MITSUBISHI ELECTRIC
Changes for the Better

for a greener tomorrow eco Changes

ROBUST CAMERA POSE ESTIMATION FOR IMAGE STITCHING

Laixi Shi[1], Dehong Liu[2], Jay Thornton[2]

Date 7/18/2021

[1]Carnegie Mellon University, PA, USA

[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

ICIP 2021

MITSUBISHI ELECTRIC
Changes for the Better

for a greener tomorrow eco Changes

Motivation: Image Stitching

- Motivation:
  - High resolution requirement, large scene.
  - Applications: Google earth mapping, panoramic image construction, video stabilization.
- Image stitching: fuse a collection of overlapped images to achieve a broad view of a scene with satisfactory resolution.
  - Pixel-based/Feature based

Challenges

- Essential Task : estimate the camera poses of the sequence of images under fusion.

- Challenges of camera pose estimation:

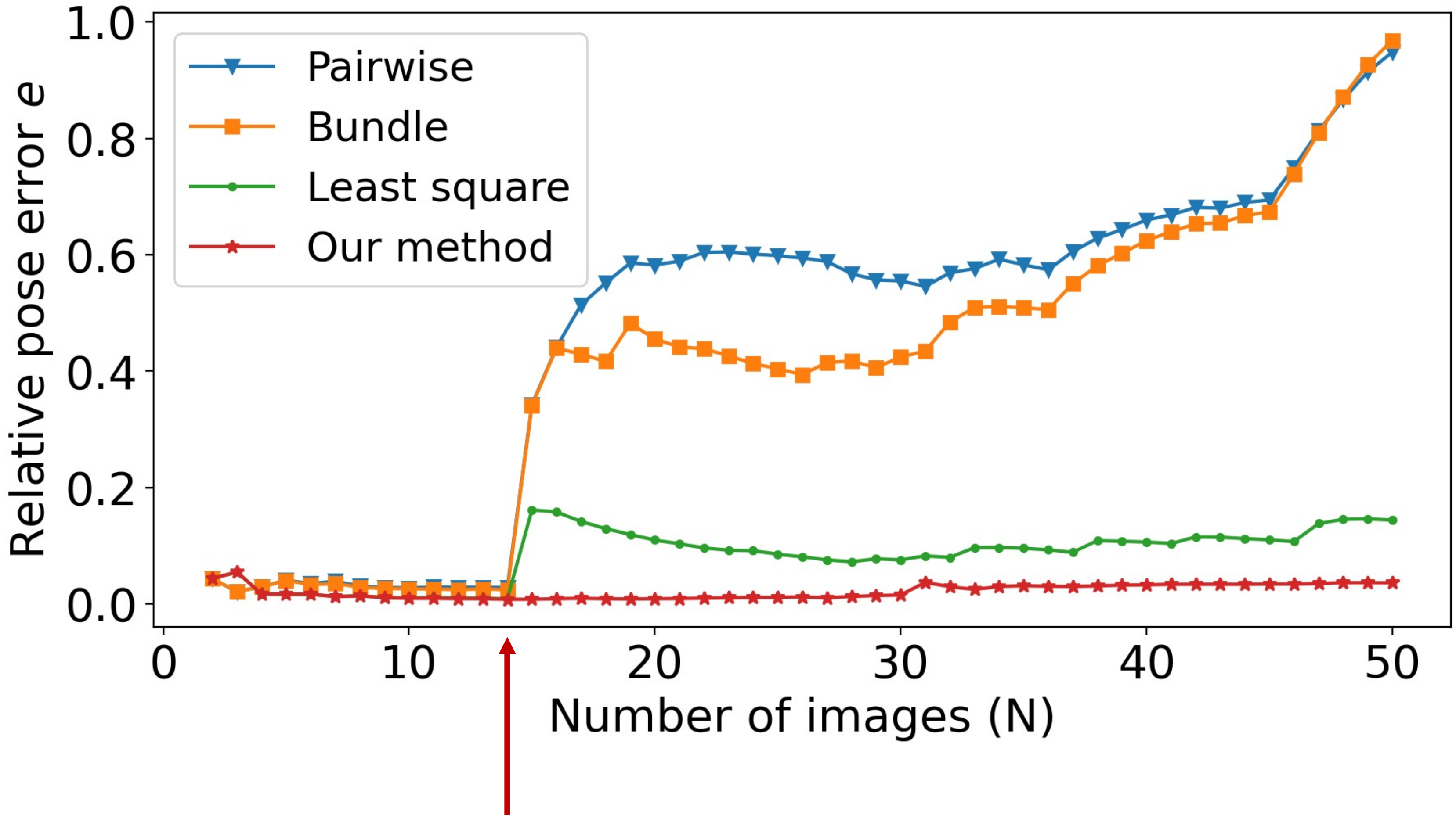
  camera pose errors → evident image visual mismatch.

  – Feature mismatch → irreversible abnormal camera pose
  – A large amount of images → accumulated pose error

Main Contribution

Methods	Challenge: abnormal pose error	Challenge: accumulated pose error
Pairwise stitching	✗	✗
Bundle adjustment	✗	MAYBE
Least square	✗	MAYBE
Ours	✓	✓

Performance



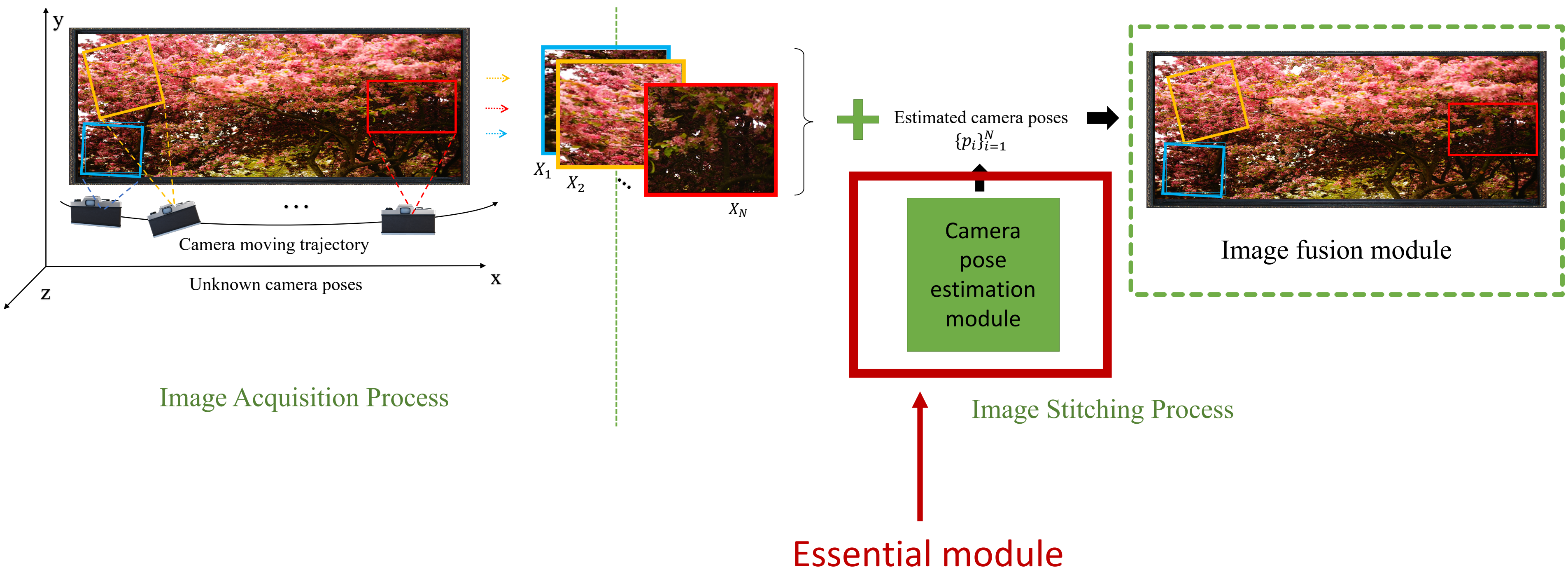
Feature mismatch leads to abnormal error

Presentation Pipeline

- Framework Pipeline
- Camera image simulation process
- Image stitching
  - Camera pose estimation module
  - Image fusion and super-resolution reconstruction module

Framework Pipeline

- Image stitching process consisting of camera pose estimation module and image fusion module.



Mitsubishi Electric
Changes for the Better

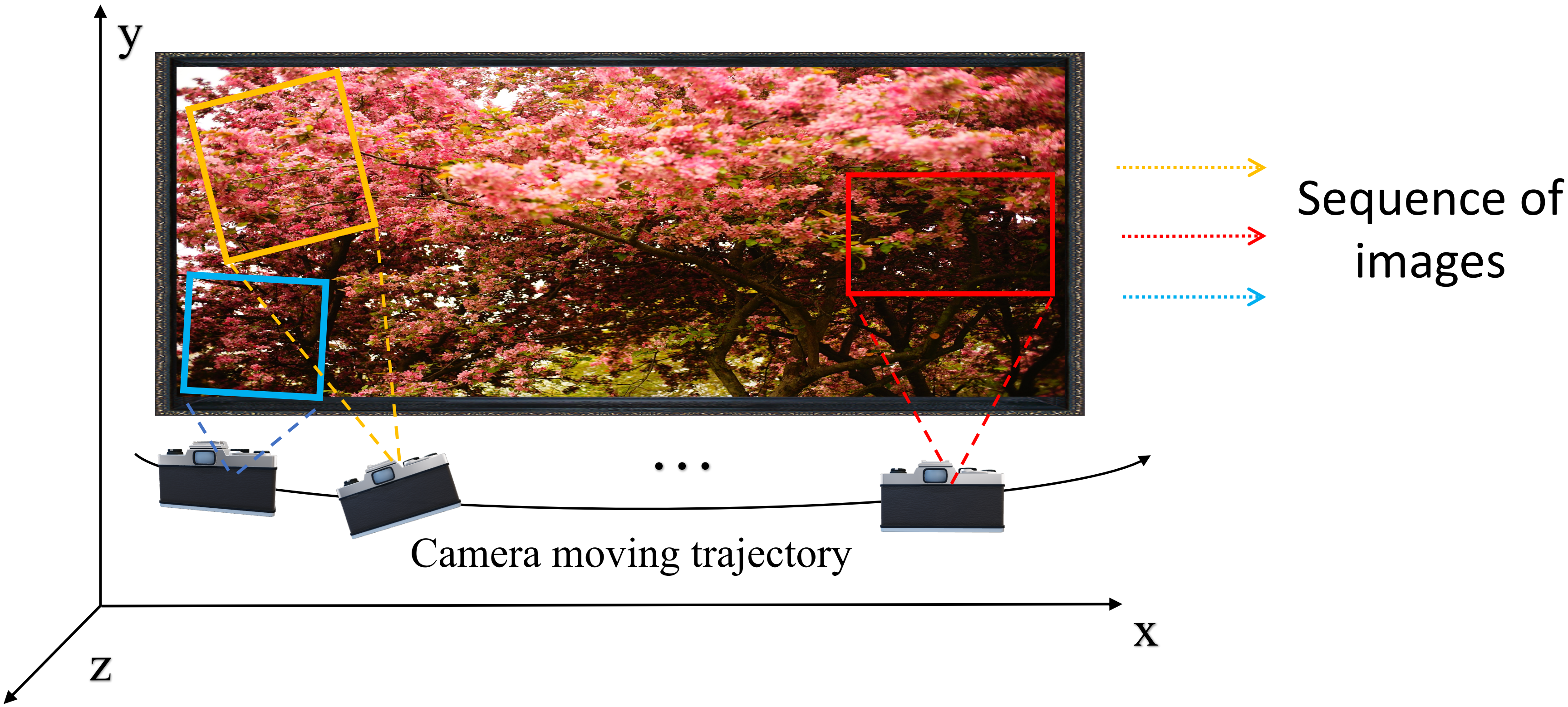
for a greener tomorrow

Presentation Pipeline

- Framework Pipeline
- Camera image simulation process
- Image stitching
  - Camera pose estimation module
  - Image fusion and super-resolution reconstruction module

Camera Image Simulation Process

- Parameters:
  - The number of images N = 50
  - Camera image resolution: 500 * 600 pixels.



Camera Image Examples

- Random camera pose: images are captured with unknown random perturbation.

Presentation Pipeline

- Framework Pipeline
- Camera image simulation process
- Image stitching
  - Camera pose estimation module
  - Image fusion and super-resolution reconstruction module

Pinhole Camera Model

$$\begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} = \frac{1}{v} \boldsymbol{P}_s \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ 1 \end{bmatrix} = \frac{1}{v} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_u \\ y_u \\ z_u \\ 1 \end{bmatrix},$$

– $\boldsymbol{R} \in \mathbb{R}^{3\times 3}, \boldsymbol{T} \in \mathbb{R}^3$: the unknown rotation and translation depend on the camera pose:

$$\boldsymbol{p} = [\theta_x, \theta_y, \theta_z, T_x, T_y, T_z]^\top \in \mathbb{R}^6$$

– $\boldsymbol{P}_s$: the known perspective matrix of the camera.
– $\boldsymbol{u} = [x_u, y_u, z_u]^\top$: a pixel $\boldsymbol{u}$ on the 3D object surface.
– $\boldsymbol{x} = [x, y]^\top$: the pixel position on the camera focal plane.
– f is the focal length, v is a pixel-dependent normalization term.

Goal: estimate all camera poses $\{p_i\}_{i=1}^N$ of N images $\{X_i\}_{i=1}^N$.

Goal: Estimate Relative camera pose matrix

- Estimate all the relative camera poses $\{p_i\}_{i=2}^{N}$ of N images (with p_1 as reference):

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \\ \cdots \\ \boldsymbol{p}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_6 \end{bmatrix} \in \mathbb{R}^{N \times 6}$$

- Relative camera pose matrix: for each dimension k of the pose

$$\boldsymbol{L}^{(k)}(i,j) = \boldsymbol{h}_k(i) - \boldsymbol{h}_k(j)$$

  - the (i,j)-th entry is the relative camera pose associated with the image pair X_i and X_j.

Camera pose estimation

→ Relative camera pose matrices estimation.

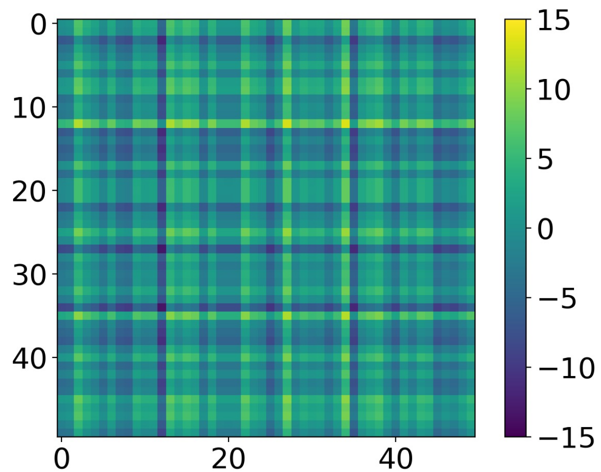
Intuition: low rank structure

- Relative camera pose matrix:

$$\boldsymbol{L}^{(k)} = \boldsymbol{h}_k \mathbf{1}^\top - \mathbf{1}\boldsymbol{h}_k^\top \in \mathbb{R}^{N\times N}$$

- For any $\boldsymbol{h}_k \neq a\mathbf{1}$,

$$\operatorname{rank}(\boldsymbol{L}^{(k)}) = \operatorname{rank}(\boldsymbol{h}_k \mathbf{1}^\top) + \operatorname{rank}(\mathbf{1}\boldsymbol{h}_k^\top) = 2$$



$\boldsymbol{L}^{(k)} \in \mathbb{R}^{N\times N}$

Problem Formulation: recover camera pose matrices

- For each dimension k of the pose: estimate relative camera pose matrices $L^{(k)}$:

$$\boldsymbol{M} \odot \widetilde{\boldsymbol{L}}^{(k)} = \boldsymbol{M} \odot (\boldsymbol{L}^{(k)} + \boldsymbol{S}^{(k)}), \quad \text{for } k = 1, \cdots, K,$$

  - $\tilde{L}^{(k)}$: the observation of the relative camera pose matrices (each (i, j)-th entry is calculated between X_i and X_j using pairwise stitching.)
  - $M \in R^{N \times N}$: the observation mask

$$\boldsymbol{M}(i,j) = \begin{cases} 1 & \text{if } \widetilde{L}^{(k)}(i,j) \text{ is observable} \\ 0 & \text{if } \widetilde{L}^{(k)}(i,j) \text{ is not observable} \end{cases}$$

  - $S^{(k)} \in R^{N \times N}$: represents sparse pose estimation errors

Problem Formulation

- Vectorize $\boldsymbol{M} \odot \widetilde{\boldsymbol{L}}^{(k)}, \boldsymbol{M} \odot \boldsymbol{L}^{(k)}, \boldsymbol{M} \odot \boldsymbol{S}^{(k)}$:

$$\widetilde{\boldsymbol{l}}^{(k)} = \mathrm{vec}(\{\widetilde{\boldsymbol{L}}^{(k)}(i,j)|_{\boldsymbol{M}(i,j)=1}\}) \in \mathbb{R}^{|\boldsymbol{M}|},$$

$$\boldsymbol{s}^{(k)} = \mathrm{vec}(\{\boldsymbol{S}^{(k)}(i,j)|_{\boldsymbol{M}(i,j)=1}\}) \in \mathbb{R}^{|\boldsymbol{M}|},$$

$$\boldsymbol{l}^{(k)} = \mathrm{vec}(\{\boldsymbol{L}^{(k)}(i,j)|_{\boldsymbol{M}(i,j)=1}\}) \in \mathbb{R}^{|\boldsymbol{M}|}$$

  - where $\boldsymbol{l}^{(k)} = \boldsymbol{A}\boldsymbol{h}_k$

- Concatenate K=6 dimensions of a camera pose:

$$\widetilde{\boldsymbol{L}} = [\tilde{\boldsymbol{l}}^{(1)}, \tilde{\boldsymbol{l}}^{(2)}, \cdots, \tilde{\boldsymbol{l}}^{(K)}] \in \mathbb{R}^{|\boldsymbol{M}| \times K},$$

$$\boldsymbol{S} = [\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}, \cdots, \boldsymbol{s}^{(K)}] \in \mathbb{R}^{|\boldsymbol{M}| \times K},$$

$$\boldsymbol{L} = [\boldsymbol{l}^{(1)}, \boldsymbol{l}^{(2)}, \cdots, \boldsymbol{l}^{(K)}] = \boldsymbol{A}\boldsymbol{P} \in \mathbb{R}^{|\boldsymbol{M}| \times K}.$$

Problem Formulation

- Optimization problem formulation:

$$\min_{\boldsymbol{S},\boldsymbol{P}} \frac{1}{2}\left\|\left(\widetilde{\boldsymbol{L}} - \boldsymbol{A}\boldsymbol{P} - \boldsymbol{S}\right)\boldsymbol{W}\right\|_F^2 + \lambda \|\boldsymbol{S}\boldsymbol{W}\|_{2,1}$$

Data fidelity

Joint sparsity regularization

– $W = \mathrm{diag}(w = [w_1, \cdots, w_6]^\top) \in R^{6\times 6}$: the magnitude normalization parameter for different dimensions of a camera pose with

$$w_k = \frac{N-1}{\sum_{i=1}^{N-1} |\widetilde{\boldsymbol{L}}^{(k)}(i, i+1)|}$$

– $\|\cdot\|_{2,1}$: the $\ell_{2,1}$ norm, i.e., for any matrix $\boldsymbol{Q} \in R^{I\times J}$

$$\|\boldsymbol{Q}\|_{2,1} = \sum_{i=1}^{I} \sqrt{\sum_{j=1}^{J} [Q(i,j)]^2}$$

MITSUBISHI ELECTRIC
Changes for the Better

for a greener tomorrow

Problem Formulation

- Alternating minimization method:

$$\min_{\boldsymbol{S},\boldsymbol{P}} \frac{1}{2} \left\| \left(\widetilde{\boldsymbol{L}} - \boldsymbol{A}\boldsymbol{P} - \boldsymbol{S} \right) \boldsymbol{W} \right\|_F^2 + \lambda \left\| \boldsymbol{S}\boldsymbol{W} \right\|_{2,1}$$

  - Subproblem of P: standard least square

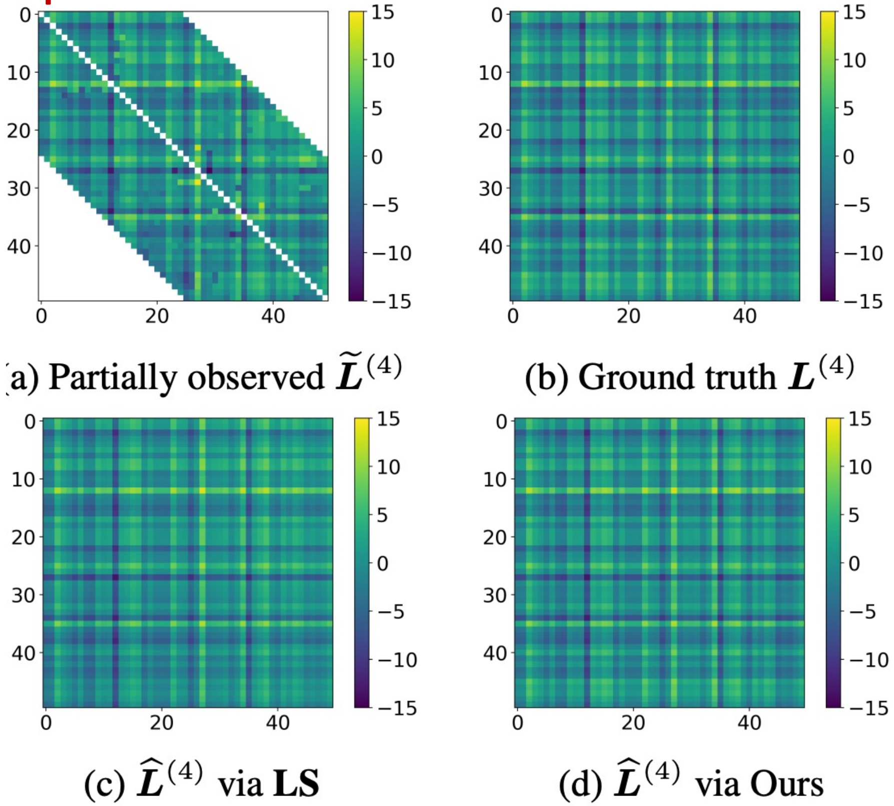
$$\boldsymbol{P}^{(t)} = \boldsymbol{A}^{\dagger} \left(\widetilde{\boldsymbol{L}} - \boldsymbol{S}^{(t-1)} \right)$$

  - Subproblem of S: row-dependent soft-thresholding

$$\boldsymbol{S}_{i,:}^{(t)} = (\widetilde{\boldsymbol{L}} - \boldsymbol{A}\boldsymbol{P}^{(t)})_{i,:} \odot \max \left(0, 1 - \frac{\lambda}{\left\| (\widetilde{\boldsymbol{L}} - \boldsymbol{A}\boldsymbol{P}^{(t)})_{i,:} \boldsymbol{W} \right\|_2} \right)$$

Camera Pose Estimation Results

- Relative camera pose matrices estimation results:



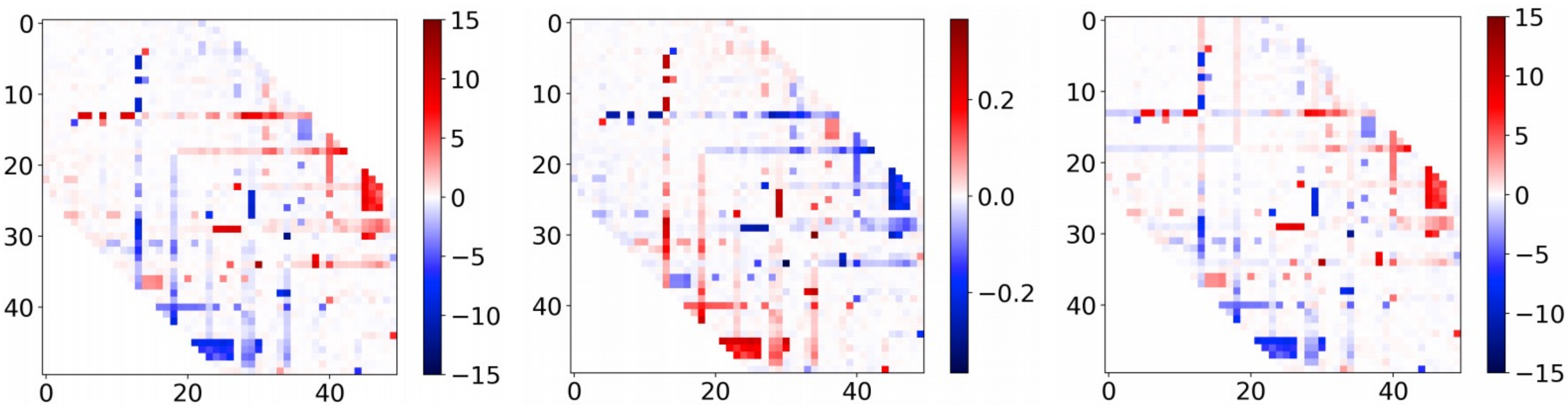
(a) Partially observed $\widetilde{\boldsymbol{L}}^{(4)}$ (b) Ground truth $\boldsymbol{L}^{(4)}$

(c) $\widehat{\boldsymbol{L}}^{(4)}$ via **LS** (d) $\widehat{\boldsymbol{L}}^{(4)}$ via Ours

- Camera pose estimation results:

	Pairwise	Bundle	LS	Ours
Relative pose error	0.948	0.968	0.140	**0.037**

Abnormal Camera Pose Error Estimation



(a) $\boldsymbol{S}^{(4)}$ (b) $\boldsymbol{S}^{(3)}$ (c) $\widehat{\boldsymbol{S}}^{(4)}$ via Ours

Side-product:

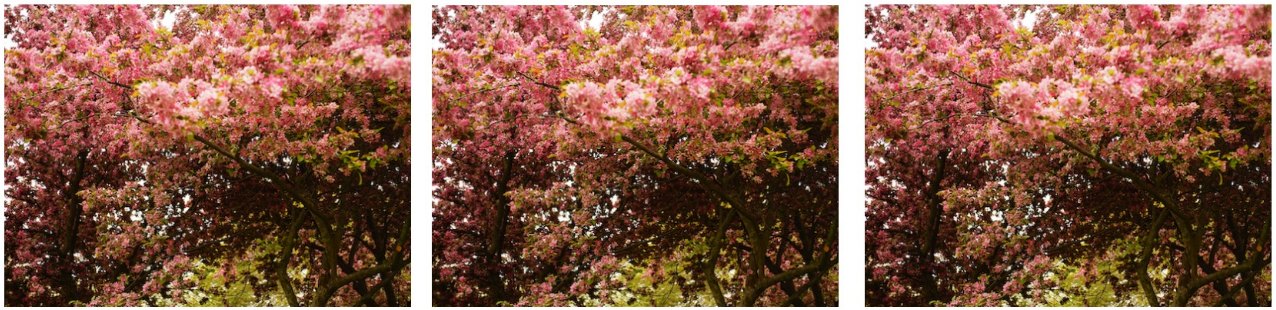
abnormal camera pose error detection

Presentation Pipeline

- Framework Pipeline
- Camera image simulation process
- Image stitching
  - Camera pose estimation module
  - Image fusion and super-resolution reconstruction module

Image stitching Results

- Final stitching results:



(a) True image $\boldsymbol{U}$ (b) $\widehat{\boldsymbol{U}}$ via **LS** (c) $\widehat{\boldsymbol{U}}$ via **Ours**

(d) Local area of (a) (e) Local area of (b) (f) Local area of (c)

- PSNR results:

	Pairwise	Bundle	LS	Ours
Relative pose error	0.948	0.968	0.140	**0.037**
PSNR of $\widehat{\boldsymbol{U}}$ (dB)	19.23	20.94	26.68	**30.29**

Conclusion

- We propose a robust camera pose estimation method for stitching a large collection of images of a 3D surface with known geometry.
  - constructed a partially observed relative pose matrix for each parameter of camera poses.
  - Estimate poses by recovering the rank-2 matrix of relative camera poses and a sparse matrix of camera pose errors by exploiting the joint sparsity of camera pose errors.
- The proposed method is capable of yielding robust camera pose estimates even if
  - abnormal pose estimates appears
  - camera pose error accumulate during the process
  - numerous images are under fusion.

Thank you for listening!

Contact:
laixis@andrew.cmu.edu
liudh@merl.com