Project Report

Peiqi Wang

Monday 25th May, 2020

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 Methods **5**
3.1 A Linear Inverse Problem . . . 5
3.2 Optimization . . . 6
3.3 Constrained Convex Optimization . . . 7
3.4 Evaluating the Proximal Operator . . . 7

4 Clarifications **8**
4.1 Structured Light Stereo . . . 8
4.2 Structured Light Coding . . . 9
4.3 Image Priors . . . 11

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

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

2.2 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 $\mathbf{Y}$,

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

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 illuination mosaics that trade spatial resolution for temporal resolution ($1, 2, \cdots, F$ 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

$$\mathbf{S} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

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

2.3 Image Formation

Per-pixel image formation model is

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

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

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

2.4 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 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.

3 Methods

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 $\mathbf{y}$,

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

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 $\boldsymbol{I}$ 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 the following unconstrained optimization problem,

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

where $\rho : \mathbb{R}^{SP} \to \mathbb{R}$ are regularizers 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 probablistic 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]

We first note that the illumination ratios are albedo quasi-invariant, and therefore smooth within object boundaries. Therefore, total variation regularization on illuination ratio images could be particularly effective. To avoid extra notations, we use $\mathbf{x}, \mathbf{y}$ as the corresponding illumination ratios that we want to reconstruct. Additionally, we adapt algorithm in [4] for imposing algorithm induced priors with state-of-the-art denoisers. In summary, we want to optimize the following constrained problem with a set of affine constraints,

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{A}\mathbf{x}_1 - \mathbf{y}\|_2^2 + \frac{\lambda_2}{2}\mathbf{x}_2^T(\mathbf{x}_2 - \mathcal{D}(\mathbf{x}_2)) + \lambda_3 \|\mathbf{x}_3\|_1 \\ \text{subject to} \quad & \mathbf{x}_1 - \mathbf{x}_2 = 0 \\ & \mathbf{G}\mathbf{x}_1 - \mathbf{x}_3 = 0 \end{aligned}$$

where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^{SP}$, $\mathbf{x}_3 \in \mathbb{R}^{2SP}$. $\lambda_2, \lambda_3 > 0$ are weights to the regularizers. $\mathbf{G} \in \mathbb{R}^{2SP \times SP}$ is the discrete image gradient for S images

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_S \otimes \mathbf{G}_x \\ \mathbf{I}_S \otimes \mathbf{G}_y \end{bmatrix}$$

where $\mathbf{G}_x, \mathbf{G}_y \in \mathbb{R}^{P \times P}$ are the discrete image gradients for a single image computed using forward difference. We can gather constraints into a single linear system

$$\mathbf{H}\mathbf{x} = 0 \qquad \text{where} \qquad \mathbf{H} = \begin{bmatrix} \mathbf{I}_{SP} & -\mathbf{I}_{SP} & 0 \\ \mathbf{G} & 0 & -\mathbf{I}_{SP} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$$

and arrive at an equivalent optimization problem

$$\begin{aligned} \text{minimize} \quad & f_1(\mathbf{x}_1) + \lambda_2 f_2(\mathbf{x}_2) + \lambda_3 f_3(\mathbf{x}_3) \\ \text{subject to} \quad & (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathcal{C} \end{aligned} \tag{5}$$

where $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{4SP} \mid \mathbf{H}\mathbf{x} = 0\}$ and

$$\begin{aligned} f_1(\mathbf{x}_1) &= \|\mathbf{A}\mathbf{x}_1 - \mathbf{y}\|_2^2 \\ f_2(\mathbf{x}_2) &= \frac{1}{2}\mathbf{x}_2^T(\mathbf{x}_2 - \mathcal{D}(\mathbf{x}_2)) \\ f_3(\mathbf{x}_3) &= \|\mathbf{x}_3\|_1 \end{aligned}$$

3.2 Optimization

As shown below, the scaled form ADMM for solving (5) is given by

$$\begin{aligned} \mathbf{x}_1^{k+1} &= \mathsf{prox}_{(1/\rho)f_1}(\mathbf{z}_1^k - \mathbf{u}_1^k) = (I + \frac{2}{\rho}A^TA)^{-1}(\mathbf{z}_1^k - \mathbf{u}_1^k + \frac{2}{\rho}A^Ty) \\ \mathbf{x}_2^{k+1} &= \mathsf{prox}_{(\lambda_2/\rho)f_2}(\mathbf{z}_2^k - \mathbf{u}_2^k) = \frac{1}{\lambda_2 + \rho}(\lambda_2\mathcal{D}(\mathbf{x}_2^k) + \rho(\mathbf{z}_2^k - \mathbf{u}_2^k)) \\ \mathbf{x}_3^{k+1} &= \mathsf{prox}_{(\lambda_3/\rho)f_3}(\mathbf{z}_3^k - \mathbf{u}_3^k) = \mathcal{S}_{\lambda_3/\rho}(\mathbf{z}_3^k - \mathbf{u}_3^k) \\ \mathbf{z}^{k+1} &= \mathsf{prox}_{(1/\rho)\boldsymbol{I}_\mathcal{C}}(\mathbf{x}^{k+1} + \mathbf{u}^k) = (I - \mathbf{H}^\dagger\mathbf{H})(\mathbf{x}^{k+1} + \mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1} \end{aligned}$$

3.3 Constrained Convex Optimization

Alternating Method of Multipliers (ADMM) is an efficient convex optimization algorithm for minimizing the sum of nonsmooth convex separable functions subject to a set of linear equality constraints. It combines the advantage of dual decomposition, where we can solve subproblems at each iteration, and that of method of multipliers, where optimization over augmented lagrangian relaxes strict convexity condition for the objective.[5] Specifically, we are interested in solving the following constrained convex optimization problem

$$\begin{aligned} \underset{\mathbf{x}_1,\cdots,\mathbf{x}_N}{\text{minimize}} \quad & \sum_{i=1}^{N} f_i(\mathbf{x}_i) \\ \text{subject to} \quad & (\mathbf{x}_1,\cdots,\mathbf{x}_N) \in \mathcal{C} \end{aligned} \tag{6}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$, $f_i : \mathbb{R}^n \to (-\infty, \infty)$ are closed proper convex functions for $i = 1, \cdots, N$, and $\mathcal{C}$ is an affine set of the form

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^{\sum_i n_i} \mid A\mathbf{x} = \mathbf{b}\}$$

which has an equivalent ADMM form [5],

$$\begin{aligned} \underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & f(\mathbf{x}) + \boldsymbol{I}_{\mathcal{C}}(\mathbf{z}) \\ \text{subject to} \quad & \mathbf{x} - \mathbf{z} = 0 \end{aligned} \tag{7}$$

where $f(\mathbf{x}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i)$. The scaled form of ADMM for (7) is

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

$\mathbf{u}^k$ is the scaled dual variable. $\rho > 0$ is the augmented lagrangian parameter. $\mathsf{prox}_{\lambda f} : \mathbb{R}^n \to \mathbb{R}^n$ is the proximal operator of λf, $\lambda > 0$,

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

3.4 Evaluating the 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 [6])

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

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

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

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

2. Let f be ℓ-2 norm of an affine function,

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

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

$$\mathsf{prox}_{(1/\rho)f}(\mathbf{v}) = \mathsf{prox}_{(2/\rho)(1/2)f} = (I + \frac{2}{\rho}A^T A)^{-1}(\mathbf{v} + \frac{2}{\rho}A^T\mathbf{y})$$

which can be solved efficiently with conjugate gradient, as $(I + \frac{2}{\rho}A^T A) \succ 0$. If ρ is fixed throughout, we can use Cholesky factorization to factor $(I + (2/\rho)A^T A)$ in $\mathcal{O}(n^3)$. Any subsequent computation of the inverse would only cost $\mathcal{O}(mn)$.

3. Let $f(\mathbf{x}) = \|\mathbf{x}\|_1$, then the proximal operator of $(\lambda/\rho)f$ is

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

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

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

4. 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.[4]. We use one fixed point iteration evaluate the approximate proximal operator for λf. Specifically, we want to evaluate

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

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

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

If we only iterate once, then

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

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

4 Clarifications

4.1 Structured Light Stereo

Some relevant reiews are [7],[8] 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.

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

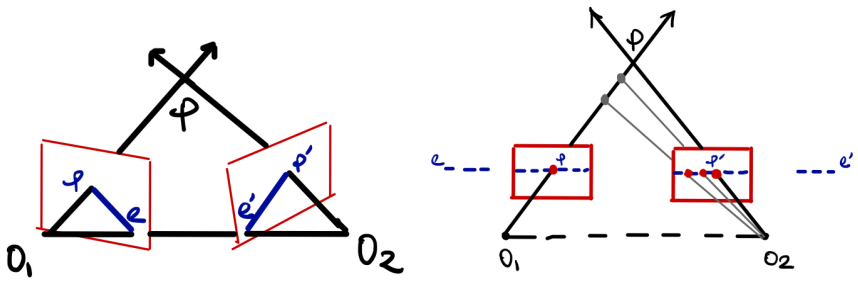

Figure 1: non parallel (left) and parallel (right) camera setup. pe and $p'e'$ are the epipolar lines

schemes serves to solve the correspondence problem.

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

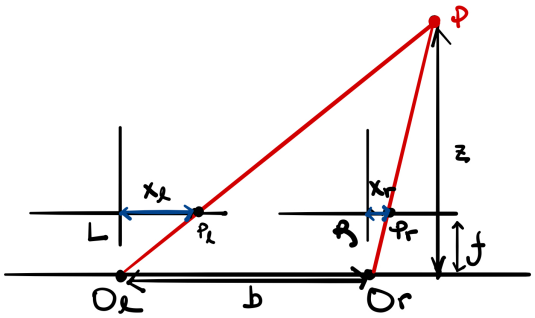

Figure 2: parallel-calibrated binocular stereo

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.

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) [9]. 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 $\hat{x}(y) = \arg\min_{x\in[L]}(f(x)-y)$ of a normalized noisy observation $y \in \mathbb{R}^K$ minimizes the probability of depth estimation error. Given the forwarding model,

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

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 $\hat{x}$,

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

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.

Phase Shifting Phase shifting is a particular coding method whereby the projector column coordinates are encoded as the phase of a series of N continuous sinusoidal pattern, each shifted by ϕ_i

One obtain a relative phase map (values in modulo 2π) from which we need to compute the absolute phase map by *phase unwrapping*.

4.3 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 task-specific priors: (1) the sparsity of $\mathbf{x}$ with ℓ-1 norm in image deblurring [2] (2) total variation in image denoising [10] (3) cross-channel correlation in color image demosaicing [11] (4) dark channel prior in image dehazing [12], etc. More exhotically, randomly initialized neural network can inject inductive bias to the optimization and act as image priors. [3]

In addition to hand-crafted priors, there has been interest in algorithm induced priors. Alternating direction method of multipliers (ADMM) is a common convex optimization method for inverse problem where the objective function is separable with respect to the *data term* and the *regularizer*. Each primal update involves an evaluation of a proximal operator, which can be interpreted as performing denoising on some iterate. [13, 14, 15] proposed plug-and-play priors where the choice of regularization is implicitly specified by the denoiser used. [4] proposed an explicit laplacian-based expression for the regularizer and generalizes the method to a number of different iterative optimization algorithms.

The convergence of plug-and-play ADMM is studied by a number of papers. [16] showed fixed point convergence of plug-and-play ADMM. [4] showed convergence of the algorithm under some mild conditions of the denoiser, which are satisfied by some state of the art denoisers like *Block-matching and 3D filtering (BM3D)* [17] and *Trainable Nonlienar Reaction Diffusion (TNRD)* [18]. Most recently, [19] established convergence given that the denoising network satisfy certain Lipschitz condition.

Some proposed to learn proximal operator from data. [20] used a CNN denoiser [21]. Instead of substituting the proximal operator with a denoiser, [22] learns a projection mapping to the space of natural images by training a single neural network and showed impressive results on a number of different linear inverse problems.

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] Yaniv Romano, Michael Elad, and Peyman Milanfar. "The Little Engine That Could: Regularization by Denoising (RED)". In: *arXiv:1611.02862 [cs]* (Nov. 2016). arXiv: 1611.02862 [cs].

[5] Stephen Boyd et al. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Found. Trends Mach. Learn.* 3.1 (Jan. 2011), pp. 1–122. ISSN: 1935-8237. DOI: 10.1561/2200000016.

[6] 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.

[7] Joaquim Salvi, Jordi Pagès, and Joan Batlle. "Pattern Codification Strategies in Structured Light Systems". In: *Pattern Recognition*. Agent Based Computer Vision 37.4 (Apr. 2004), pp. 827–849. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2003.10.002.

[8] 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).

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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*. Dec. 2013, pp. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048.

[14] 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.

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

[16] 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 2016). arXiv: 1605.01710 [cs].

[17] 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.

[18] 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 2017), pp. 1256–1272. ISSN: 0162-8828, 2160-9292. DOI: 10.1109/TPAMI.2016.2596743. arXiv: 1508.02848.

[19] Ernest Ryu et al. "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers". en. In: *International Conference on Machine Learning*. May 2019, pp. 5546–5557.

[20] Tim Meinhardt et al. "Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems". In: *arXiv:1704.03488 [cs]* (Apr. 2017). arXiv: 1704.03488 [cs].

[21] 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.

[22] J. H. Rick Chang et al. "One Network to Solve Them All — Solving Linear Inverse Problems Using Deep Projection Models". en. In: *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.