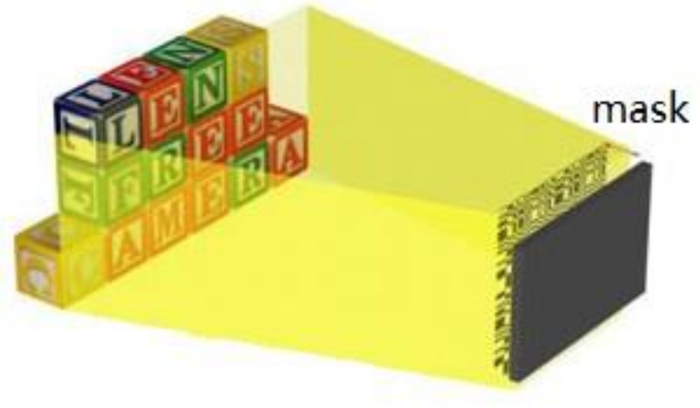
DiffuserCam: Lensless Single-exposure 3D Imaging

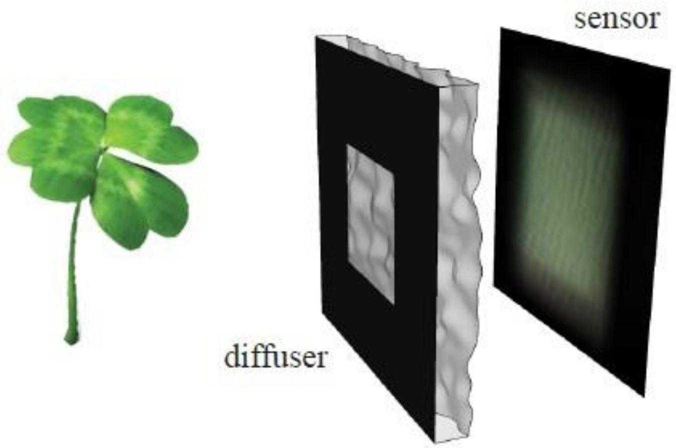
Author: Nick Antipa,
Grace Kuo,
Reinhard Heckell,
Ben Mildenhall,
Emrah Bostan,
Ren Ng,
Laura Waller

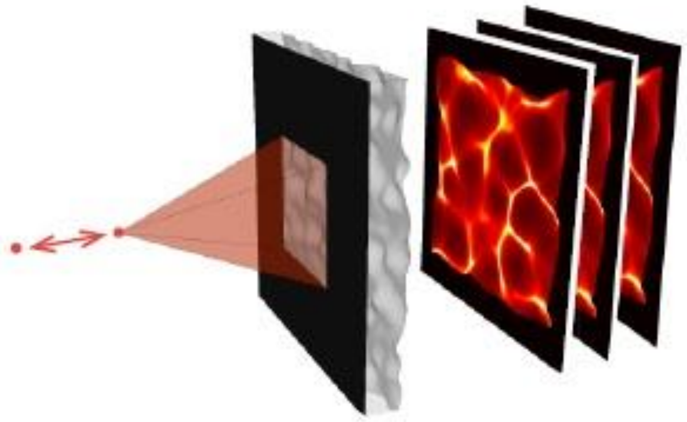
Presenter: Wang Zi

- Introduction
- System overview
- Convolutional Forward Model
- Inverse Algorithm
- System Analysis
- Experimental Results

Introduction







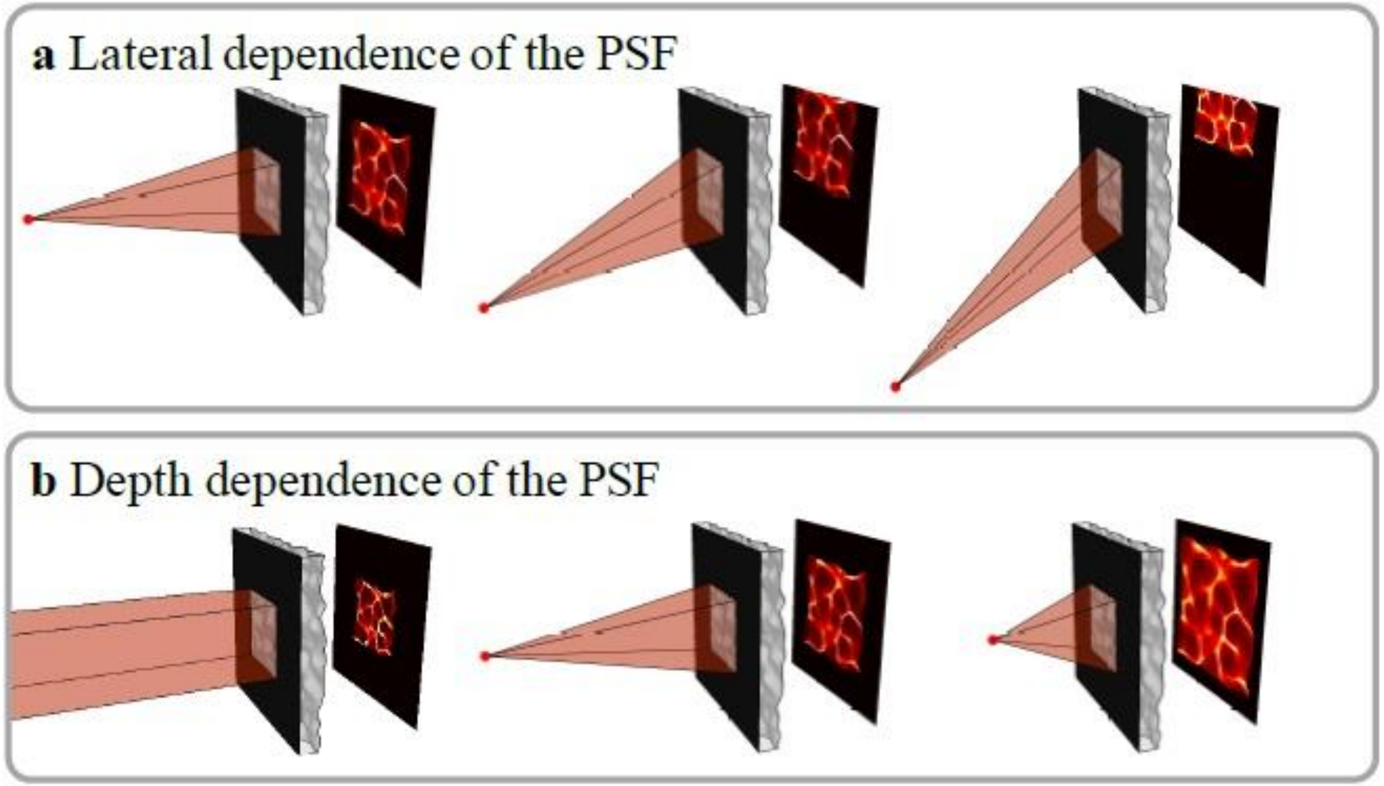
Mask $\Longrightarrow$ Diffuser

A point source:(x_0, y_0, z_0)

$\downarrow$

Image: $h_{x_0,y_0,z_0}(x, y)$

System overview



PSF of each point source is unique

This ensure 3-D reconstruction!

System overview

Object: $v(x, y, z) \leftrightarrow$ reshaped to $\boldsymbol{v}$

Image: $b(x', y') \leftrightarrow$ reshaped to $\boldsymbol{b}$

$$b(x', y') = \sum_{(x,y,z)} v(x, y, z) h(x', y'; x, y, z)$$

$$\boldsymbol{b} = \boldsymbol{H}\boldsymbol{v}$$

System overview

Calibration:

Estimate all $h(x', y'; x, y, z)$

Reconstruction:

$$\hat{v} = \underset{v \geq 0}{argmin} \frac{1}{2} \|b - Hv\|_2^2 + \lambda \|\Psi v\|_1$$

$\boldsymbol{\Psi}$ is a transform matrix to make $\boldsymbol{\Psi v}$ sparse

If $\boldsymbol{v}$ is sparse in 3-D space, we choose $\boldsymbol{\Psi}$ as identity matrix

If the gradient of $\boldsymbol{v}$ is sparse, we choose $\boldsymbol{\Psi}$ as finite difference operator and then $\|\boldsymbol{\Psi v}\|_{\mathbf{1}}$ becomes total variation(TV) norm

Q

Convolutional Forward Model

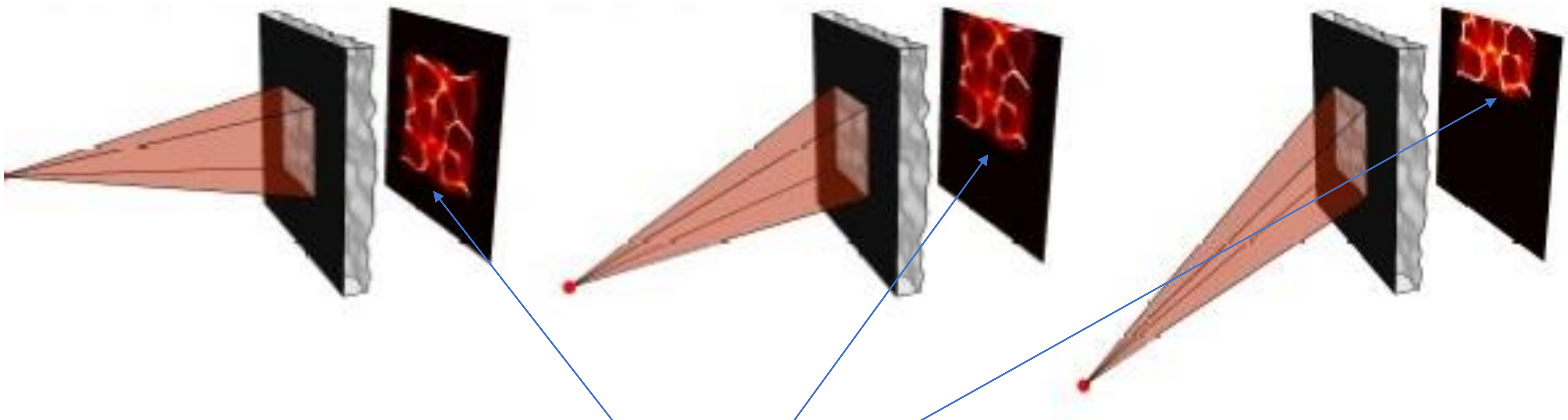
$$b(x', y') = \sum_{(x,y,z)} v(x, y, z)h(x', y'; x, y, z)$$

Suppose the size of object is $N_x \times N_y \times N_z$

We have to do $N_x N_y N_z$ times imaging to get all h

How to simplify it?

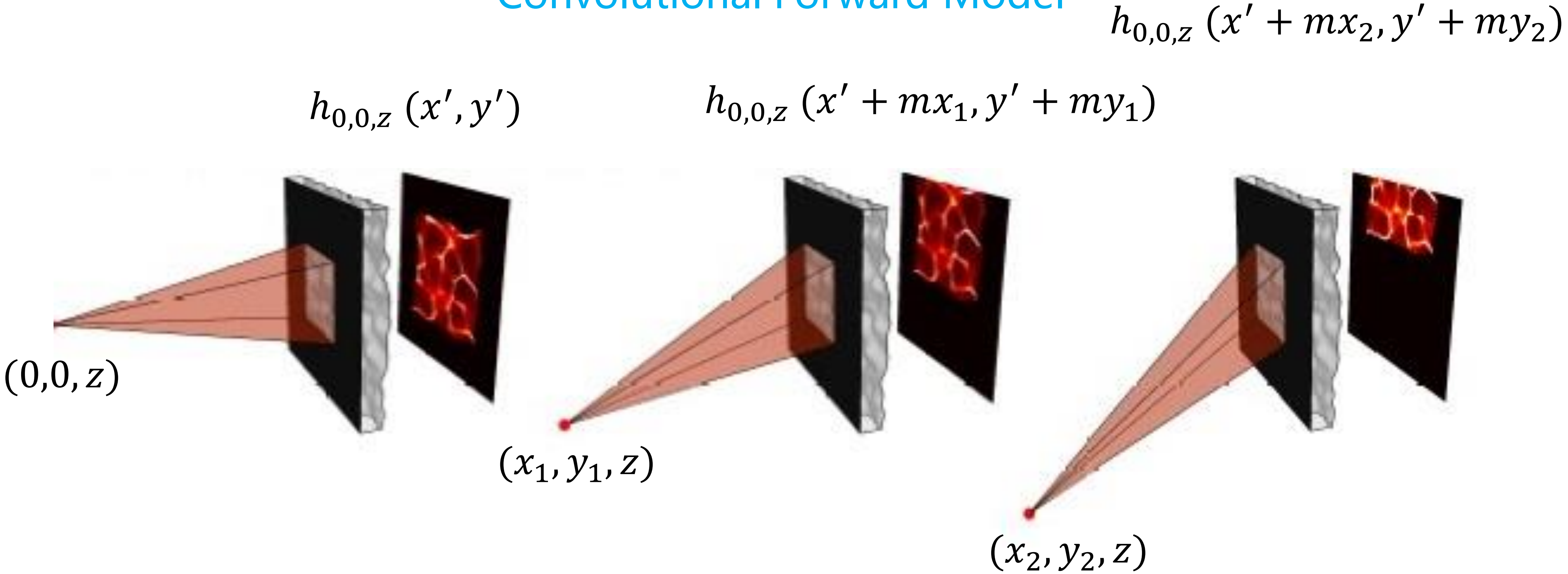
Convolutional Forward Model



They are almost translated results!

Can we describe them with only one PSF?

Convolutional Forward Model



With a shift invariance assumption, we have

$$h(x', y'; x, y, z) = h(x' + mx, y' + my; 0,0, z)$$

Then we only need to do N_z times imaging to calibrate

Convolutional Forward Model

Model:

$$b = Hv$$

We can do a composition:

$$b = DMv$$

Where D is a diagonal matrix, M is a convolution matrix

Inverse Algorithm

Reconstruction:

$$\hat{v} = \underset{v\geq 0}{argmin}\frac{1}{2}||b - Hv||_2^2 + \lambda||\Psi v||_1$$

With variable-splitting method, it can be written as

$$\hat{v} = \underset{w\geq 0,u,v}{argmin}\frac{1}{2}||b - D\mu||_2^2 + \lambda||u||_1\,, s.t\ \mu = Mv,\ u = \Psi v, w = v$$

ADMM algorithm is used to solved the problem

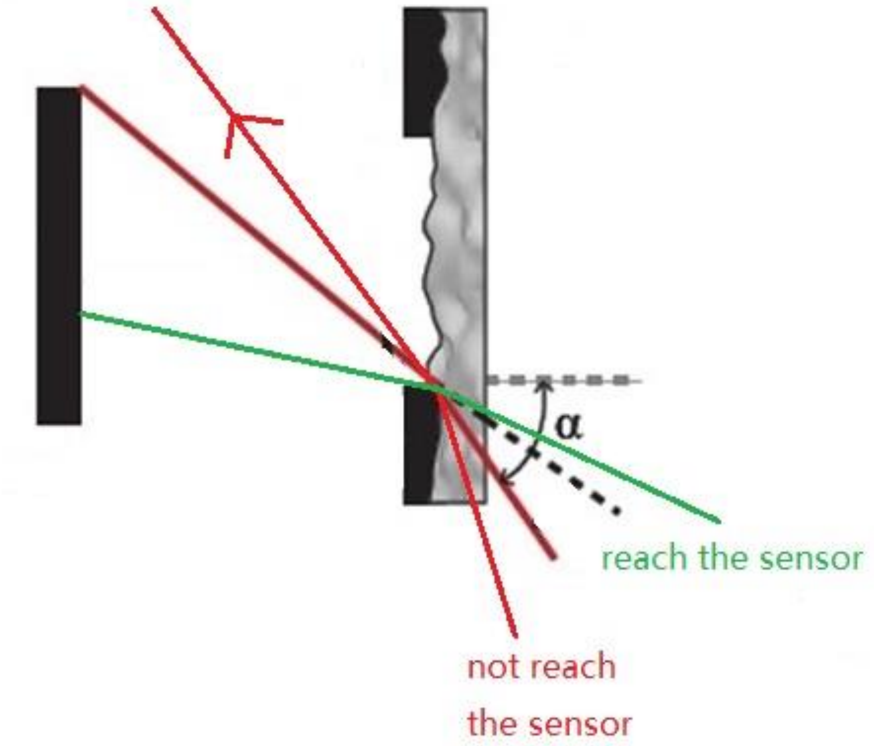
System Analysis

Field of View(FoV)

1. geometric angular cutoff

   Not all rays can reach the sensor!

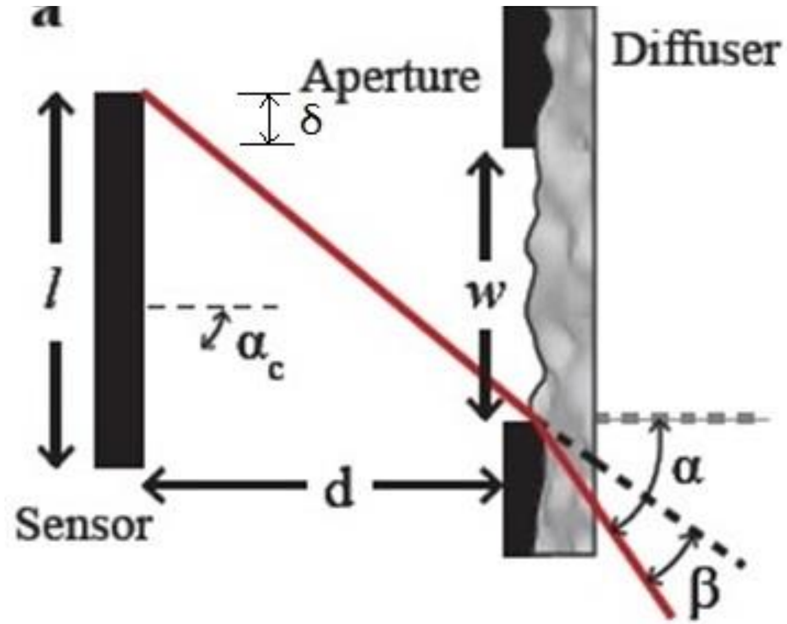
   $\boldsymbol{\alpha}$ is the angular



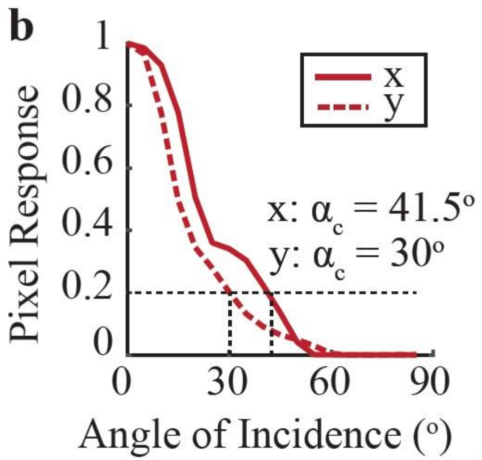
$$l - w = 2\delta$$

$$\tan(\alpha - \beta) = \frac{l - \delta}{d}$$

$$\alpha = \beta + arctan\,(\frac{l + w}{2d})$$



System Analysis



Field of View(FoV)

2. Angular response limitation:

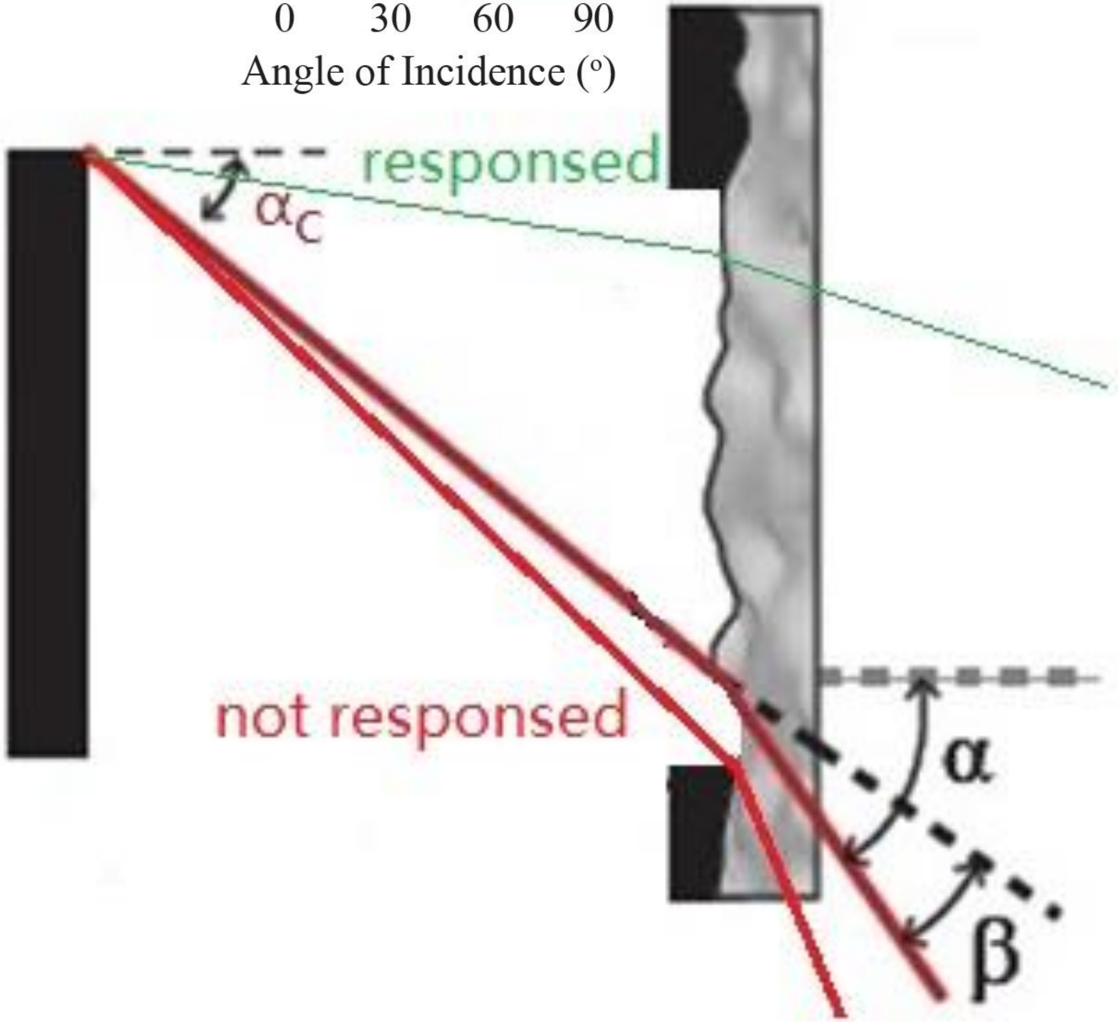
Not all rays will have response!

The limit is α_c, determined by the sensor

$$\alpha - \beta = \alpha_c \Rightarrow \alpha = \beta + \alpha_c$$

Above all, the final FoV is

$$FoV = \beta + \min\{\alpha_c, arctan(\frac{l+w}{2d})\}$$



Q

System Analysis

resolution:

We need to design a proper object grid,

Which means: divide object as voxels

- Two-point resolution
  - Lateral resolution
    - A single depth plane
    - Different depth
  - Axis resolution
- multipoint resolution

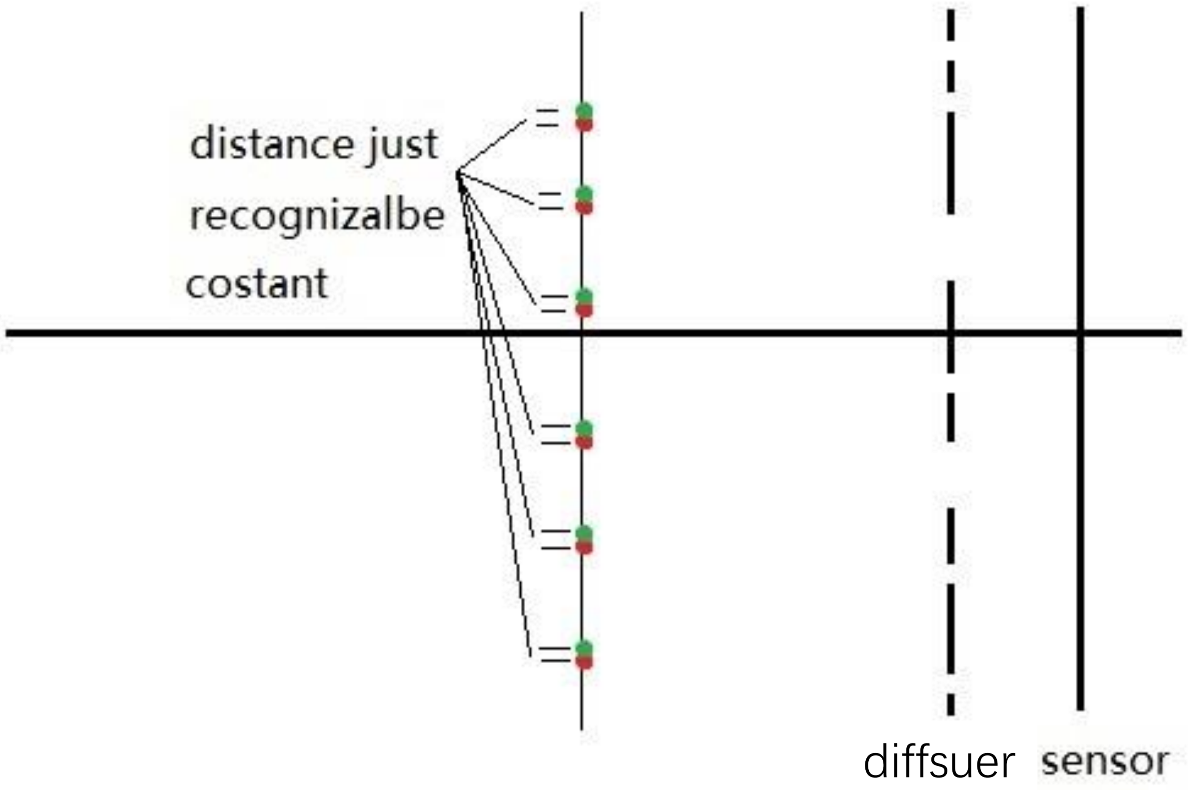
System Analysis

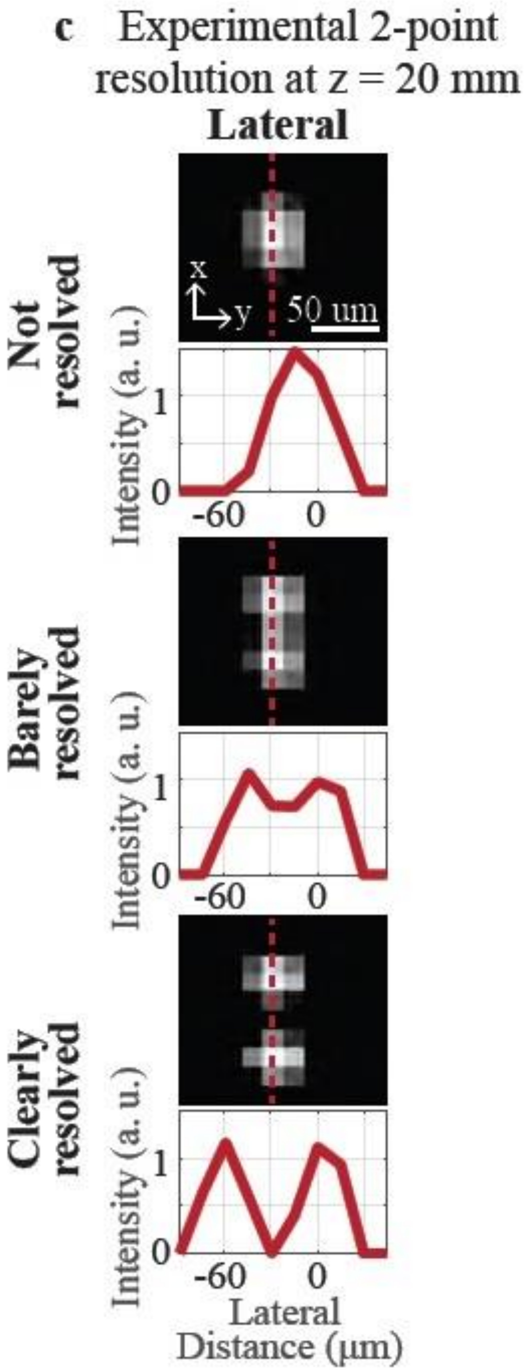
Two-point resolution:

1. Lateral resolution

   Single depth plane:

   constant due to shift invariance assumption





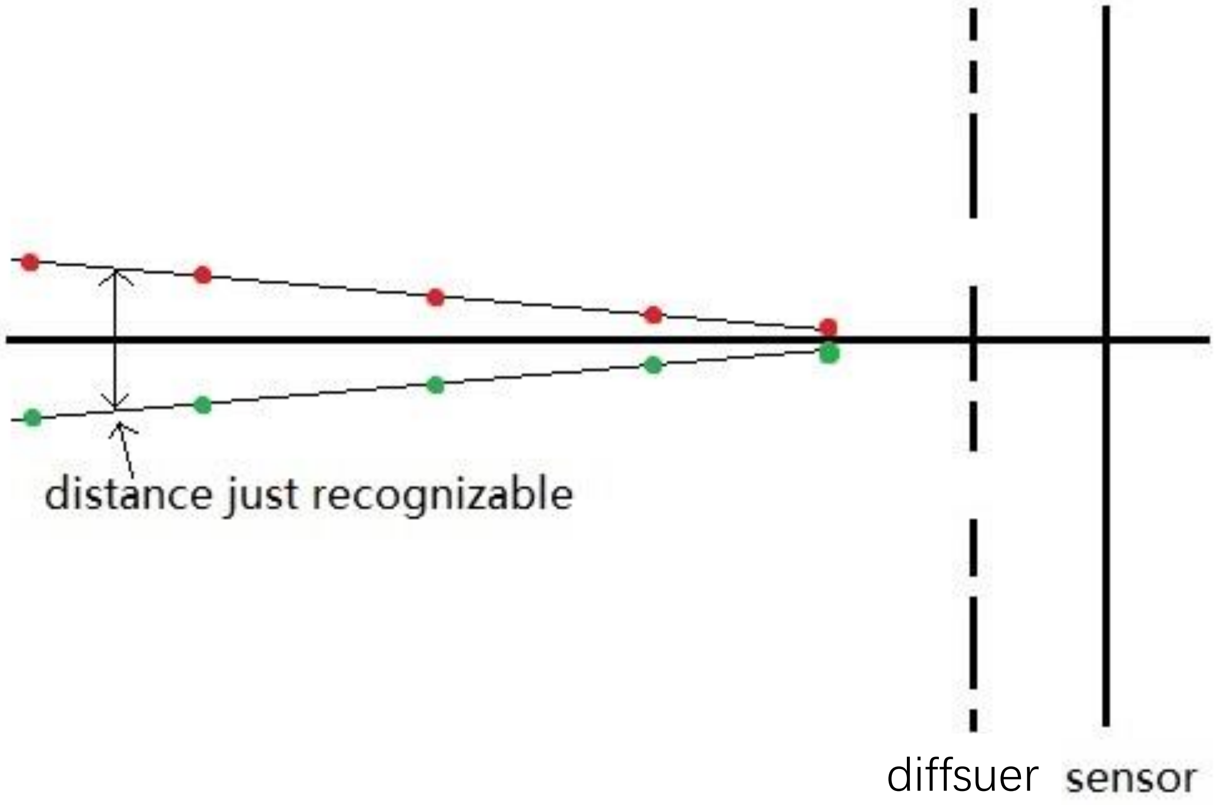
System Analysis

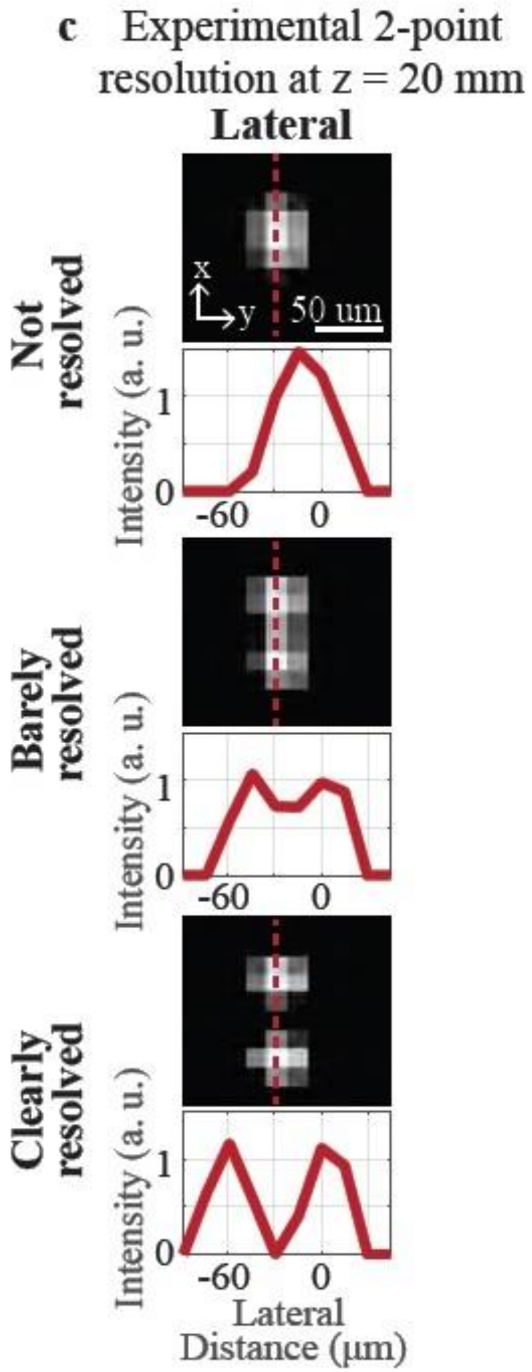
Two-point resolution:

1. Lateral resolution

   Multi-depth plane:

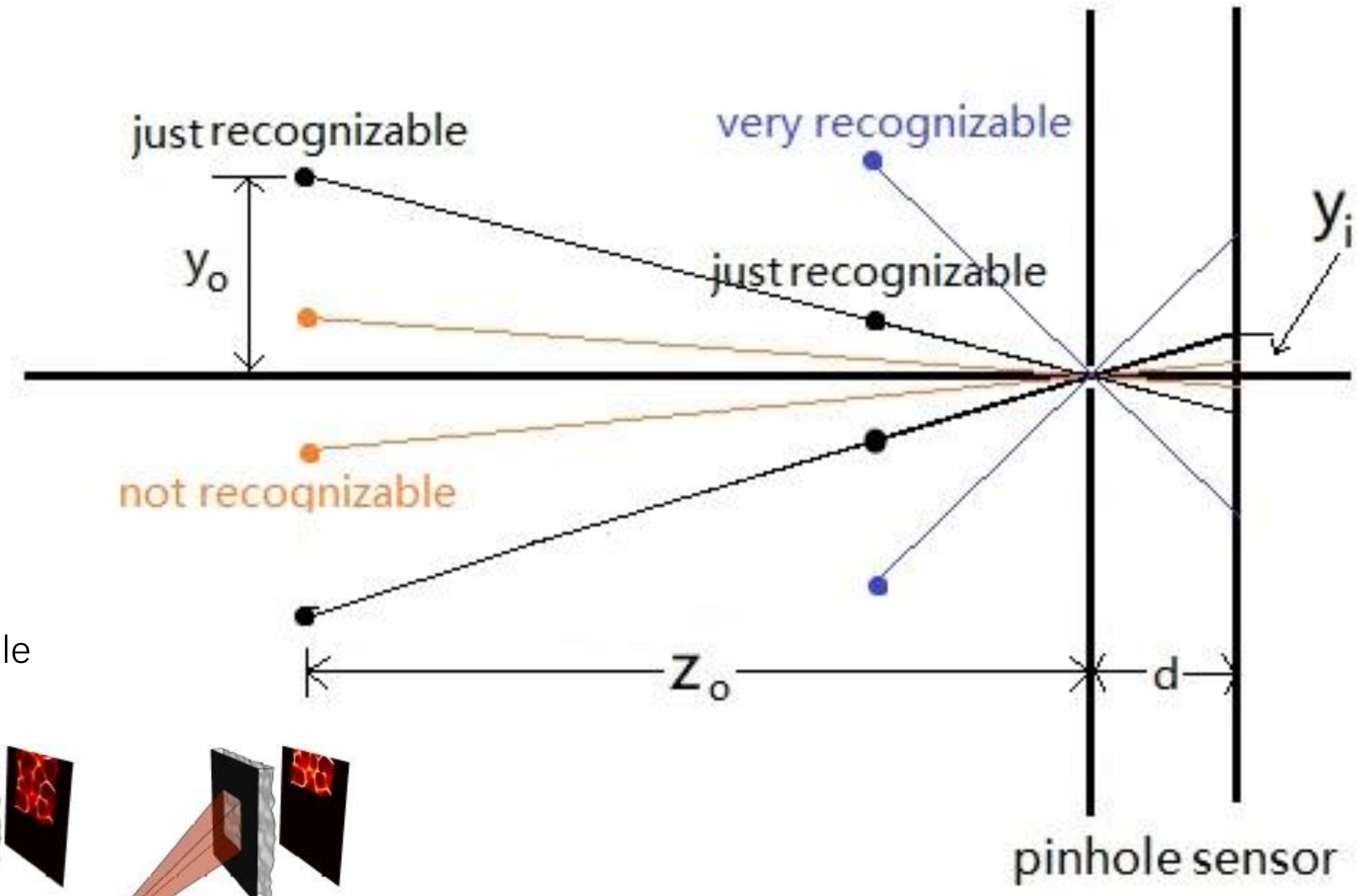
   At different depth: linearly with depth



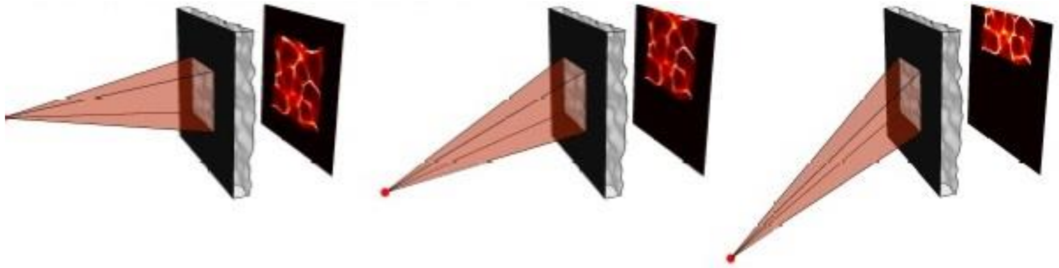


System Analysis

$$\frac{y_o}{z_0} = \frac{y_i}{d} = costant$$

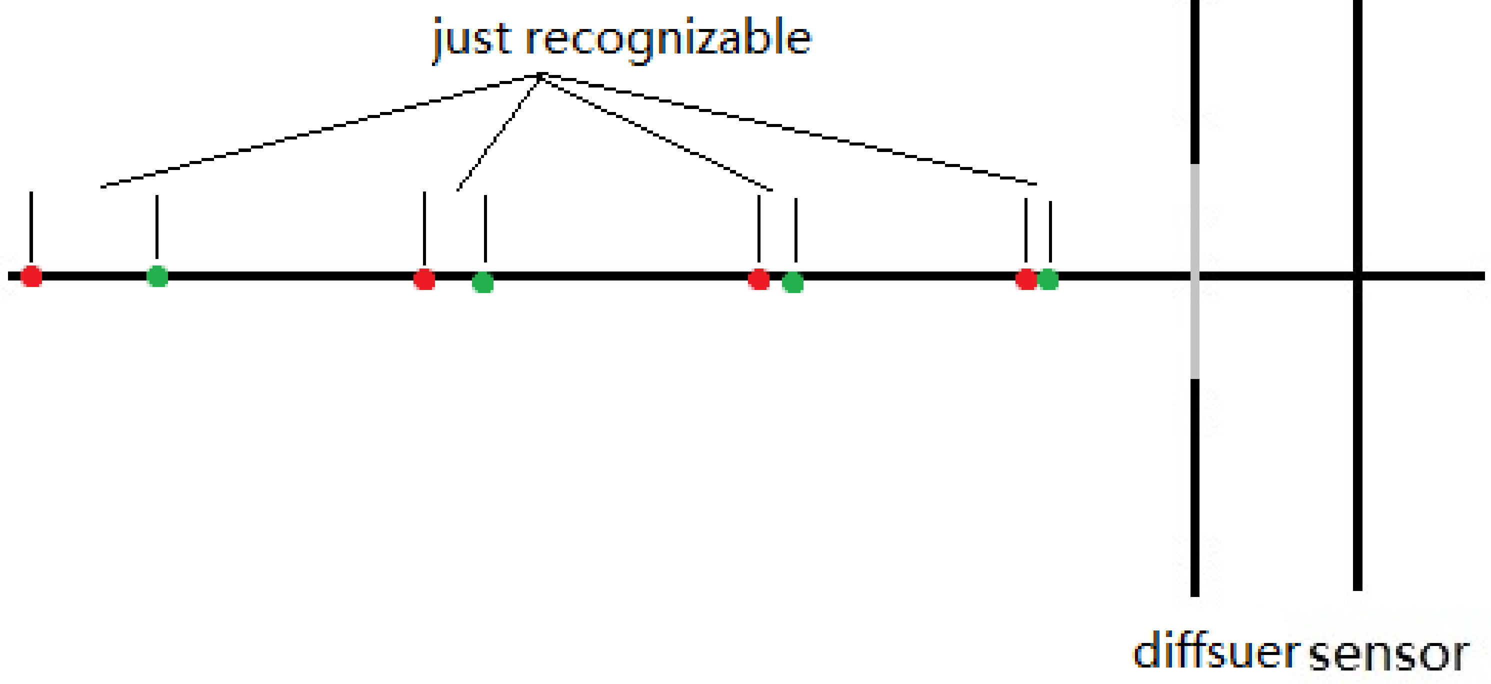


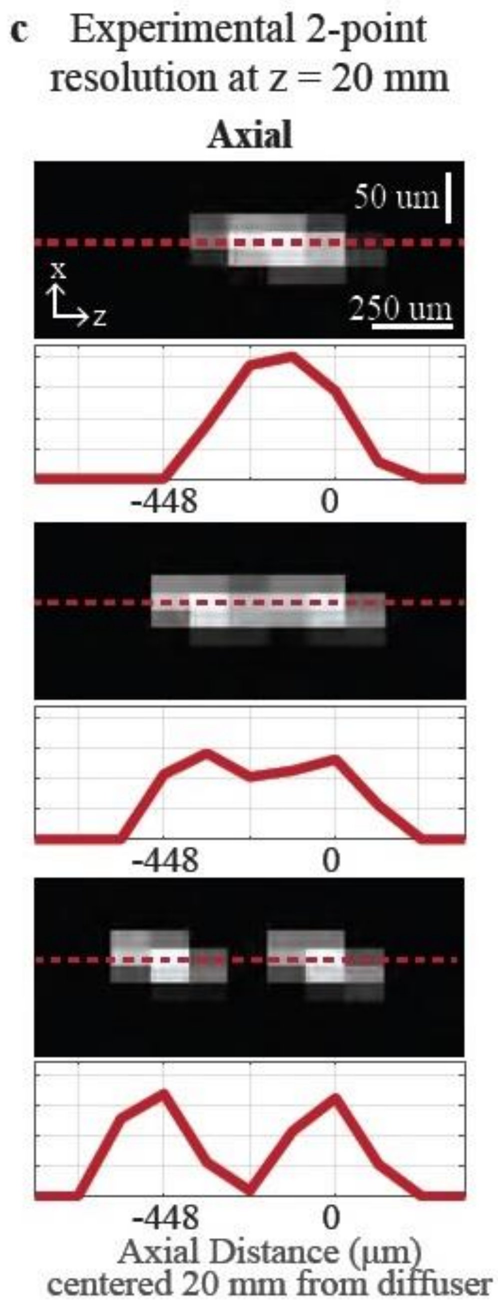
Pinhole: images recognizable
Diffsuer: PSFs recognizable



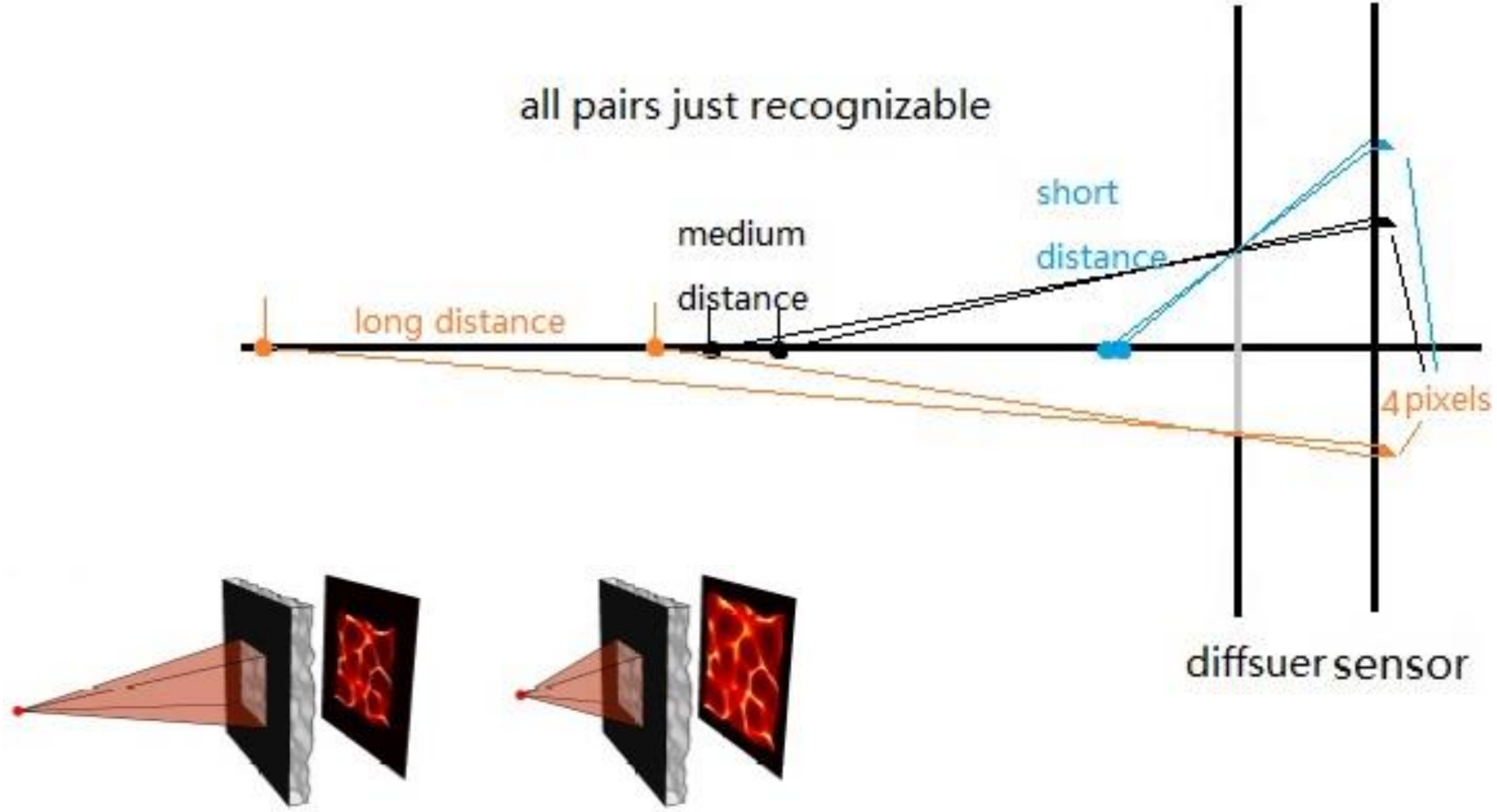
System Analysis

2. Axis resolution



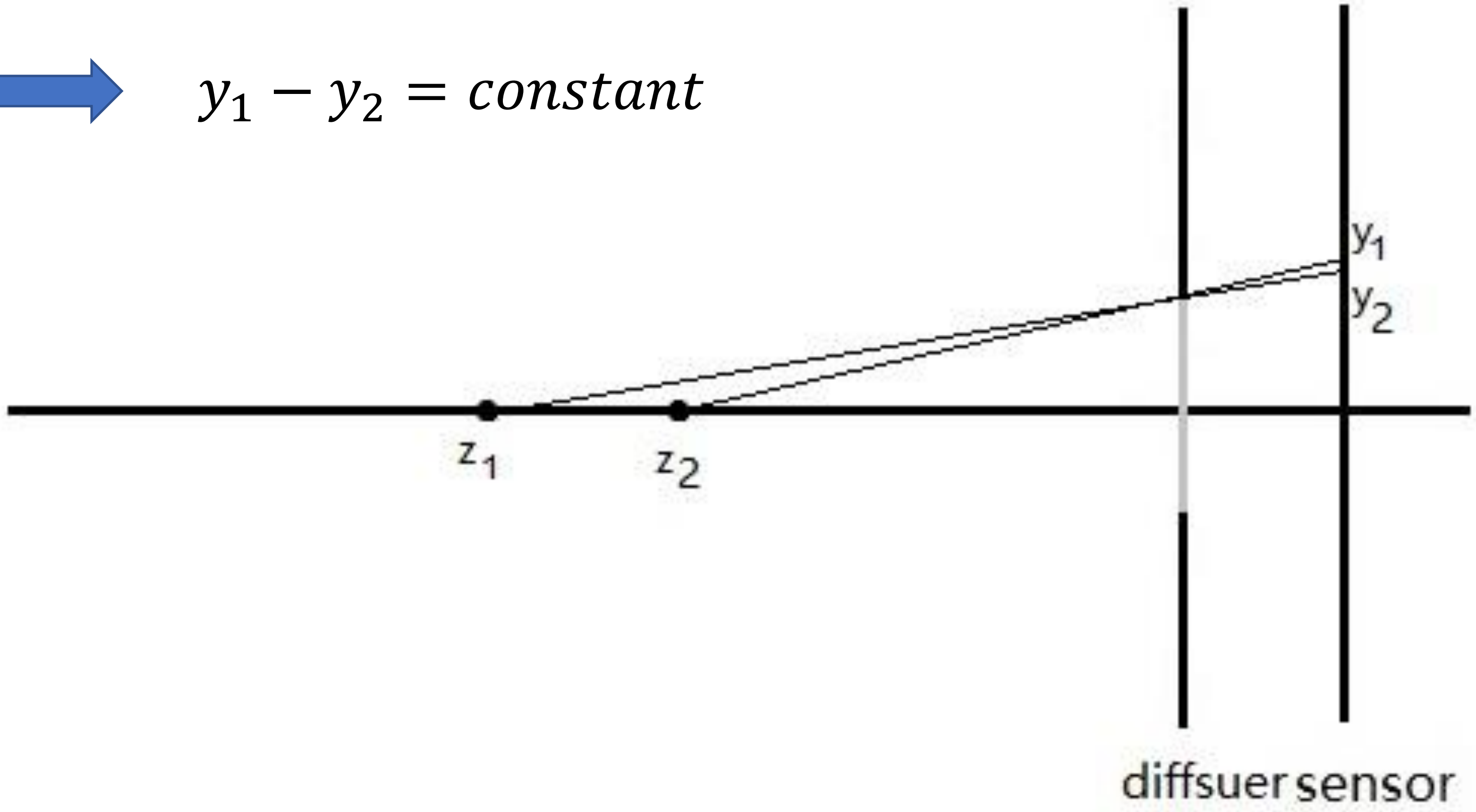


System Analysis

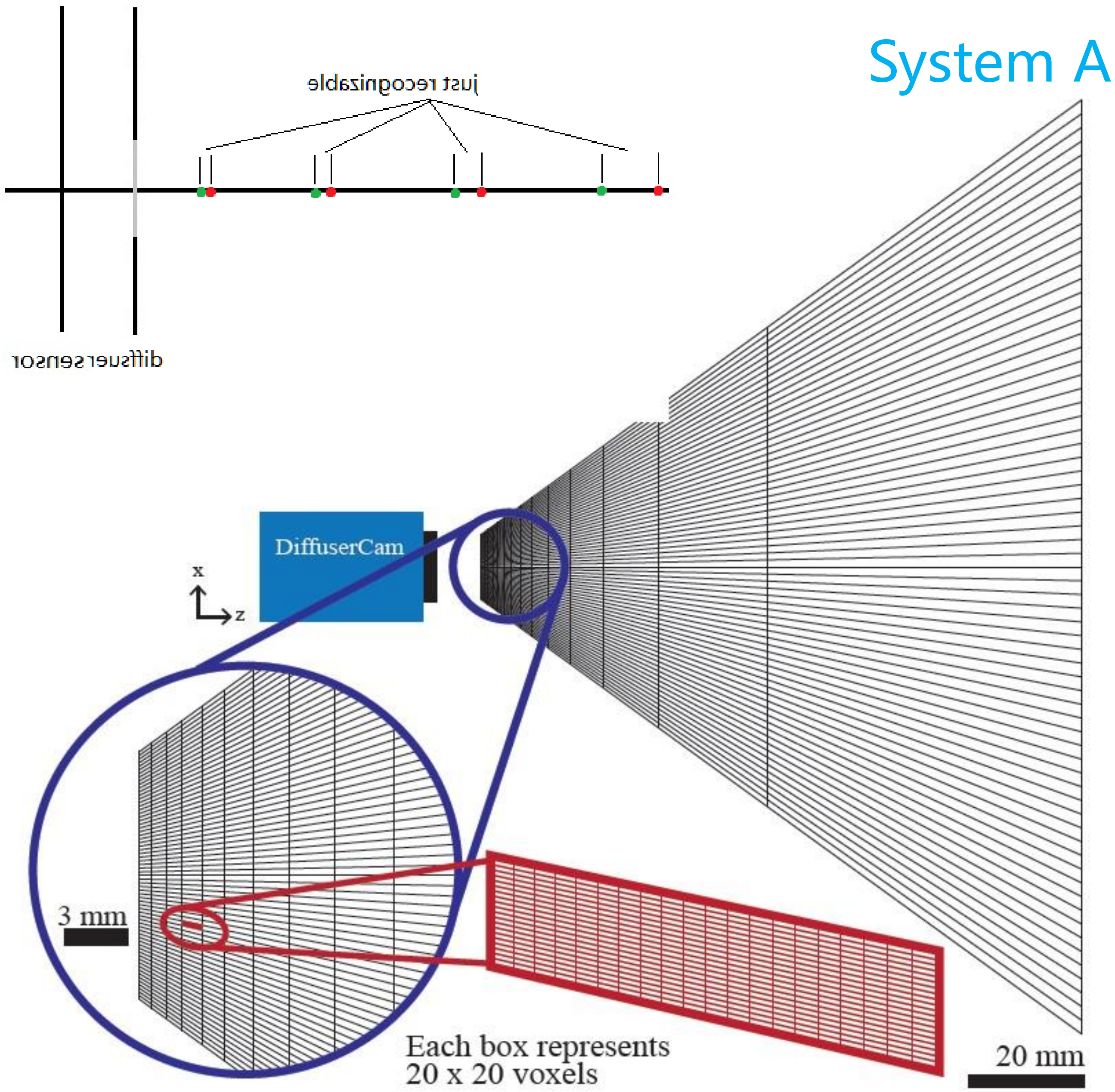


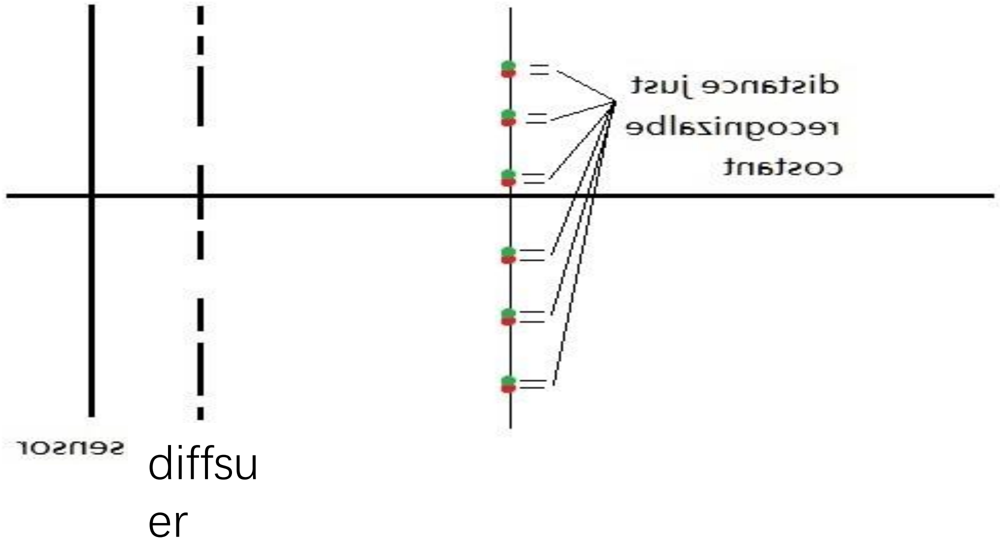
System Analysis

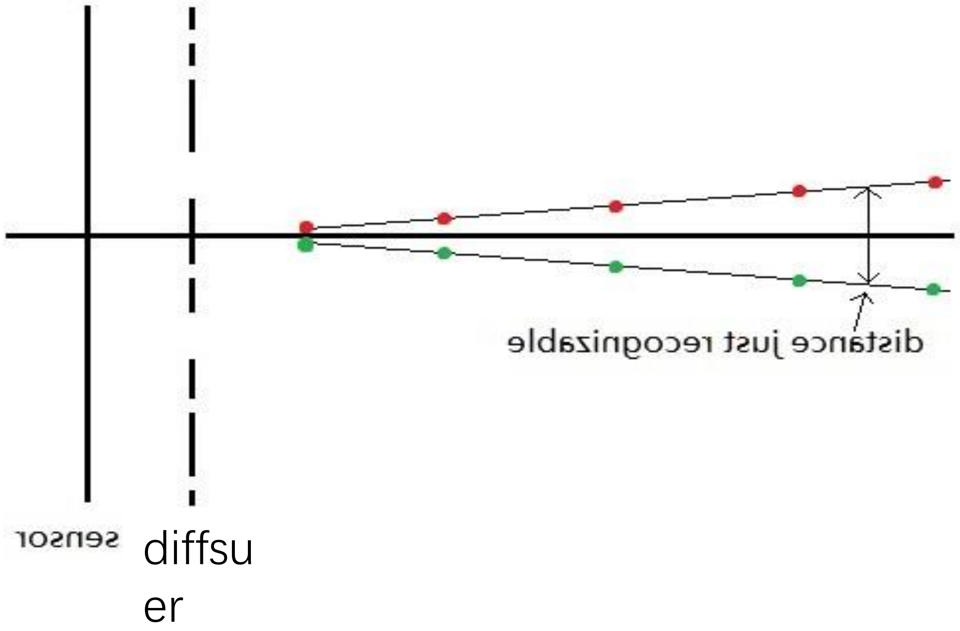
$$\frac{1}{z_1} - \frac{1}{z_2} = constant \Rightarrow y_1 - y_2 = constant$$



System Analysis







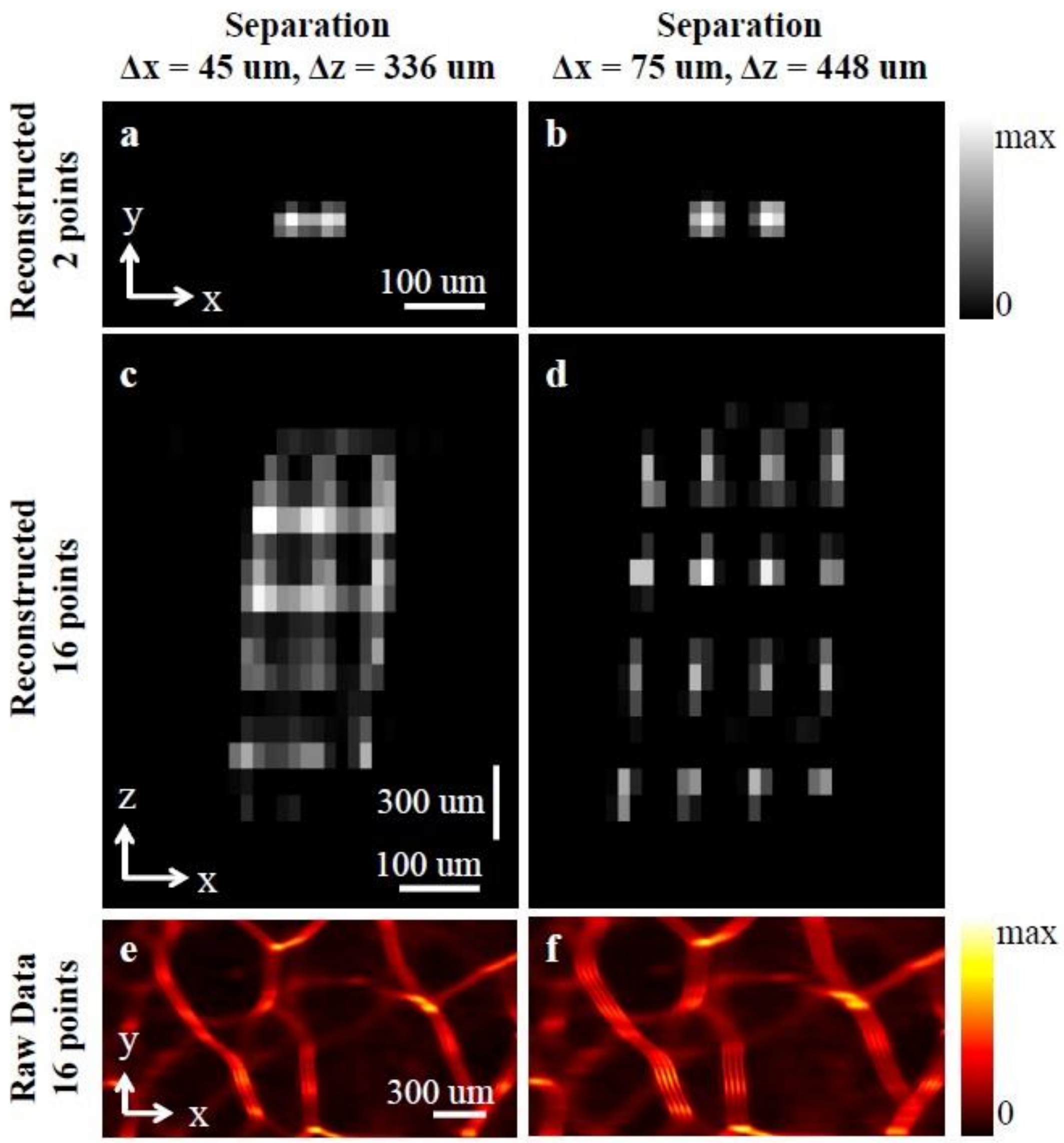
Q

System Analysis

Multi-point resolution:

Resolution varies from images

Why?



System Analysis

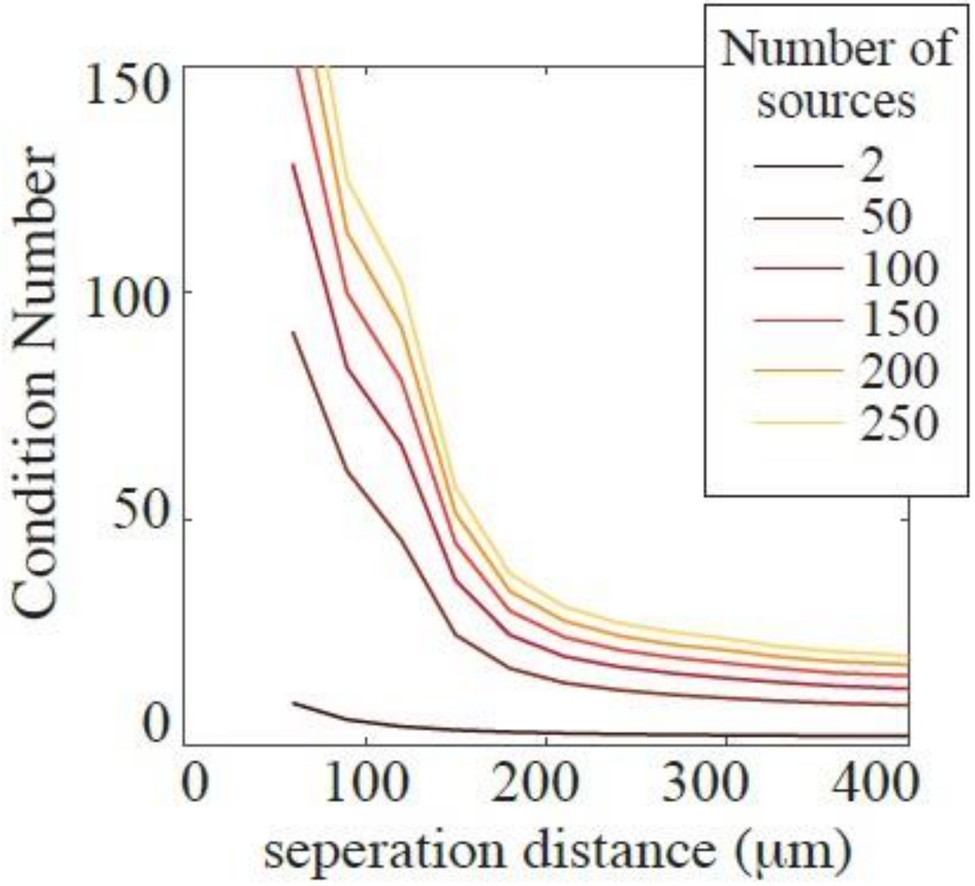
Local condition number theory

Sources are contiguous block and close

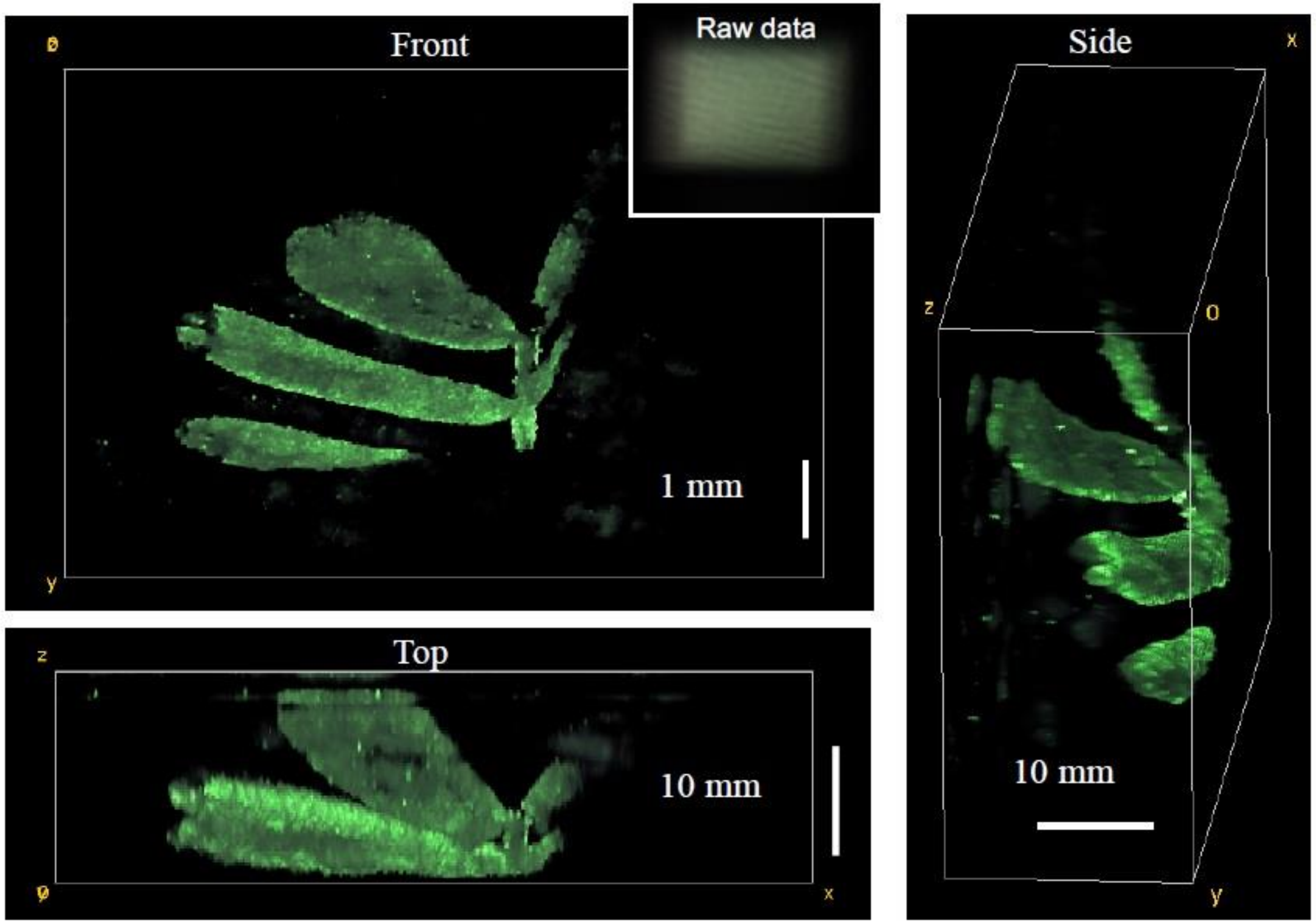
sub-matrices of H are ill-conditioned

reconstruction problem is ill-posed

Noise sensitive and long reconstruction time



Experimental Results



3-D object!

Q

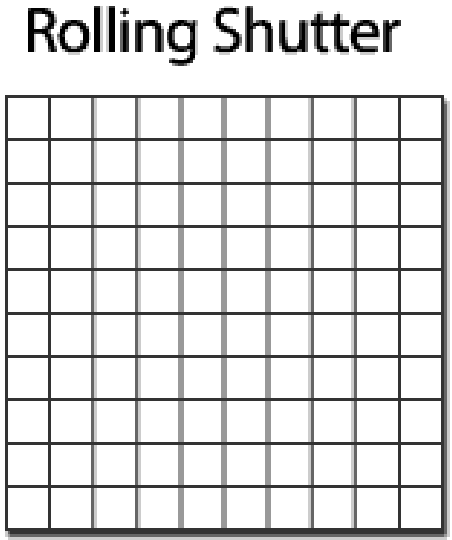
optional

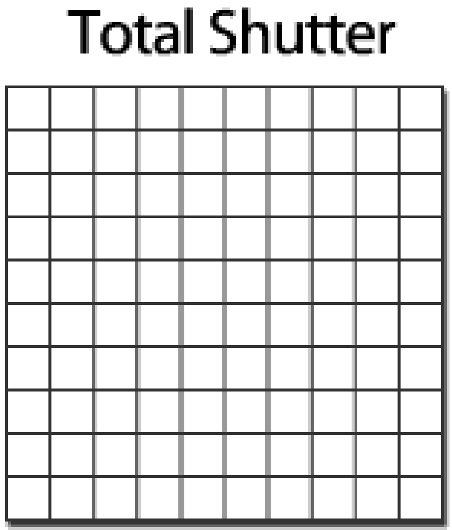
Video from Stills:
Lensless Imaging with Rolling Shutter

Author: Nick Antipa
Patrick Oare
Emrah Bostan
Ren Ng
Laura Waller

Presenter: Wang Zi

Shutter in traditional lens camera



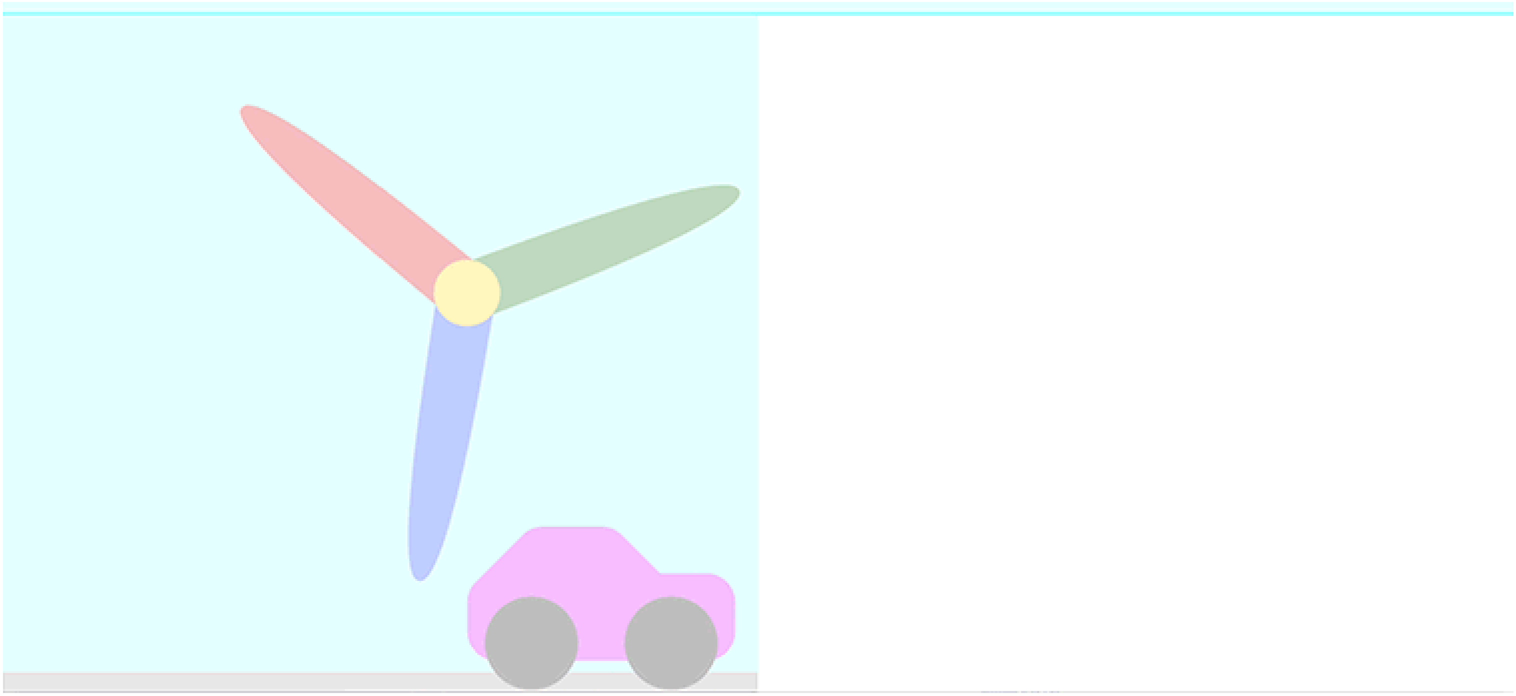


Expose line by line

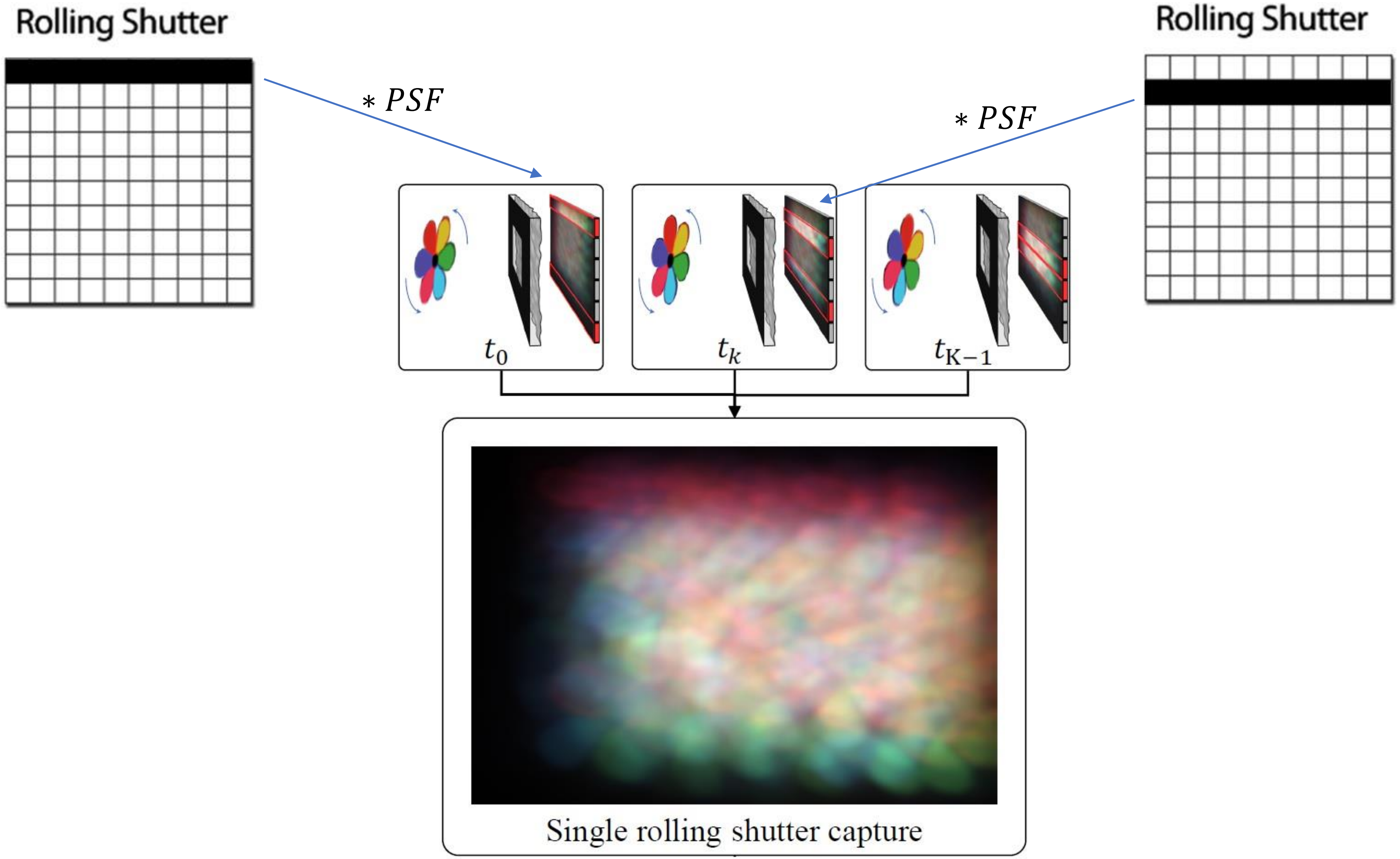
Disadvantage: jello effect

Expose at the same time

Jello effect in lens camera

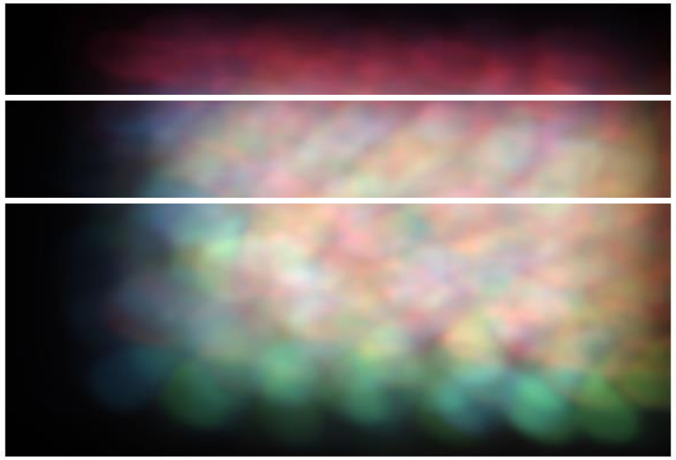


Jello effect is bad for lens camera,
but may be good for DiffuserCam



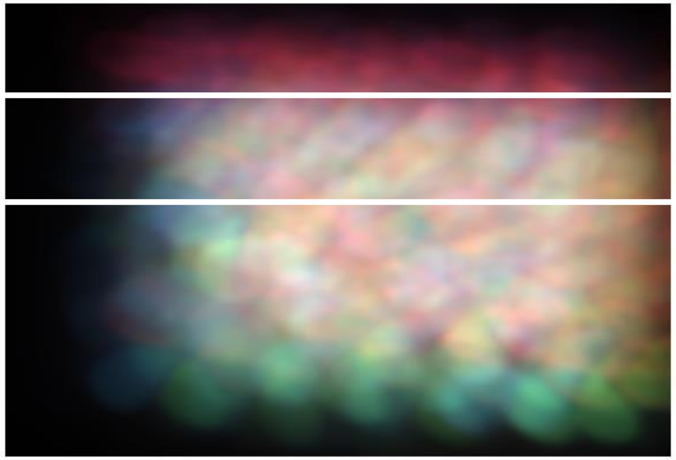
Rolling Shutter
Rolling Shutter
$* \, PSF$
$* \, PSF$
t_0
t_k
t_{K-1}
Single rolling shutter capture

Total shutter capture



t_0

+

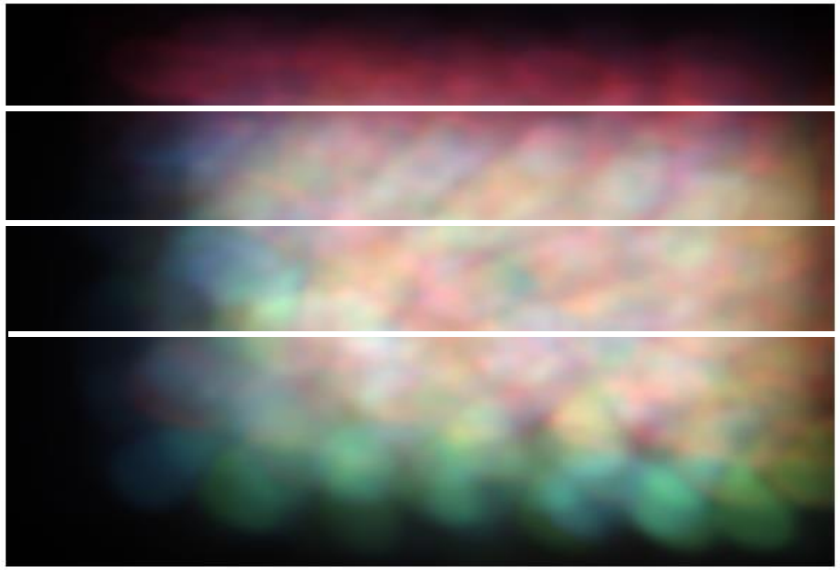


t_1

+

...

=



Single rolling shutter capture

Each capture of rolling shutter

contains information of **K frames**!

Better to recover video

Rolling shutter model

$$L(x,y) = \int_0^{\infty} S(t|x,y) \cdot \tilde{v}(x,y,t)dt$$

Lensless imaging model

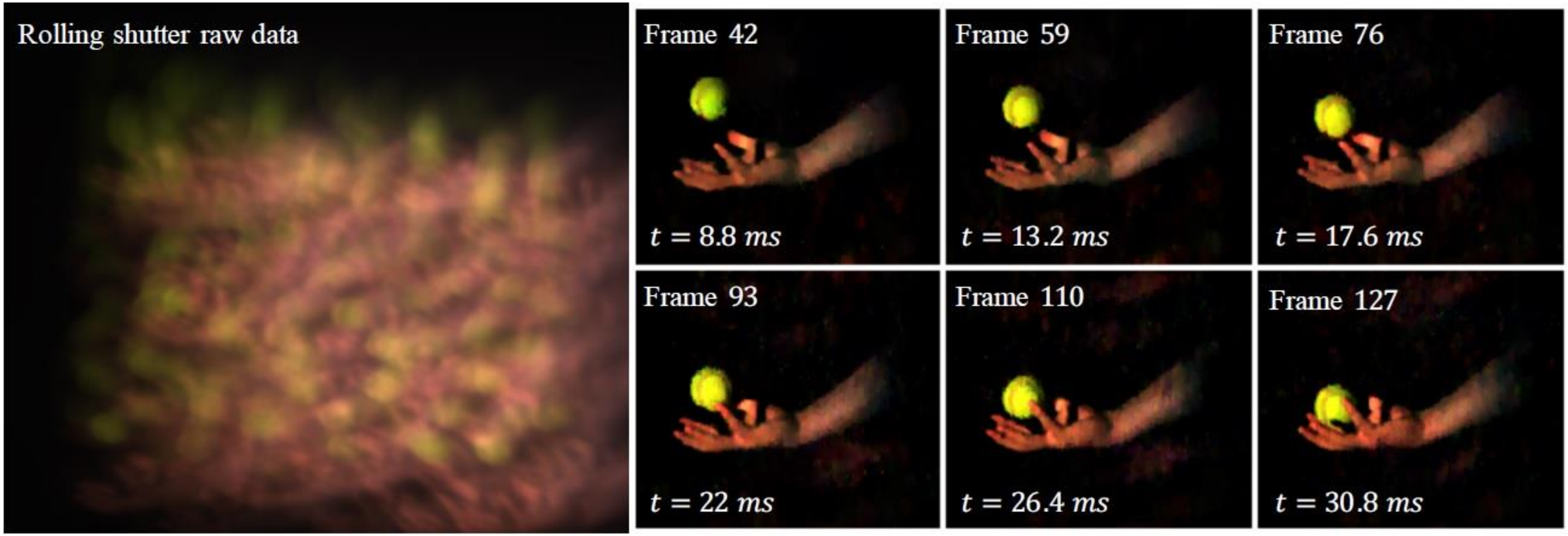
$$\tilde{v}(x,y,t) = v\left(\frac{x}{m}, \frac{y}{m}, t\right) \overset{(x,y)}{*} h(x,y)$$

Combined imaging model

$$\begin{aligned} \boldsymbol{b} &= \sum_{k=0}^{K-1} \overline{S}_k[i] \cdot \left(h[i,j] \overset{[i,j]}{*} \boldsymbol{v}[i,j,k] \right) \\ &= A\boldsymbol{v} \end{aligned}$$

Reconstruction

$$\boldsymbol{v}^* = \arg\min_{\boldsymbol{v} \geq 0} \frac{1}{2} \|A\boldsymbol{v} - \boldsymbol{b}\|_2^2 + \tau \|\nabla_{xyt}\boldsymbol{v}\|_1$$



3790 frames per second(fps)!

The daily life fps is 60 or 144

Thank you!

Thank you!

System overview

Suppose the size of object is $N_x \times N_y \times N_z$
the size of sensor is $M_x \times N_y$

Object: $v(x, y, z) \leftrightarrow$ reshaped to v, which is a vector

Image: $b(x', y') \leftrightarrow$ reshaped to b, which is a vector

$$b(x', y') = \sum_{(x,y,z)} v(x, y, z) h(x', y'; x, y, z)$$

For each fixed (x, y, z) there are $N_x N_y N_z$ variables and there are $M_x N_y$ different $h(x', y'; x, y, z)$ for different (x, y, z)

$$b = Hv$$

$$H \in R^{N_x N_y \times N_z M_x N_y}$$

Such two models are equivalent

Convolutional Forward Model

$$b(x', y') = \sum_{(x,y,z)} v(x, y, z) h(x', y'; x, y, z)$$

$$= \sum_{(x,y,z)} v(x, y, z) h(x' + mx, y' + my; 0,0, z)$$

$$= C \sum_{z} v\left(-\frac{x'}{m}, -\frac{y'}{m}, z\right) * h(x', y'; 0,0, z)$$

Where $*$ is 2-D convolution, C is a crop matrix because the size of the result of convolution is larger than the original matrix.

We can rewrite it as matrix form:

$$b = DMv$$

Where D is a diagonal matrix, M is a convolution matrix

Inverse Algorithm

$$u^{k+1} \leftarrow \mathcal{T}_{\frac{\lambda}{\mu_2}}\left(\Psi \mathbf{v}^k + \eta^k/\mu_2\right)$$
$$v^{k+1} \leftarrow (\mathbf{D}^\intercal \mathbf{D} + \mu_1 I)^{-1}\left(\xi^k + \mu_1 \mathbf{M}\mathbf{v}^k + \mathbf{D}^\intercal \mathbf{b}\right)$$
$$w^{k+1} \leftarrow \max\left(\rho^k/\mu_3 + \mathbf{v}^k, 0\right)$$
$$\mathbf{v}^{k+1} \leftarrow (\mu_1 \mathbf{M}^\intercal \mathbf{M} + \mu_2 \Psi^\intercal \Psi + \mu_3 I)^{-1} r^k$$
$$\xi^{k+1} \leftarrow \xi^k + \mu_1(\mathbf{M}\mathbf{v}^{k+1} - v^{k+1})$$
$$\eta^{k+1} \leftarrow \eta^k + \mu_2(\Psi \mathbf{v}^{k+1} - u^{k+1})$$
$$\rho^{k+1} \leftarrow \rho^k + \mu_3(\mathbf{v}^{k+1} - w^{k+1}),$$

where

$$r^k = (\mu_3 w^{k+1} - \rho^k) + \Psi^\intercal\left(\mu_2 u^{k+1} - \eta^k\right) + \mathbf{M}^\intercal\left(\mu_1 v^{k+1} - \xi^k\right)$$

D is diagonal

so $(D^T D + \mu_1 I)$ is also diagonal

The complexity is $O(n)$

$(\mu_1 M^T M + \mu_2 \Psi^T \Psi + \mu_3 I)$ is

diagonalized by 3-D DFT matrix

The complexity is $O(n^3 logn)$

Lens camera with rolling shutter

scene

↓ Imaging

Sensor measurements

↓ **Post-processing**

images

Lensless camera

scene

↓ Imaging

Sensor measurements

↓ **Reconstruction**

images

Lensless camera with rolling shutter

scene

↓ Imaging

Sensor measurements

↓ **Reconstruction including post-processing**

images

It is possible to obtain high SNR with lensless rolling shutter camera