian noise model ($\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ where $\mathbf{e} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$), and prior $p_\mathsf{x}(x) \propto \exp\left\{-(\lambda/\sigma^2)\Gamma(\mathbf{x})\right\}$ where $\Gamma : \mathbb{R}^{SP} \to \mathbb{R}$ is some regularizer for $\mathbf{x}$. Given noisy measurement $\mathbf{y}$, *max a posterior* estimate $\hat{\mathbf{x}}$ can be obtained by solving the following

$$\text{minimize}_\mathbf{x} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\Gamma(\mathbf{x}) \tag{1}$$

where λ is weight for the regularizer.

Optimization Using PnP/RED We solve (1) using PnP/RED [3, 4], which use denoisers in place of a proximal operator under a ADMM style updates. We can specialize the update equations to our problem. First we notice that $\mathbf{A}\mathbf{A}^T$ is diagonal when $\mathbf{W}$ is the optimal bucket multiplexing matrix constructed from Hadamard matrix specified in [5]. The x-update can be simplified [6]

$$\mathbf{x}^{k+1} = \tilde{\mathbf{x}} + \mathbf{A}^T\left[\frac{(\mathbf{y}-\mathbf{A}\tilde{x})_1}{\zeta_1+\rho}, \cdots, \frac{(\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}})_{SP}}{\zeta_{SP}+\rho}\right]^T \tag{2}$$

where ρ is the parameter for augmented lagrangian, $\tilde{\mathbf{x}} = \mathbf{z}^k - \mathbf{u}^k$, and $\boldsymbol{\zeta} = \mathbf{diag}\left(\mathbf{A}\mathbf{A}^T\right)$, which can be precomputed. (2) is fast because it consists of 2 sparse matix-vector multiply and a few element-wise vector operations. The update equations is then

$$\mathbf{x}^{k+1} = \tilde{\mathbf{x}} + \mathbf{A}^T\left[\frac{(\mathbf{y}-\mathbf{A}\tilde{x})_1}{\zeta_1+\rho}, \cdots, \frac{(\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}})_{SP}}{\zeta_{SP}+\rho}\right]^T \qquad \tilde{\mathbf{x}} = \mathbf{z}^k - \mathbf{u}^k \tag{3}$$

$$\mathbf{z}^{k+1} = \frac{1}{\rho + \lambda}\left(\lambda\mathcal{D}(\mathbf{z}^k) + \rho(\mathbf{x}^{k+1} + \mathbf{u}^k)\right) \tag{4}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1} \tag{5}$$

Note this method is general enough to allow for both multispectral imaging, structured light reconstruction, and perhaps a mix of the two modalities.

1.2 Correspondence Refinement

In applications where we are purely interested in structured light reconstruction. There are artifacts associated with spatial subsampling when number of illuminated patterns gets larger. Here we exploit a simple image formation model under active illumination to refine correspondence.
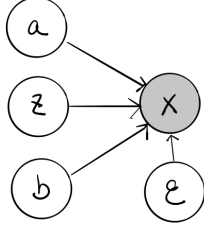
Figure 1: Reconstructed Images $\mathbf{X}$ and its relationship to albedo, phase, and ambient illumination.

Image Formation Under Active Illumination   Let $\mathbf{a} \in \mathbb{R}_+^P$ be albedo, $\mathbf{b} \in \mathbb{R}_+^P$ be ambient illumination, $\mathbf{z} \in [0,1]^P$ be correspondence/phase. We consider the following probabilistic model as depicted in Figure 1. In particular, we assume Gaussian noise $\boldsymbol{\epsilon} \in \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{I})$ capturing errors in model misspecification, then we have per-pixel image formation model

$$\boldsymbol{i}^p = \mathbf{a}^p \mathbf{l}(\mathbf{z}^p) + \mathbf{b}^p + \boldsymbol{\epsilon}^p \qquad p = 1, 2, \cdots, P \tag{6}$$

where $\mathbf{l} : [0,1] \to [0,1]^S$ where $\mathbf{l}(z) = (\mathbf{l}_1(z), \cdots, \mathbf{l}_S(z))$ is vector valued function that specifies illumination condition at a particular projector index z, normalized to lie in $[0,1]$. From perspective of structured light coding, $\mathbf{l}$ is the coding scheme for the system. It might be interesting to extend (6) to include effect of global illumination and defocus, similar to [7]. Note the relationship in (6) is nonlinear and nonconvex, due to coupling of albedo and illumination conditions. Given $\mathbf{z}$, the relationship becomes linear

$$vec\left(\mathbf{X}^T\right) = \begin{bmatrix} \boldsymbol{i}^1 \\ \vdots \\ \boldsymbol{i}^P \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{diagl}(\mathbf{z}^1) & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{diagl}(\mathbf{z}^P) & \mathbf{I} \end{bmatrix}}_{\mathbf{M}(\mathbf{z})} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}^1 \\ \vdots \\ \boldsymbol{\epsilon}^P \end{bmatrix} \tag{7}$$

MAP Inference of Correspondence   For arbitrary patterns, the MAP estimate of a discrete set of correspondences is very hard, as it involves a combinatorial problem over a very high dimensional space, in particular L^P possible combinations. For convenience of optimization, we assume $\mathbf{l}$ be a coordinate-wise continuous function, and we are interested in finding correspondence in the real numbers. A similar bayesian setup for solving depth and albedo is detailed in [8]. Assuming uniform prior for $\mathbf{a}, \mathbf{b}$, we have

$$\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{z}} = \arg\max_{\mathbf{a}, \mathbf{b}, \mathbf{z}} p_{\mathsf{a,b,z|x}}(\mathbf{a}, \mathbf{b}, \mathbf{z} | \mathbf{x}) = \arg\min_{\mathbf{a}, \mathbf{b}, \mathbf{z}} -\log p_{\mathsf{x|a,b,z}}(\mathbf{x} | \mathbf{a}, \mathbf{b}, \mathbf{z}) - \log p_{\mathsf{z}}(\mathbf{z}) \tag{8}$$

Assumption of Gaussian noise in (6) and $p_{\mathsf{z}}(\mathbf{z}) \propto \exp\left\{-(\lambda/\sigma_\epsilon^2) \|\mathbf{Gz}\|_1\right\}$ ($\mathbf{G} = [\mathbf{G}_x; \mathbf{G}_y] \in \mathbb{R}^{2P \times P}$ are the differencing matrix for computing total variation norm) yield the following optimization problem

$$\underset{\mathbf{a} \in \mathbb{R}_+^P, \mathbf{b} \in \mathbb{R}_+^P, \mathbf{z} \in [0,1]^P}{\text{minimize}} \quad \frac{1}{2} \sum_{p=1}^P \|\boldsymbol{i}^p - \mathbf{a}^p \mathbf{l}(\mathbf{z}^p) - \mathbf{b}^p\|_2^2 + \lambda \|\mathbf{Gz}\|_1 \tag{9}$$

Without assumption of coordinate-wise convexity of $\mathbf{l}$, the objective function is subject to bad local minimas. Idea is to seed the nonlinear optimization procedure with ZNCC decoding of $\mathbf{x}$ solved by (5) and hope that this refinement will get us to the destination. To make optimization easier (9), similar to [9], we use block coordinate descent to optimize for

$\mathbf{a}, \mathbf{b}$ and $\mathbf{z}$ in an alternating fashion,

$$\mathbf{a}_{k+1}, \mathbf{b}_{k+1} = \arg\min_{\mathbf{a} \in \mathbb{R}^P_+, \mathbf{b} \in \mathbb{R}^P_+} \left\| \mathbf{M}(\mathbf{z}_k) \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - vec\left(\mathbf{X}^T\right) \right\|_2^2 \tag{10}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z} \in [0,1]^P} \frac{1}{2} \sum_{p=1}^{P} \left\| \boldsymbol{i}^p - \mathbf{a}_{k+1}^p \mathbf{l}(\mathbf{z}^p) - \mathbf{b}_{k+1}^p \right\|_2^2 + \lambda \left\| \mathbf{Gz} \right\|_1 \tag{11}$$

where (10) can be solved with any sparse linear solver, and (11) can be solved using some nonlinear solver such as lbfgs as it emits computable sub-gradients

$$-\sum_{s=1}^{S} \left(\boldsymbol{i}^p - \mathbf{a}^p \mathbf{l}_s(\mathbf{z}^p) - \mathbf{b}^p \right) \mathbf{a}^p g_s(\mathbf{z}^p) + \lambda \left(\mathbf{G}^T \mathsf{sgn}\left(\mathbf{Gz}\right) \right)_p \in \partial_{\mathbf{z}^p}(\cdot) \tag{12}$$

where $g_s(\mathbf{z}^p) \in \partial_z \mathbf{l}_s(\mathbf{z}^p)$. In case when $\mathbf{l}$ is coordinate-wise convex, the objective for (9) is a bi-convex problem, whereby block coordinate descent converges monotonically to stationary points [10]. Optimizing multi-convex function has been explored in applications such as time-of-flight reconstruction [11].

Applicable Coding Schemes Empirically, the choice of $\mathbf{l}$ is very important for optimization. Some coding schemes are more well behaved, for example spatial sinusoids with frequency of 1 shifted multiple times. Spatial sinusoids with spatial frequency > 2, micro phase shifting code [7], Hamiltonian code [12], a la carte code [1] all give rise to objective functions in (9) with bad local minima. One potential way to alleviate this is perhaps to design coding schemes that are coordinate-wise convex, which will ensure that the objective function is bi-convex.

1.3 Revised Noise Model

We might argue the Gaussian model assumption does actually reflect the physical reality. In particular, sensor noise happens when the images $\mathbf{Y}$ are acquired, as depicted in Figure 2.
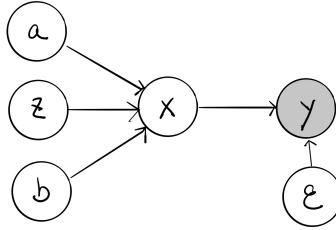

Figure 2: c2b measurements $\mathbf{Y}$ and its relationship to albedo, phase, and ambient illumination

1.4 Experimental Results

Ground Truth Ground truth phase/disparity is obtained by computing ZNCC decoding on sinusoids of periods 1,2,5,17,31, each shifted 30 times. Each measurement under a unique illumination pattern is stacked form 250 noisy images.

Empirical Performance Upper Bound In Figure 3, we experimentally determine what is empirically possible for the decoding algorithm, without either noise nor recosntruction error, when given measurement under different number of phase shifts, assuming sinusoidal pattern with period 1.
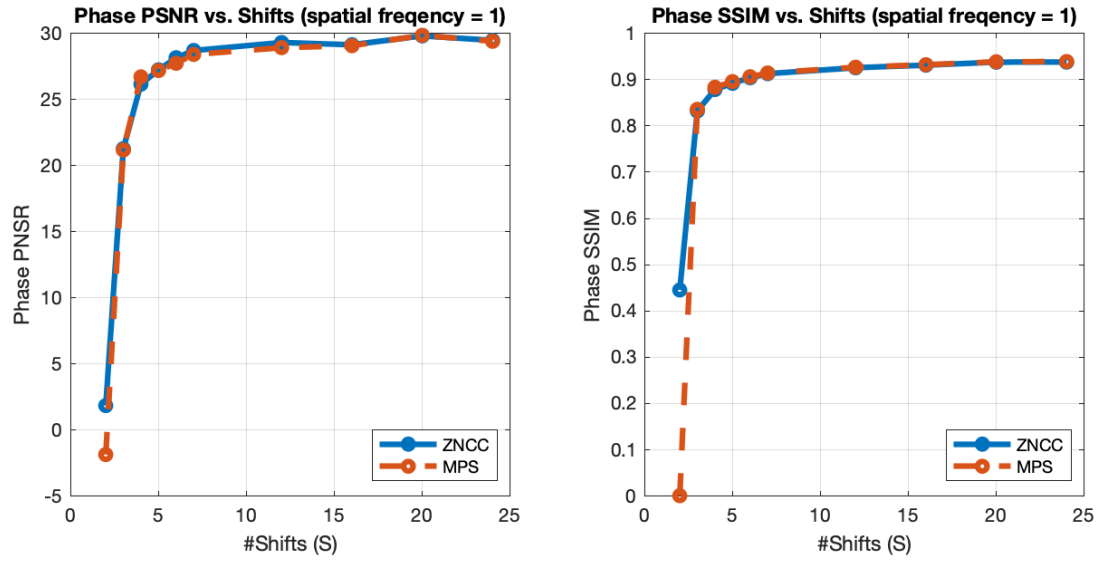
3

Figure 3: PSNR/SSIM computed by comparing phase obtained by ZNCC/MPS decoding on stacked noiseless images for sinusoidal pattern with period of 1 with varying number of phase shifts to the ground-truth phase. Both decoder performed similarly. We see large improvement when $S = 3, 4$ and minimal improvement when $S > 7$.