Project Report

Peiqi Wang

Wednesday 10th June, 2020

Contents

1 Abstract

We aim to improve upon low level image proecssing 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
$P \times 1 \times S$	$\widetilde{\mathbf{C}}$	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

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

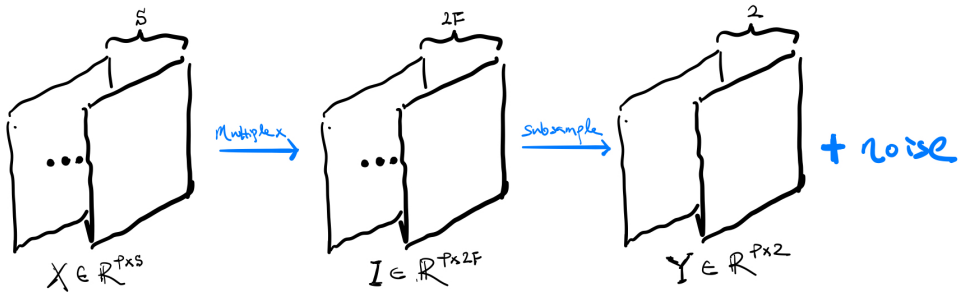

Figure 1: Image Formation Sketch

2.2 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 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

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

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 $\mathbf{y}$,

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

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.

2.3 Image Processing Pipeline

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 phase and albedo using ZNCC decoding [2]

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.

3 Methods

3.1 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 [2]

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 [3]. 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

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\Gamma(\mathbf{x}) \tag{4}$$

where λ is weight for the regularizer.

Optimization Using PnP/RED We solve (4) using PnP/RED [4, 5], 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 [1]. The x-update can be simplified [6]

$$\mathbf{x}^{k+1} = \tilde{\mathbf{x}} + \mathbf{A}^T \left[\frac{(\mathbf{y}-\mathbf{A}\tilde{x})_1}{\boldsymbol{\zeta}_1+\rho}, \cdots, \frac{(\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}})_{SP}}{\boldsymbol{\zeta}_{SP}+\rho}\right]^T \tag{5}$$

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. (5) is fast because it consists of 2 sparse matix-vector multiply and a few element-wise vector operations. The update equations is then

$$\mathbf{x}^{k+1} = \tilde{\mathbf{x}} + \mathbf{A}^T \left[\frac{(\mathbf{y}-\mathbf{A}\tilde{x})_1}{\boldsymbol{\zeta}_1+\rho}, \cdots, \frac{(\mathbf{y}-\mathbf{A}\tilde{\mathbf{x}})_{SP}}{\boldsymbol{\zeta}_{SP}+\rho}\right]^T \qquad \tilde{\mathbf{x}} = \mathbf{z}^k - \mathbf{u}^k \tag{6}$$

$$\mathbf{z}^{k+1} = \frac{1}{\rho+\lambda}\left(\lambda\mathcal{D}(\mathbf{z}^k) + \rho(\mathbf{x}^{k+1} + \mathbf{u}^k)\right) \tag{7}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1} \tag{8}$$

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

3.1.1 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 2, 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.

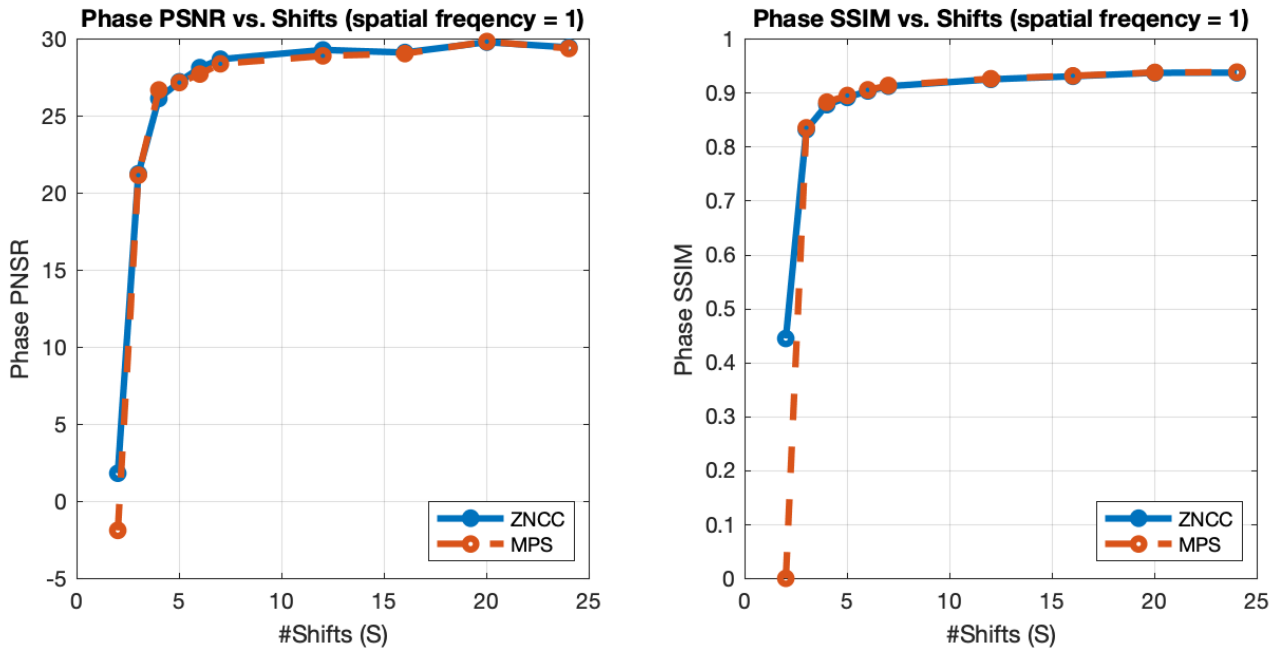

Figure 2: PSNR/SSIM computed by comparing phase obtained by ZNCC/MPS decoding on stacked noiseless images for sinusoidal pattern with period of 1 with varying number of phase shifts to the ground-truth phase. Both decoder performed similarly. We see large improvement when $S = 3, 4$ and minimal improvement when $S > 7$.

3.2 Correspondence Refinement

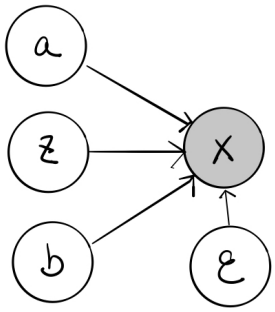

Figure 3: Signal and its relationship to albedo, phase, and ambient illumination.

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.

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 3. 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

$$\boldsymbol{i}^p = \mathbf{a}^p \mathbf{l}(\mathbf{z}^p) + \mathbf{b}^p + \boldsymbol{\epsilon}^p \qquad p = 1, 2, \cdots, P \tag{9}$$

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. Note the relationship in (9) is nonlinear and nonconvex, due to coupling of albedo and illumination conditions. Given $\mathbf{z}$, the relationship becomes linear

$$vec\left(\mathbf{X}^T\right) = \begin{bmatrix} \boldsymbol{i}^1 \\ \vdots \\ \boldsymbol{i}^P \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{diagl}(\mathbf{z}^1) & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{diagl}(\mathbf{z}^P) & \mathbf{I} \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}}_{\mathbf{v}} \tag{10}$$

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.

$$\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{z}} = \underset{a \mathbb{R}_+^P, b \in \mathbb{R}_+^P, z \in [L]^P}{\arg\max} p_{\mathsf{a},\mathsf{b},\mathsf{z}|\mathsf{x}}(a, b, z|x) = \underset{a \in \mathbb{R}_+^P, b \in \mathbb{R}_+^P, z \in [L]^P}{\arg\min} p_{\mathsf{z}|\mathsf{a},\mathsf{b},\mathsf{z}}(x|a, b, z) p_{\mathsf{a}}(a) p_{\mathsf{b}}(b) p_{\mathsf{z}}(z) \tag{11}$$

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 [7]. Assumption of Gaussian noise in (9) and uniform prior for $\mathbf{a}, \mathbf{b}$ and $p_{\mathsf{z}}(z) \propto \exp\left\{-(\lambda/\sigma_\epsilon^2) \left\|\mathbf{G}\mathbf{z}\right\|_1\right\}$ ($\mathbf{G} = [\mathbf{G}_x; \mathbf{G}_y] \in \mathbb{R}^{2P \times P}$ are the differencing matrix for computing total variation norm) yield the following optimization problem

$$\underset{\mathbf{a} \mathbb{R}_+^P, \mathbf{b} \in \mathbb{R}_+^P, \mathbf{z} \in [0,1]^P}{\text{minimize}} \quad \frac{1}{2} \sum_{p=1}^{P} \left\|\boldsymbol{i}^p - \mathbf{a}^p \mathbf{l}(\mathbf{z}^p) - \mathbf{b}^p\right\|_2^2 + \lambda \left\|\mathbf{G}\mathbf{z}\right\|_1 \tag{12}$$

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 (8) and hope that this refinement will get us to the destination. To make optimization easier (12), similar to [8], we use block coordinate descent to optimize for $\mathbf{a}, \mathbf{b}$ and $\mathbf{z}$ in an alternating fashion,

$$\mathbf{a}^{k+1}, \mathbf{b}^{k+1} = \underset{\mathbf{v}}{\arg\min} \left\|\mathbf{M}\mathbf{v} - vec\left(\mathbf{X}^T\right)\right\|_2^2 \tag{13}$$

$$\mathbf{z}^{k+1} = \underset{\mathbf{z} \in [0,1]^P}{\arg\min} \frac{1}{2} \sum_{p=1}^{P} \left\|\boldsymbol{i}^p - \mathbf{a}^p \mathbf{l}(\mathbf{z}^p) - \mathbf{b}^p\right\|_2^2 + \lambda \left\|\mathbf{G}\mathbf{z}\right\|_1 \tag{14}$$

where (13) can be solved with any sparse linear solver, and (14) can be solved using some nonlinear solver such as lbfgs as it emits computable sub-gradients

$$\frac{\partial}{\partial \mathbf{z}}(\cdot) = \sum_{p=1}^{P} \sum_{s=1}^{S} \left(\boldsymbol{i}^p - \mathbf{a}^p \mathbf{l}_s(\mathbf{z}^p) - \mathbf{b}^p\right) \mathbf{a}^p \frac{d}{dz} \mathbf{l}_s(\mathbf{z}^p) + \lambda \mathbf{G}^T \mathsf{sgn}\left(\mathbf{G}\mathbf{z}\right) \tag{15}$$

In case when $\mathbf{l}$ is coordinate-wise convex, the objective for (12) is a bi-convex problem, which enjoys certain conergence properties. Optimizing multi-convex function has been explored in time-of-flight reconstruction [9].

4 Details

4.1 Structured Light Stereo

Some relevant reiews are [10],[11] and slides. A *structured light stereometric system* is similar to a passive stereo system where one of the camera is replaced by a projector. A light source projects light a vertical plane of light that creates a narrow stripe on the scene. The intersection of an illumination plane of known spatial position (corresponds to a projector column) and a line of sight (corresponds to a camera pixel) determines a point. For dense reconstruction of the scene, many images must be taken. To speed up the scanning process, spatially modulated light projector has been suggested, in which multiple illumination planes or rays can be projected simultaneously as part of a single illumination pattern. Spatial-temporal modulation of illumination, i.e. sequentially projecting several patterns, can be used for reliable identification of light planes. To enable acquisition of dynamic scenes, the number of projected patterns used should be as small as possible. Intuitively, the projected pattern impose illusion of texture on the object, increasing the number of correspondences, which enables reconstruction. Structured light stereo is equivalent to (1) solving the correspondence problem and then (2) computing stereo using triangulation.

4.1.1 The Correspondence Problem

The correspondence problem can be stated simply

For each point in the left image, find the corresponding point in the right image

We first note that search space for the corresponding point can be restricted to pixels lying on the epipolar line. Epipolar plane is plane formed from points p, o_1, o_2 and epipolar line is intersection of epipolar plane with the image plane. For structured light stereo systems, the art of designing robust, fast, reliable coding schemes serves to solve the correspondence problem.

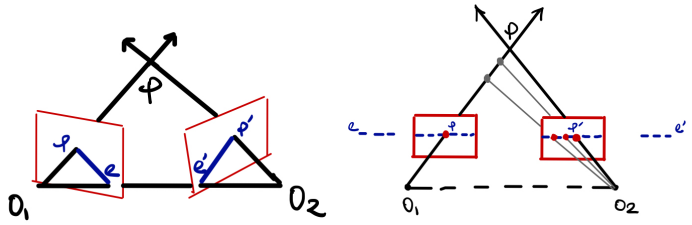

Figure 4: non parallel (left) and parallel (right) camera setup. pe and $p'e'$ are the epipolar lines

4.1.2 Triangulation

Given correspondence between projector and camera images, triangulation refers to the process of computing the distance of object relative to camera. Since a 3D point can be obtained by intersecting a ray (pixel of camera image) with a projector plane (a single code), it is necessary to encode a single axis in projected image to ensure unique reconstruction. *disparity* is the displacement between points of a conjugate pair p_l, p_r (points in different images of projection of same point in the scene) when the two images are superimposed. In the context of parallel calibrated cameras, disparity is inversely proportional to depth, when baseline b and focal length f are known

$$\frac{b}{z} = \frac{b + x_l - x_r}{z - f} \quad \text{implies} \quad z = \frac{bf}{x_l - x_r}$$

4.2 Structured Light Coding

The goal of structured light coding is to find coding schemes (maps pattern intensity to indices of the projector light planes) that enables robust, reliable algorithms for finding correspondence.

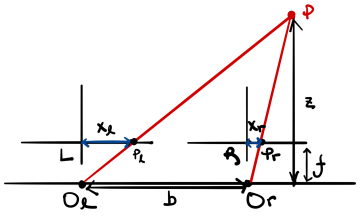

Figure 5: parallel-calibrated binocular stereo

4.2.1 Horn & Kiryati

This paper generalizes Gray code to n-ary code in order to reduce the number of patterns that needs to be projected, (L^K instead of 2^K code words) [12]. The authors draw inspirations from communication theory, where the projector projects unique temporal codes, received at each image plane through a noisy channel and subsequently decoded. Let there be K patterns and L code words (distinct planes of light), we want to encode the indices of vertical light planes $x \in [L]$ using some encoding scheme $f : [L] \to \mathbb{R}^K$ such that the nearest neighbor decoding $\hat{x}(y) = \arg\min_{x\in[L]}(f(x) - y)^2$ of a normalized noisy observation $y \in \mathbb{R}^K$ minimizes the probability of depth estimation error. Given the forwarding model,

$$\mathsf{y} = f(\mathsf{x}) + \mathsf{n} \qquad \text{where} \qquad \mathsf{x} \sim \text{Cat}(1/L) \quad \mathsf{n} \sim p_{\mathsf{n}}$$

implying $\mathsf{y}|\mathsf{x} = x \sim p_{\mathsf{n}}(y - f(x))$. We want to minimize the probability of depth estimation error, which roughly proportional to difference between true index x of the plane of light and the estimated index $\hat{x}$,

$$\text{minimize}_f \;\; \mathbb{E}_{\mathsf{x},\mathsf{y}}\left[(x - \hat{x}(y))^2\right] = \sum_{x=1}^{L} p_{\mathsf{x}}(x) \int p_{\mathsf{y}|\mathsf{x}}(y|x)(x - \hat{x}(y))^2 dy \tag{16}$$

$$\propto \sum_{x=1}^{L} \int (x - \hat{x}(y))^2 p_{\mathsf{n}}(y - f(x)) dy \tag{17}$$

This optimization problem is hard. The paper suggest the use of space filling curves as the encoding function and established that Gray code is a special limiting case of the space filling curve. Note here we assume there is no *mutual illumination*, i.e. there is no interval reflection and so the projected codes $f(x)$ is proportional to observation y.

4.2.2 Phase Shifting

Phase shifting is a particular coding method whereby the projector column coordinates are encoded as the (absolute) phase of a spatial sinusoidal pattern. Note we want to encode column coordinates because we only need to search along the horizontal epipolar lines. Let N be number of columns to be encoded. When the scene is projected with a cosine pattern of period T (measured in number of pixels) and therefore frequency $f = \frac{1}{T}$, an idealized image formation model for any pixel is

$$I = I_0 + A\cos(\Phi) = I_0 + A\cos(\phi) \tag{18}$$

where I_0 is the pixel intensity resulting from ambient illumination; I is the pixel intensity measured; A is amplitude (albedo/reflectance) of the signal corresponding to intensity of scene assuming unit intensity for projector patterns; $\Phi = 2\pi f x = 2\pi n + \phi \in [0, 2\pi f N]$ for some number of period $n \in \mathbb{N}$ is the absolute phase, $\phi \in [0, 2\pi]$ is the relative phase. When measured w.r.t. number of pixels, x is the corresponding absolute phase and $\tilde{x}$ the corresponding relative phase, satisfying

$$x = Tn + \tilde{x} \qquad \text{or} \qquad x \equiv \tilde{x} \mod T \tag{19}$$

where $\tilde{x} = \frac{T\phi}{2\pi}$. In (18), I_0, A, ϕ are unknown and so ϕ cannot be determined. To solve for ϕ, phase shifting method projects K sinusoidal patterns of same frequency, each shifted by $\varphi_k = \frac{2\pi(k-1)}{K}$ for $k = 1, \cdots, K$.

Thereby obtaining a system of K equations in 3 unknowns.

$$I_k = I_0 + A\cos\left(\phi + \varphi_k\right) \qquad \text{for} \qquad k = 1, \cdots, K \tag{20}$$

Although $K = 3$ suffices, larger values of K makes determination of relative phase ϕ more robust to noise. We determine ϕ using least squares

$$\text{minimize}_\phi \quad \left[\epsilon(\phi, I_0, A) := \sum_{k=1}^{K} \left[I_k - \left(I_0 + A\cos\left(\phi + \varphi_k\right)\right)\right]^2\right] \tag{21}$$

Similar to appendix in [13] and results shown in [14, 13], we can show the following

$$\begin{aligned} 0 = \frac{\partial \epsilon}{\partial \phi} = 2A\sum_{k=1}^{K} I_i \sin\left(\phi + \varphi_k\right) &\propto \cos(\phi)\sum_{k=1}^{K} I_k \sin(\varphi_k) + \sin(\phi)\sum_{k=1}^{K} I_k \cos(\varphi_k) \\ \phi = \tan^{-1}&\left[-\frac{\sum_{k=1}^{K} I_k \sin(\varphi_k)}{\sum_{k=1}^{K} I_k \cos(\varphi_k)}\right] \end{aligned} \tag{22}$$

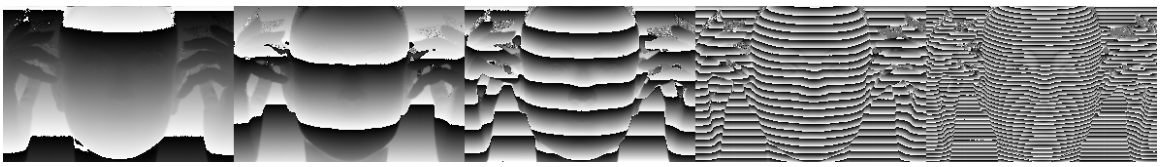

Figure 6: Relative phase ϕ for spatial sinusoids with period 1, 2, 5, 17, 31 solved via (22)

phase unwrapping aims to recover absolute phase x from relative phase $\tilde{x}$, which is non-trivial unless $T \geq N$. One particular choice of phase unwrapping method relies results in number theory [15]. The idea is to project patterns whose periods $T_1, \cdots, T_F$ are relative co-prime, each shifted by K times such that relative phase $\tilde{x}_1, \cdots, \tilde{x}_F$ can be solved using (21) or (24), from which we can use the Chinese Reminder Theorem to solve a system of congruences

$$\begin{aligned} x &= \tilde{x}_1 \mod T_1 \\ &\vdots \\ x &= \tilde{x}_F \mod T_F \end{aligned} \tag{23}$$

Intuitively, projecting F spatial sinusoids with period $T_1, \cdots, T_F$ with different frequency emulates the projection of a low frequency spatial sinusoid with period $T = T_1 \times \cdots \times T_F$, shown in Figure (7). As specified in [16] and experimented in Figure (8), phase unwrapping is unstable when the relatives phases $\tilde{x}_1, \cdots, \tilde{x}_F$ are noisy. Simple application of medium filtering of the relative phase does not seem to help with phase unwrapping. Wavelet denoising using `wdenoise` seems to help to a limited extend, as shown in Figure (9)

4.2.3 Micro Phase Shifting

An alternative formulation as mentioned in [16], we can write (20) as a linear system of equations

$$I_k = I_0 + A\cos(\phi)\cos(\varphi_k) - A\sin(\phi)\sin(\varphi_k) = \begin{bmatrix} 1 & \cos(\varphi_k) & -\sin(\varphi_k) \end{bmatrix} \begin{bmatrix} I_0 \\ A\cos(\phi) \\ A\sin(\phi) \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} I_1 \\ \vdots \\ I_K \end{bmatrix}}_{\mathbf{I}} = \underbrace{\begin{bmatrix} 1 & \cos(\varphi_1) & -\sin(\varphi_1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(\varphi_K) & -\sin(\varphi_K) \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} I_0 \\ A\cos(\phi) \\ A\sin(\phi) \end{bmatrix}}_{\mathbf{u}}$$

Figure 7: Two sinusoids with $T_1 = 17, T_2 = 31$ emulates a low frequency sinusoid with $T = 527$

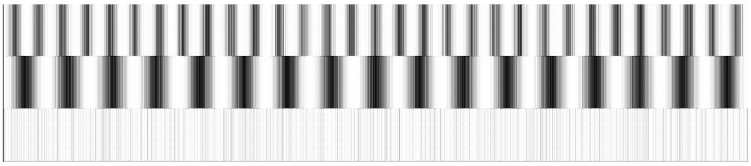

Figure 8: When relative phase $\tilde{x}_1, \tilde{x}_2$ are corrupted with i.i.d. additive Gaussian noise $\mathcal{N}(\mathbf{0}, 0.1 \cdot \mathsf{range}(\tilde{x}) \cdot \mathbf{I})$, phase unwrapping is unable to recover the absolute phase x robustly.

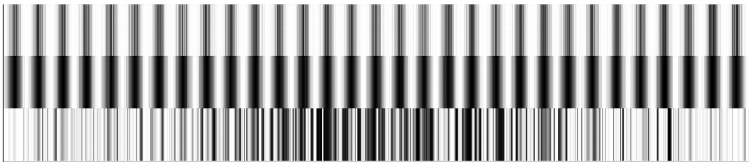

Figure 9: $\tilde{x}_1$ and wavelet denoised $\tilde{x}_1$ ($x_2, \tilde{x}_2$ not shown) helps with phase unwrapping to some extend

and solve for $\mathbf{u}$ using least squares,

$$\text{minimize}_{\mathbf{u}} \quad \|\mathbf{I} - \mathbf{M}\mathbf{u}\|_2^2 \tag{24}$$

The relative phase is then given by

$$\phi = \cos^{-1}\left(\frac{\mathbf{u}_2}{A}\right) \qquad \text{where} \qquad A = \sqrt{\mathbf{u}_2^2 + \mathbf{u}_3^2} \quad \text{and} \quad \mathbf{u} = M^{\dagger}\mathbf{I}$$

Note the optimization problem formulated in (21,24) are not equivalent.

4.2.4 ZNCC Decoding

Let L be number of vertical light planes to be encoded, $f : [L] \to \mathbb{R}^K$ is some encoding scheme that encodes index x to a code $f(x)$, y is per pixel measurement. Similar to nearest neighbor decoder used in [12]

$$\hat{x}(y) = \arg\min_{x \in [L]} \left(f(x) - y\right)^2 \tag{25}$$

[2] defines a decoding function that is optimal in a sense in small noise regimen,

$$\hat{x}(y) = \arg\max_{x \in [L]} \mathsf{ZNCC}\left(y, f(x)\right) \tag{26}$$

Note the quality of decoding depends heavily on the noise level. Additionally, number of shifts is crucial for good reconstruction of absolute phase, see Figure 10

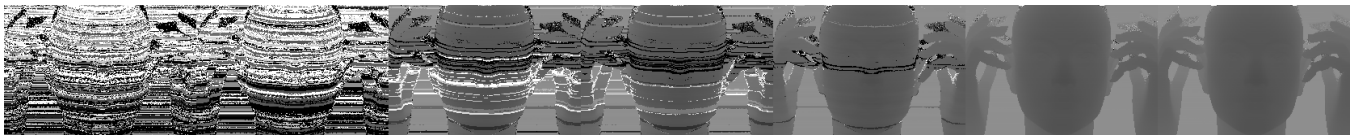

Figure 10: Here we have spatial sinusoids with period 1,2,5,17,31, each shifted 30 times projected onto a static scene. Resulting in $5 \cdot 30 = 150$ measurement, each stacked from 250 noisy images. We see ZNCC decoding taking 5,7,9,11,13,50,149 randomly chosen measurements respectively yield progressively better disparity reconstruction

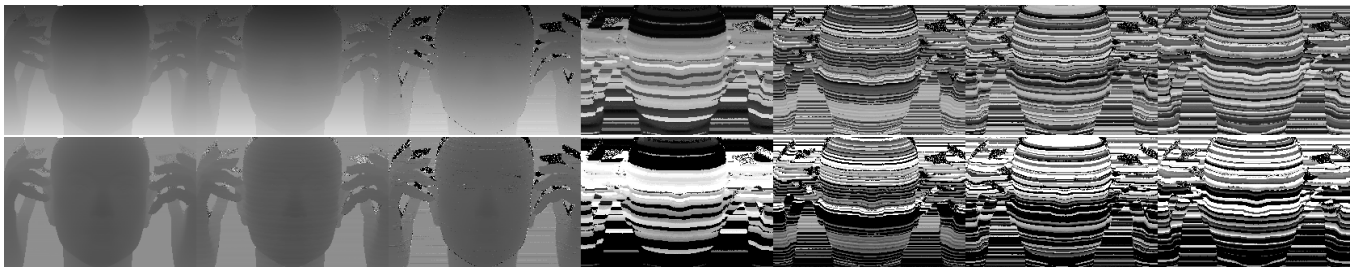

Figure 11: ZNCC decoded phase and disparity for (from left to right) ground truth shown in Figure 10, Hamiltonian, MPS, Optimized-MDE, Optimized-Top0, Optimized-Top1, Optimized-Top2. Apart from ground truth, reconstruction takes intensity of scene under 7 projector patterns, each stacked from 250 noisy measurements

4.2.5 Geometric Perspective & Hamiltonian Code

When trying to find optimal encoding scheme that minimizes bayesian decoding/correspondence error, a surrogate objective is the length of the coding curve. [17] argues that an optimal coding scheme should have a coding curve that is long, non-intersecting, and distance preserving (points far along the curve should also be far in euclidean distance). Sinusoidal codes has intersection with itself, i.e. multiple column indices are coded to the same location in measurement space implying there is 2π phase ambiguity when trying to compute its inverse decoding function. The distance preserving property is especially important in noisy regime, where non-distance preserving coding scheme would potentially yield large error as the decoding function maps (w.r.t. Euclidean distance) noisy measurement to indices far away from true value. For example, A La Carte code is not distance preserving as shown in Figure 12.

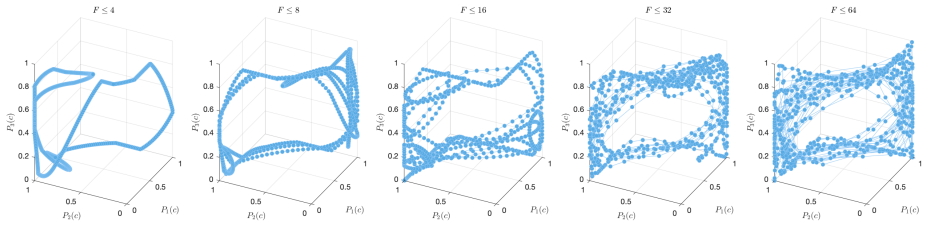

Figure 12: A La Carte codes for 3 patterns with varying max spatial frequency [2] is not distance preserving, i.e. points far along the curve are close in measurement space in the Euclidean sense. So we'd expect it to have sub-optimal performance when measurements are noisy

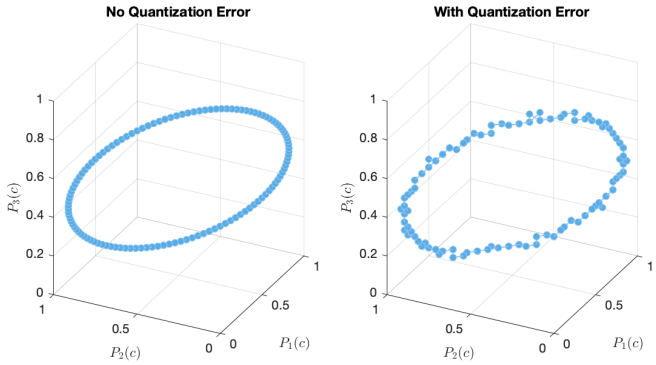

Figure 13: Projector quantization error translates to jitters in the coding curve for sinusoidal patterns

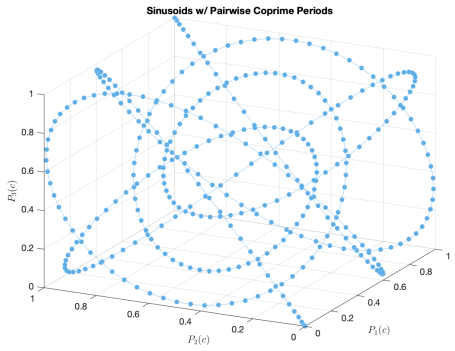

Figure 14: Sinusoids with pairwise co-prime periods (5, 11, 13) is non-intersecting and therefore emits unique decoder using the Chinese Reminder Theorem. However, certain points along the curve does not satisfy the distance preserving property, implying the code is not optimal when measurements are noisy. For sinusoidal codes, non-intersection and distance preserving is a conflicting objective so it seems.

4.3 Alternating Direction Method of Multipliers

4.3.1 Formulation

Consider the following equality constrained convex optimization problem

$$\text{minimize } f(x) + g(z) \tag{27}$$
$$\text{subject to } Ax + Bz = c \tag{28}$$

assuming f, g are both convex. The augmented lagrangian L_ρ makes the objective better behaved,

$$L_\rho(x, z, y) = f(x) + g(z) + y^T (Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2 \tag{29}$$

where y is the dual variable, and $\rho > 0$ is the penalty parameter. The method of multipliers is simply doing gradient ascent (using ρ as stepsize) on the dual objective computed from the augmented lagrangian

$$(x^{k+1}, z^{k+1}) = \arg\min_{x,z} L_\rho(x, z, y^k) \tag{30}$$
$$y^{k+1} = y^k + \rho \nabla L_\rho(x^{k+1}, z^{k+1}, y) = y^k + \rho \left(Ax^{k+1} + Bz^{k+1} - c \right) \tag{31}$$

When regular lagrangian is used and when objective is separable, the minimization step (30) is also separable with respect to the primal variables x, z. The addition of Euclidean norm on the residual of

primal equlity constraints makes the optimization coupled. ADMM is method of multiplier where a single Gauss-Seidel pass over the primal variable is used instead of jointly optimizing for x, z.

$$\begin{aligned} x^{k+1} &= \arg\min_x L_\rho(x, z^k, y^k) \\ z^{k+1} &= \arg\min_z L_\rho(x^{k+1}, z, y^k) \\ y^{k+1} &= y^k + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right) \end{aligned} \tag{32}$$

We can gather the residual term $r = Ax + Bz - c$ and completing the squares and get

$$L_\rho(x, z, y) = f(x) + g(z) + \frac{\rho}{2}\|Ax + Bz - c + u\|_2^2 \tag{33}$$

where $u = \frac{1}{\rho}y$ is scaled dual variable. The scaled form ADMM is then

$$\begin{aligned} x^{k+1} &= \arg\min_x \left(f(x) + \frac{\rho}{2}\left\|Ax + Bz^k - c + u^k\right\|_2^2\right) \\ z^{k+1} &= \arg\min_z \left(g(z) + \frac{\rho}{2}\left\|Ax^{k+1} + Bz - c + u^k\right\|_2^2\right) \\ u^{k+1} &= u^k + Ax^{k+1} + Bz^{k+1} - c \end{aligned} \tag{34}$$

4.3.2 LASSO Problem

The lasso problem is

$$\text{minimize} \quad \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1 \tag{35}$$

We can apply variable splitting and obtain an equivalent constrained optimization problem

$$\text{minimize} \quad \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|z\|_1 \tag{36}$$

$$\text{subject to} \quad x - z = 0 \tag{37}$$

Specialize (34) to ridge regression to obtain iterates of the form

$$x^{k+1} = \arg\min_x \left(\frac{1}{2}\|Ax - y\|_2^2 + \frac{\rho}{2}\left\|x - z^k + u^k\right\|_2^2\right) = \left(A^TA + \rho I\right)^{-1}\left(A^Ty + \rho(z^k - u^k)\right) \tag{38}$$

$$z^{k+1} = \arg\min_z \left(\lambda\|z\|_1 + \frac{\rho}{2}\left\|x^{k+1} - z + u^k\right\|_2^2\right) = \mathcal{S}_{\lambda/\rho}\left(x^{k+1} + u^k\right) \tag{39}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{40}$$

which essentially performs lasso (x update) repeatedly.

4.3.3 Affine Constrained Convex Optimization Problem

$$\begin{aligned} \underset{\mathbf{x}_1,\cdots,\mathbf{x}_N}{\text{minimize}} \quad & \left[f(\mathbf{x}) = \sum_{i=1}^N f_i(\mathbf{x}_i)\right] \\ \text{subject to} \quad & (\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \mathcal{C} \end{aligned} \tag{41}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$, $f_i : \mathbb{R}^n \to (-\infty, \infty)$ are closed proper convex functions for $i = 1, \cdots, N$. Here we assume $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{n_1 \times \cdots \times n_N} \mid A\mathbf{x} = \mathbf{b}\}$ is an affine set. (41) has an equivalent form,

$$\begin{aligned} \underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & f(\mathbf{x}) + \delta_{\mathcal{C}}(\mathbf{z}) \\ \text{subject to} \quad & \mathbf{x} - \mathbf{z} = 0 \end{aligned} \tag{42}$$

which can be solved using ADMM iterates of the form,

$$\begin{aligned}
\mathbf{x}_i^{k+1} &= \arg\min_{\mathbf{x}_i} \left(f_i(\mathbf{x}_i) + \frac{\rho}{2} \left\| \mathbf{x}_i - (\mathbf{z}_i^k - \mathbf{u}_i^k) \right\|_2^2 \right) = \mathsf{prox}_{(1/\rho) f_i}(\mathbf{z}_i^k - \mathbf{u}_i^k) \\
\mathbf{z}^{k+1} &= \arg\min_{\mathbf{z}} \left(\delta_{\mathcal{C}}(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{z} - (\mathbf{x}^{k+1} + \mathbf{u}^k) \right\|_2^2 \right) = \mathsf{prox}_{(1/\rho)\delta_{\mathcal{C}}}(\mathbf{x}^{k+1} + \mathbf{u}^k) \\
\mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}
\end{aligned}$$

where $\mathsf{prox}_{\lambda f} : \mathbb{R}^n \to \mathbb{R}^n$ is the proximal operator of λf, $\lambda > 0$,

$$\mathsf{prox}_{\lambda f}(\mathbf{v}) = \arg\min_{\mathbf{x}} \left(f(\mathbf{x}) + \frac{1}{2\lambda} \|\mathbf{x} - \mathbf{v}\|_2^2 \right)$$

4.3.4 Some Common Proximal Operator

Evaluating the proximal operator involves solving a convex optimization problem. We will show how we can compute the proximal operators relevant to our methods. (See section 6 of [18])

Projection Onto Convex Set The proximal operator of an indicator function onto a convex set $\mathcal{C}$ is simply the projection onto $\mathcal{C}$.

$$\mathsf{prox}_{(1/\rho)\delta_{\mathcal{C}}}(\mathbf{x}^{k+1} + \mathbf{u}^k) = \textstyle\prod_{\mathcal{C}}(\mathbf{x}^{k+1} + \mathbf{u}^k)$$

When $\mathcal{C}$ is affine, there is an analytic expression for the projection

$$\begin{aligned}
\textstyle\prod_{\mathcal{C}}(\mathbf{v}) &= \mathbf{v} - A^{\dagger}(A\mathbf{v} - \mathbf{b}) \\
&= \mathbf{v} - A^T(A^TA)^{-1}(A\mathbf{v} - \mathbf{b}) \qquad \text{(if } A \in \mathbb{R}^{m\times n} \text{ has } m < n \text{ and full rank)}
\end{aligned}$$

Quadratic Function Let f be ℓ-2 norm of an affine function, assuming $A \in \mathbb{R}^{p\times n}$

$$f(x) = \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2 = \frac{1}{2}\mathbf{x}^TA^TA\mathbf{x} - \mathbf{y}^TA\mathbf{x} + \frac{1}{2}\mathbf{y}^T\mathbf{y}$$

Then proximal operator of $(1/\rho)f$ has a closed form expression

$$\mathsf{prox}_{(1/\rho)f}(\mathbf{v}) = (\rho I + A^TA)^{-1}(\rho\mathbf{v} + A^T\mathbf{y})$$

which can be solved efficiently with conjugate gradient, as $(\rho I + A^TA) \succ 0$. If ρ is fixed throughout, we can use Cholesky factorization to factor $(\rho I + A^TA)$ in $\mathcal{O}(n^3)$. Any subsequent computation of the inverse with back-solve would only cost $\mathcal{O}(n^2)$. When $p << n$, we can exploit this by using the matrix inversion lemma,

$$(\rho I + A^TA)^{-1} = \frac{1}{\rho}I - \frac{1}{\rho}A^T\left(\rho I + AA^T\right)^{-1}A$$

The dominant cost is computing $(AA^T)^{-1} \in \mathbb{R}^{p\times p}$. Each x update now costs $\mathcal{O}(np^2)$. If use Cholesky factorization to factor $(\rho I + AA^T)$ once, subsequent iteration can be carried out in $\mathcal{O}(np)$ flops, which is essentially costs for matrix-vector multiply.

ℓ-1 Norm Let $f(\mathbf{x}) = \|\mathbf{x}\|_1$, then the proximal operator of λf is

$$\mathsf{prox}_{\lambda f}(\mathbf{v}) = \mathcal{S}_{\lambda}(\mathbf{v})$$

where $\mathcal{S}$ is element-wise soft shrinkage operator

$$(\mathcal{S}_{\lambda}(\mathbf{v}))_i = (1 - \lambda/|\mathbf{v}_i|)_+\mathbf{v}_i \qquad \mathbf{x}_+ = \max(\mathbf{x}, 0)$$

RED Explicit Regularizer Let $f(\mathbf{x}) = (1/2)\mathbf{x}^T(\mathbf{x} - \mathcal{D}(\mathbf{x}))$ for some denoiser $\mathcal{D}$, the explicit regularizer in RED.[5]. We use one fixed point iteration evaluate the approximate proximal operator for λf. Specifically, we want to evaluate

$$\mathsf{prox}_{(\lambda/\rho)f}(\mathbf{v}) = \arg\min_{\mathbf{x}} \frac{\lambda}{2}\mathbf{x}^T(\mathbf{x} - \mathcal{D}(\mathbf{x})) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{v}\|_2^2 \tag{43}$$

Setting the gradient to zero, we arrive at the fixed point iteration

$$\mathbf{x}^{(k)} \leftarrow \frac{1}{\rho + \lambda}(\lambda\mathcal{D}(\mathbf{x}^{(k-1)}) + \rho\mathbf{v}) \tag{44}$$

If we only iterate once, then

$$\mathsf{prox}_{(\lambda/\rho)f}(\mathbf{v}) = \frac{1}{\rho + \lambda}(\lambda\mathcal{D}(\mathbf{x}^{(0)}) + \rho\mathbf{v}) \tag{45}$$

for some initialization value $\mathbf{x}^{(0)}$

4.4 Linear Inverse Problem

Numerous inverse problem in image processing are modeled using the following relationship

$$y = Ax + \epsilon \tag{46}$$

wherer $y \in \mathbb{R}^m$ is noisy measurement, $x \in \mathbb{R}^n$ are unknown image, $A \in \mathbb{R}^{m\times n}$ captures the forward relationship, $\epsilon \in \mathbb{R}^m$ is noise. The goal is to estimate x from noisy y by making assumptions about the noise distributioon ϵ and exploiting structure in the latent variable x. For example, denoising, super-resolution, demosaicking, deblurring, deconvolution can all be formulated as solving the problem of the form (46)

4.4.1 MAP Inference Under Gaussian Noise

A typical approach to solve the linear inverse problem relies on computing max a posterior (MAP) estimator while assuming somer prior distirbution p_{x}. In particular given observation y, we want to estimate $\hat{x}$ where

$$\hat{x} = \arg\max_x p_{\mathsf{x}|\mathsf{y}}(x|y) \tag{47}$$

Assuming isotropic Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, the conditional likelihood depends on the noise distribution, $\mathsf{y} \mid \mathsf{x} = x \sim \mathcal{N}(Ax, \sigma^2 I)$. The log likelihood is then

$$\log p_{\mathsf{y}|\mathsf{x}}(y|x) = -\frac{m}{2}\log(2\pi) - \frac{1}{2\sigma^2}\|Ax - y\|_2^2 \tag{48}$$

Then the problem (47) reduces to

$$\hat{x}(y) = \arg\max_x \left(\log p_{\mathsf{y}|\mathsf{x}}(y|x) + \log p_{\mathsf{x}}(x)\right) = \arg\min_x \left(\frac{1}{2\sigma^2}\|Ax - y\|_2^2 - \log p_{\mathsf{x}}(x)\right) \tag{49}$$

The choice of regularization has been an important research topic in image processing. Numerous task-specific prior have been successful in a number of image recovery problems: (1) the sparsity of $\mathbf{x}$ with ℓ-1 norm in image deblurring [19] (2) total variation in image denoising [20] (3) cross-channel correlation in color image demosaicing [21]. To solve (49) over a class of tractable priors, algorithms that exploits variable splitting and imploys proximal algorithms have shown to be very efficient [22, 23, 3, 24]. There are works to learn proximal operator from data [25, 26], uses a CNN denoiser [27].

4.4.2 Plug-and-Play Prior

PnP prior is a method that exploits denoisers as implicit regularizers for the inverse problem that has shown good performance on a range of inverse problems in imaging [28],[29]. Fixed point convergence can be proved for a class of bounded denoisers [28]. Most recently, [30] established convergence given that the denoising network satisfy certain Lipschitz condition.

Given problem

$$\text{minimize}_x \ \frac{1}{2}\|Ax-y\|_2^2 + \lambda R(x) \tag{50}$$

where $R(\cdot)$ is some arbitrary prior for the model, which can be functionally be interpreted as the regularizer for the denoising subproblem (54). Similar to the lasso probelm (35), PnP methods uses variable splitting but instead considers z-update as a denoising subproblem where the prior $R(\cdot)$ is implicitly imposed by the denoiser,

$$x^{k+1} = \arg\min_x \left(\frac{1}{2}\|Ax-y\|_2^2 + \frac{\rho}{2}\left\|x - z^k + u^k\right\|_2^2\right) = \left(A^TA+\rho I\right)^{-1}\left(A^Ty + \rho(z^k-u^k)\right) \tag{51}$$

$$z^{k+1} = \arg\min_z \left(\lambda R(z) + \frac{\rho}{2}\left\|x^{k+1} - z + u^k\right\|_2^2\right) = \mathcal{D}_{\lambda/\rho}(x^{k+1}+u^k) \tag{52}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{53}$$

where denoiser $\mathcal{D}_{\lambda/\rho}(v)$ denoises $v = x+u$ by making the assumption that v is a noisy measurement of z, i.e. $v = z+e$ where $e \sim \mathcal{N}(0,\sigma^2)$ and $\sigma^2 = \frac{\lambda}{\rho}$, i.e. $\mathcal{D}_{\sigma^2}(v)$ solves the following denoising subproblem

$$\mathcal{D}_{\sigma^2}(v) = \arg\min_z \frac{1}{2\sigma^2}\|v-z\|_2^2 + R(z) \tag{54}$$

For any non-trivial A, the noise model for primal variable z is not necessarily Gaussian.

4.4.3 Regularization By Denoising

Regularization by denoising (RED) came up with an explicit form for the regularizer $R(x) = \frac{1}{2}x^T(x - \mathcal{D}(x))$ [5, 31]. With some assumptions on the denoiser, the objective function emits computable gradients ($\nabla R(x) = x - \mathcal{D}(x)$) and is therefore amenable to arbitrary first order optimization methods. A similar ADMM updates can be obtained

$$x^{k+1} = \left(A^TA+\rho I\right)^{-1}\left(A^Ty + \rho(z^k-u^k)\right) \tag{55}$$

$$z^{k+1} = \frac{1}{\rho+\lambda}\left(\lambda\mathcal{D}(z^k) + \rho(x^{k+1}+u^k)\right) \tag{56}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{57}$$

where z-update is simply a single fixed point iteration given by (45). In practice, PnP and RED yield similar performance but RED is more robust to the choice of hyper-parameters. Possible denoisers include block-matching and 3D filtering (BM3D) [32] and trainable nonlienar reaction diffusion (TNRD) [33].

References

[1] Mian Wei et al. "Coded Two-Bucket Cameras for Computer Vision". 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. URL: http://link.springer.com/10.1007/978-3-030-01219-9_4 (visited on 07/19/2019).

[2] P. Mirdehghan, W. Chen, and K. N. Kutulakos. "Optimal Structured Light a la Carte". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 2018, pp. 6248–6257. DOI: 10.1109/CVPR.2018.00654.

[3] Felix Heide et al. "FlexISP: A Flexible Camera Image Processing Framework". In: *ACM Trans. Graph.* 33.6 (Nov. 2014), 231:1–231:13. ISSN: 0730-0301. DOI: 10.1145/2661229.2661260. URL: http://doi.acm.org/10.1145/2661229.2661260 (visited on 07/19/2019).

[4] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. "Plug-and-Play priors for model based reconstruction". In: *2013 IEEE Global Conference on Signal and Information Processing*. 2013 IEEE Global Conference on Signal and Information Processing. Dec. 2013, pp. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048.

[5] Yaniv Romano, Michael Elad, and Peyman Milanfar. "The Little Engine that Could: Regularization by Denoising (RED)". In: *arXiv:1611.02862 [cs]* (Nov. 9, 2016). arXiv: 1611.02862. URL: http://arxiv.org/abs/1611.02862 (visited on 07/19/2019).

[6] Yang Liu et al. "Rank Minimization for Snapshot Compressive Imaging". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.12 (Dec. 1, 2019), pp. 2990–3006. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2018.2873587. arXiv: 1807.07837. URL: http://arxiv.org/abs/1807.07837 (visited on 11/08/2019).

[7] Amit Adam et al. "Bayesian Time-of-Flight for Realtime Shape, Illumination and Albedo". In: *arXiv:1507.06173 [cs]* (July 22, 2015). arXiv: 1507.06173. URL: http://arxiv.org/abs/1507.06173 (visited on 06/10/2020).

[8] Guy Rosman, Anastasia Dubrovina, and Ron Kimmel. "Sparse Modeling of Shape from Structured Light". In: *Visualization Transmission 2012 Second International Conference on 3D Imaging, Modeling, Processing*. Visualization Transmission 2012 Second International Conference on 3D Imaging, Modeling, Processing. ISSN: 1550-6185. Oct. 2012, pp. 456–463. DOI: 10.1109/3DIMPVT.2012.20.

[9] Felix Heide et al. "Non-line-of-sight Imaging with Partial Occluders and Surface Normals". In: *arXiv:1711.07134 [cs]* (Nov. 19, 2017). arXiv: 1711.07134. URL: http://arxiv.org/abs/1711.07134 (visited on 10/06/2019).

[10] Joaquim Salvi, Jordi Pagès, and Joan Batlle. "Pattern codification strategies in structured light systems". In: *Pattern Recognition*. Agent Based Computer Vision 37.4 (Apr. 1, 2004), pp. 827–849. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2003.10.002. URL: http://www.sciencedirect.com/science/article/pii/S0031320303003303 (visited on 07/19/2019).

[11] Joaquim Salvi et al. "A state of the art in structured light patterns for surface profilometry". In: *Pattern Recognition* 43.8 (Aug. 1, 2010), pp. 2666–2680. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2010.03.004. URL: http://www.sciencedirect.com/science/article/pii/S003132031000124X (visited on 05/24/2020).

[12] E. Horn and N. Kiryati. "Toward optimal structured light patterns". In: *Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134)*. Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134). ISSN: null. May 1997, pp. 28–35. DOI: 10.1109/IM.1997.603845.

[13] D. Moreno, K. Son, and G. Taubin. "Embedded phase shifting: Robust phase shifting with embedded signals". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 2015, pp. 2301–2309. DOI: 10.1109/CVPR.2015.7298843.

[14] Tomislav Pribanić, Saša Mrvoš, and Joaquim Salvi. "Efficient multiple phase shift patterns for dense 3D acquisition in structured light scanning". In: *Image and Vision Computing* 28.8 (Aug. 1, 2010), pp. 1255–1266. ISSN: 0262-8856. DOI: 10.1016/j.imavis.2010.01.003. URL: http://www.sciencedirect.com/science/article/pii/S0262885610000107 (visited on 05/26/2020).

[15] V.I. Gushov and Yu.N. Solodkin. "Automatic processing of fringe patterns in integer interferometers". In: *Optics and Lasers in Engineering* 14.4 (1991), pp. 311–324. ISSN: 01438166. DOI: 10.1016/0143-8166(91)90055-X. URL: https://linkinghub.elsevier.com/retrieve/pii/014381669190055X (visited on 02/09/2020).

[16] M. Gupta and S. K. Nayar. "Micro Phase Shifting". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI: IEEE, June 2012, pp. 813–820. ISBN: 978-1-4673-1228-8 978-1-4673-1226-4 978-1-4673-1227-1. DOI: 10.1109/CVPR.2012.6247753. URL: http://ieeexplore.ieee.org/document/6247753/ (visited on 07/19/2019).

[17] Mohit Gupta and Nikhil Nakhate. "A Geometric Perspective on Structured Light Coding". In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, pp. 87–102. URL: http://openaccess.thecvf.com/content_ECCV_2018/html/Mohit_Gupta_A_Geometric_Perspective_ECCV_2018_paper.html (visited on 11/26/2019).

[18] Neal Parikh and Stephen Boyd. "Proximal Algorithms". In: *Found. Trends Optim.* 1.3 (Jan. 2014), pp. 127–239. ISSN: 2167-3888. DOI: 10.1561/2400000003.

[19] 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*. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Apr. 2009, pp. 693–696. DOI: 10.1109/ICASSP.2009.4959678.

[20] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. "Nonlocal Image and Movie Denoising". In: *International Journal of Computer Vision* 76.2 (Feb. 1, 2008), pp. 123–139. ISSN: 1573-1405. DOI: 10.1007/s11263-007-0052-1. URL: https://doi.org/10.1007/s11263-007-0052-1 (visited on 07/19/2019).

[21] 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*. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 3. May 2004, pp. iii–485. DOI: 10.1109/ICASSP.2004.1326587.

[22] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. "Fast Image Recovery Using Variable Splitting and Constrained Optimization". In: *IEEE Transactions on Image Processing* 19.9 (Sept. 2010), pp. 2345–2356. ISSN: 1057-7149. DOI: 10.1109/TIP.2010.2047910.

[23] Mário A. T. Figueiredo and José M. Bioucas-Dias. "Restoration of Poissonian Images Using Alternating Direction Optimization". In: *IEEE Transactions on Image Processing* 19.12 (Dec. 2010), pp. 3133–3145. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2010.2053941. arXiv: 1001.2244. URL: http://arxiv.org/abs/1001.2244 (visited on 07/19/2019).

[24] Felix Heide et al. "ProxImaL: efficient image optimization using proximal algorithms". In: *ACM Transactions on Graphics* 35.4 (July 11, 2016), pp. 1–15. ISSN: 07300301. DOI: 10.1145/2897824.2925875. URL: http://dl.acm.org/citation.cfm?doid=2897824.2925875 (visited on 07/19/2019).

[25] Tim Meinhardt et al. "Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems". In: *arXiv:1704.03488 [cs]* (Apr. 11, 2017). arXiv: 1704.03488. URL: http://arxiv.org/abs/1704.03488 (visited on 07/19/2019).

[26] J. H. Rick Chang et al. "One Network to Solve Them All — Solving Linear Inverse Problems Using Deep Projection Models". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, Oct. 2017, pp. 5889–5898. ISBN: 978-1-5386-1032-9. DOI: 10.1109/ICCV.2017.627. URL: https://ieeexplore.ieee.org/document/8237889/ (visited on 07/19/2019).

[27] K. Zhang et al. “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising”. In: *IEEE Transactions on Image Processing* 26.7 (July 2017), pp. 3142–3155. ISSN: 1057-7149. DOI: 10.1109/TIP.2017.2662206.

[28] Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. “Plug-and-Play ADMM for Image Restoration: Fixed Point Convergence and Applications”. In: *arXiv:1605.01710 [cs]* (May 5, 2016). arXiv: 1605.01710. URL: http://arxiv.org/abs/1605.01710 (visited on 07/19/2019).

[29] Stanley H. Chan. “Algorithm-Induced Prior for Image Restoration”. In: *ArXiv* abs/1602.00715 (2016). arXiv: 1602.00715.

[30] Ernest Ryu et al. “Plug-and-Play Methods Provably Converge with Properly Trained Denoisers”. In: *International Conference on Machine Learning*. International Conference on Machine Learning. May 24, 2019, pp. 5546–5557. URL: http://proceedings.mlr.press/v97/ryu19a.html (visited on 07/19/2019).

[31] Edward T. Reehorst and Philip Schniter. “Regularization by Denoising: Clarifications and New Interpretations”. In: *arXiv:1806.02296 [cs]* (June 6, 2018). arXiv: 1806.02296. URL: http://arxiv.org/abs/1806.02296 (visited on 07/19/2019).

[32] K. Dabov et al. “Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering”. In: *IEEE Transactions on Image Processing* 16.8 (Aug. 2007), pp. 2080–2095. ISSN: 1057-7149. DOI: 10.1109/TIP.2007.901238.

[33] Yunjin Chen and Thomas Pock. “Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.6 (June 1, 2017), pp. 1256–1272. ISSN: 0162-8828, 2160-9292. DOI: 10.1109/TPAMI.2016.2596743. arXiv: 1508.02848. URL: http://arxiv.org/abs/1508.02848 (visited on 07/22/2019).