SIGGRAPH ASIA 2016 MACAO

Sponsored by acm

# Video Stitching for Handheld Inputs via Combined Video Stabilization

Tan Su[1], Yongwei Nie[2], Zhensong Zhang[1], Hanqiu Sun[1], Guiqing Li[2]

[1]The Chinese University of Hong Kong
[2]South China University of Technology

SA2016.SIGGRAPH.ORG

# VIDEO STITCHING: OVERVIEW

- Video Stitching
  - Larger field of view: more immersive viewing experience
  - One of the key techniques in producing panoramic video

SIGGRAPH ASIA 2016 MACAO

# VIDEO STITCHING: OVERVIEW

- Video Stitching
  - Larger field of view: more immersive viewing experience
  - One of the key techniques in producing panoramic video

- Types of Input Settings
  1. Pre-calibrated camera arrays
     - 360 camera
  2. Unstructured camera
     - Perazzi *et. al.* – *ctured Camera Arrays”* [1]
     - Jiang and Gu – *emporal Content-Preserving Warping”* [2]
  3. Handheld / free-mo

*GoPro Odyssey. www.gopro.com/odyssey

# VIDEO STITCHING: OVERVIEW

- Video Stitching
  - Larger field of view: more immersive viewing experience
  - One of the key techniques in producing panoramic video

- Types of Input Settings
  1. Pre-calibrated camera arrays
     - 360 cameras on the markets, for either profess… consumers
  2. Unstructured camera array
     - Perazzi *et. al.* – *“Panoramic Video from Unstructured Camera Arrays”* [1]
     - Jiang and Gu – *“Video Stitching with Spatial-Temporal Content-Preserving Warping”* [2]
  3. Handheld / free-moving cameras

# VIDEO STITCHING: OVERVIEW

- Video Stitching
  - Larger field of view: more immersive viewing experience
  - One of the key techniques in producing panoramic video

- Types of Input Settings
  1. Pre-calibrated camera arrays
     - 360 cameras on the markets, for either professionals and average consumers
  2. Unstructured camera array
     - Perazzi *et. al.* – *"Panoramic Video from Unstructured Camera Arrays"* [1]
     - Jiang and Gu – *"Video Stitching with Spatial-Temporal Content-Preserving Warping"* [2]
  3. Handheld / free-moving cameras

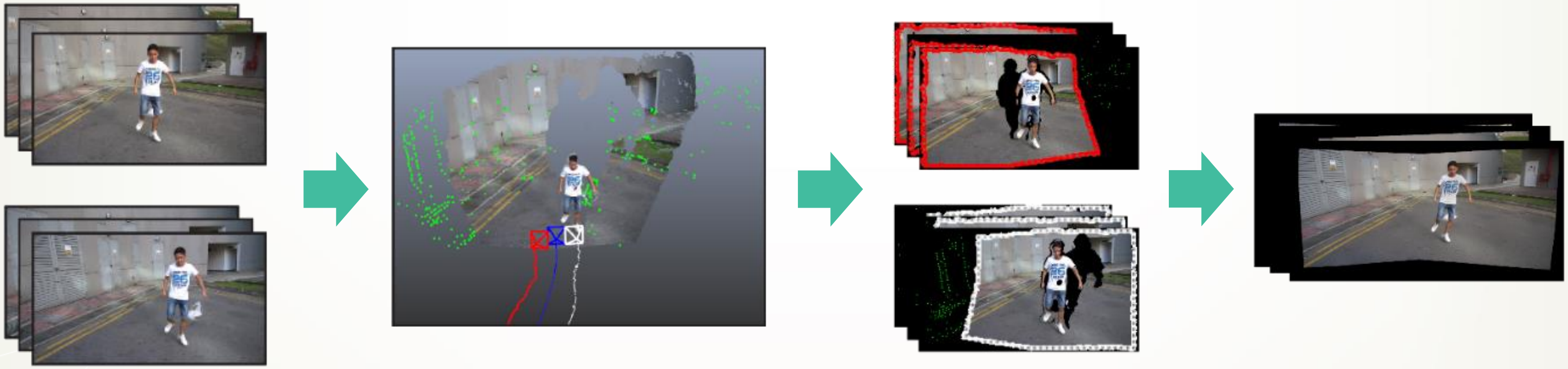
# DIFFICULTIES OF HANDHELD CAMERAS INPUTS

- Difficulties of Handheld Cameras Inputs
  - The relative pose and parallax between cameras are dynamic
  - Input videos are shaky
- Frame-by-frame image stitching?
  - Cannot preserve temporal coherence – jittering

# DIFFICULTIES OF HANDHELD CAMERAS INPUTS

Previous work that handle free-moving cameras

- Lin *et. al.* – *“Seamless Video Stitching from Handheld Camera Inputs”* [3]



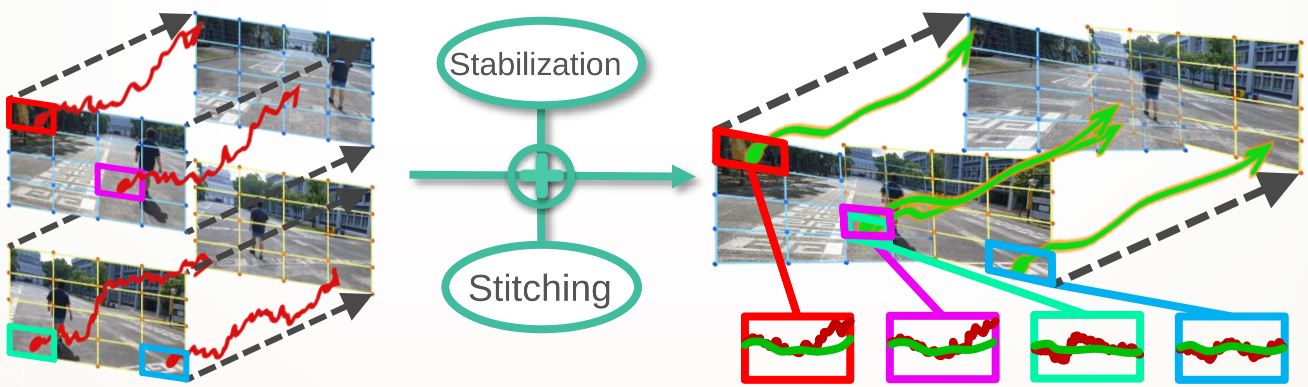
SIGGRAPH ASIA 2016 MACAO

# Our Method

# OUR METHOD

Highlights of our method

- Stitching and stabilization in a unified framework
- ONE global alignment operation for ALL the frames
- Preserving temporal coherence



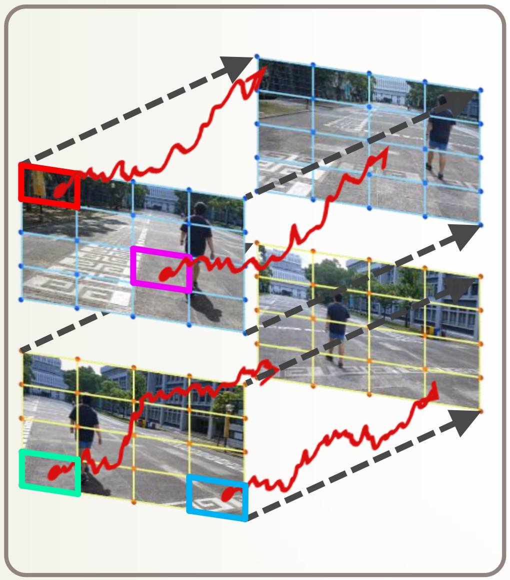
# OUR METHOD

## Video Capturing setting

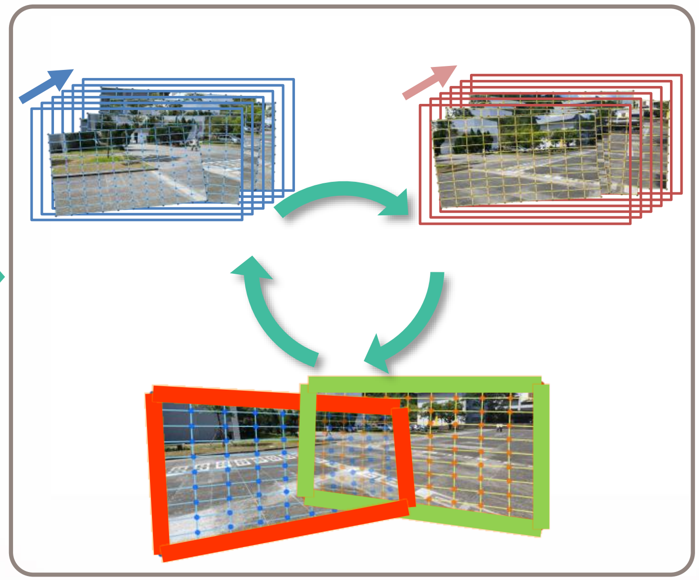
- Use two handheld cameras (DVs or phones) to capture different parts of the same scene
- Cameras can be freely moved, but the parallax cannot be too large

# VIDEO STITCHING FRAMEWORK PIPELINE

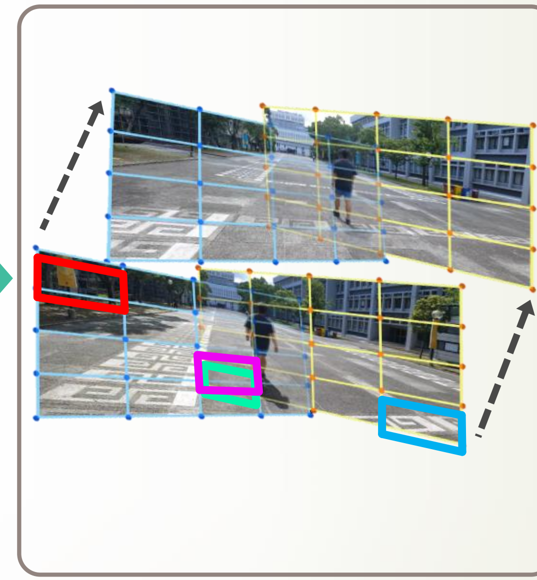
## Framework Pipeline



Bundled Camera Path Estimation



Iterative Camera Path Optimization

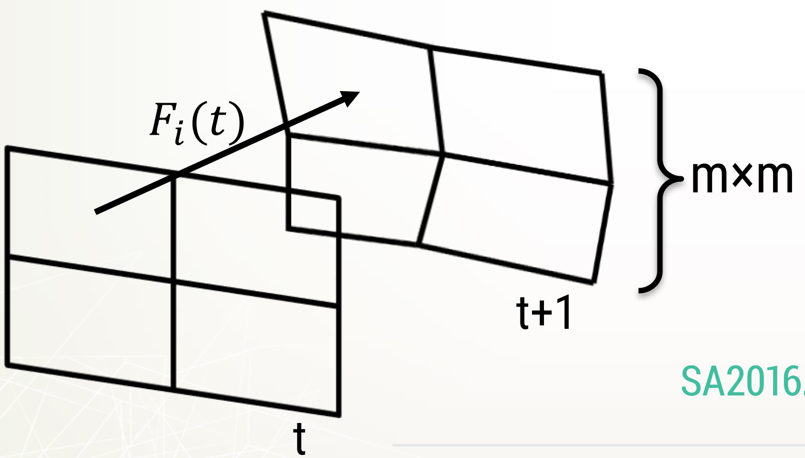


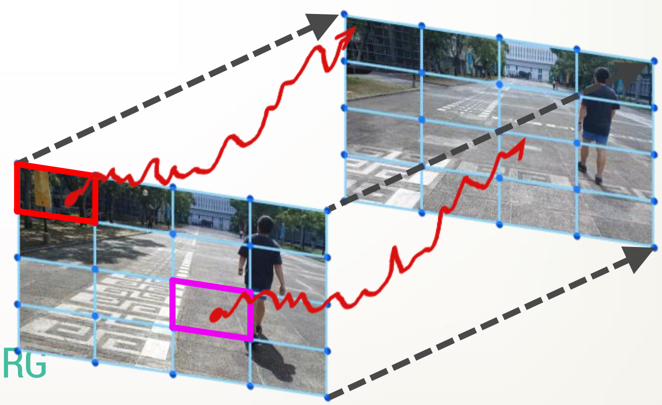
Warping frames to new camera paths

# Camera Path Estimation

- The "*Bundled Camera Path*"
    - Liu *et. al.* –"*Bundled Camera Paths for Video Stabilization*" [4]
    - Camera paths are computed as $m \times m$ accumulated homography transformation

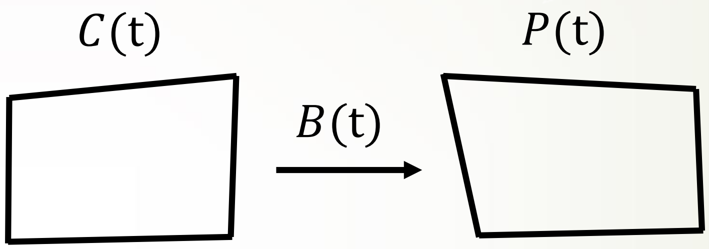
$$C_i(t) = F_i(t) \cdot F_i(t-1) \cdots F_i(1), 1 \leq t \leq T, 1 \leq i \leq m^2$$



# BUNDLED PATH STABILIZATION

Optimize $C_i(t) \rightarrow P_i(t)$




$$E^{stable} = \sum_i \sum_t \left( \|P_i(t) - C_i(t)\|^2 + \lambda \cdot \sum_{r \in \Omega_t} \omega_{t,r} \cdot \|P_i(t) - P_i(r)\|^2 + \sum_{j \in N(i)} \|P_i(t) - P_j(t)\|^2 \right)$$

Data term

Smooth term

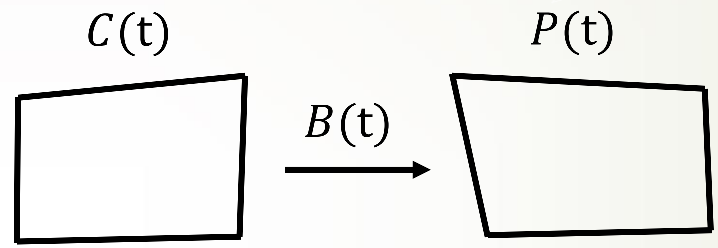
Bundle term

Stabilize frames by the warping transformation

$$B_i(t) = P_i(t) \cdot C_i(t)^{-1}$$

# BUNDLED PATH STABILIZATION

Optimize $C_i(t) \rightarrow P_i(t)$




$$E^{stable} = \sum_i \sum_t \left( \|P_i(t) - C_i(t)\|^2 + \lambda \cdot \sum_{r \in \Omega_t} \omega_{t,r} \cdot \|P_i(t) - P_i(r)\|^2 + \sum_{j \in N(i)} \left\|P_i(t) - B_j(t) C_i(t)\right\|^2 \right)$$

Data term

Smooth term

Our Improvement

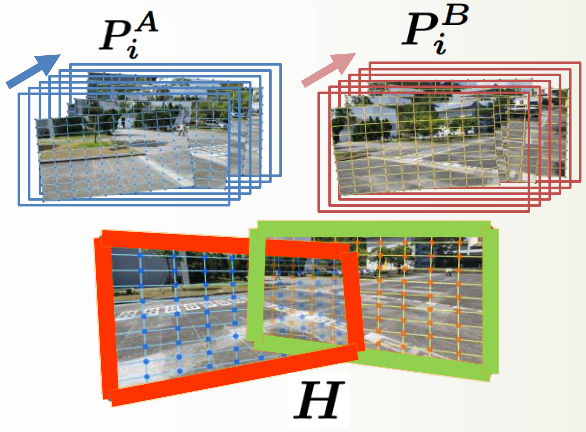
Our modification:

- Let $B_i(t)$ be similar to neighboring grids ➔ reduce distortion

$$B_j(t) = P_j(t) \cdot C_j(t)^{-1}$$

# VIDEO STITCHING VIA THE STABILIZATION

Optimize camera paths in a unified video stitching and stabilization framework




- 2+1 variables to be optimized
  - ✓ Camera path $P_i^A$ and $P_i^B$
  - ✓ ONE single and global homography $H$ for alignment
    - Enough for stitching ALL the frames
    - Temporal coherence can be well preserved

# VIDEO STITCHING VIA THE STABILIZATION

- Optimize two camera paths $P^A$ and $P^B$ in a unified video stabilization framework

$$E\left(P_i^A, P_i^B, H\right) = \underbrace{E_A^{stable}(P_i^A) + E_B^{stable}(P_i^B)}_{\text{Stabilization Terms}} + \beta \cdot \underbrace{E^{stitch}(P_i^A, P_i^B, H)}_{\text{Stitching Term}}$$

# VIDEO STITCHING VIA THE STABILIZATION

- Stitching Term
  - Ensure the matched SIFT features as close as possible in the stabilized and aligned videos

$$E^{stitch}(P_i^A, P_i^B, H) = \sum_t \sum_k \left\| P_u^A(t) \cdot C_u^A(t)^{-1} \cdot p_k^A(t) - H \cdot P_v^B(t) \cdot C_v^B(t)^{-1} \cdot p_k^B(t) \right\|^2$$

Feature's new location in video *A*

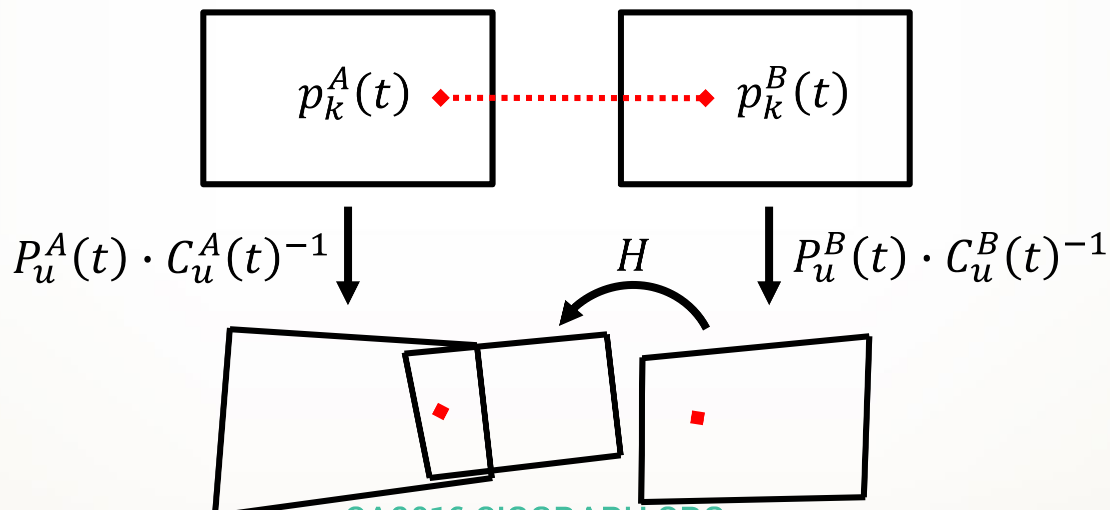
Feature's new location in video *B*

# VIDEO STITCHING VIA THE STABILIZATION

$$E^{stitch}(P_i^A, P_i^B, H) = \sum_t \sum_k \left\| P_u^A(t) \cdot C_u^A(t)^{-1} \cdot p_k^A(t) - H \cdot P_v^B(t) \cdot C_v^B(t)^{-1} \cdot p_k^B(t) \right\|^2$$

Feature's new location in video *A*

Feature's new location in video *B*

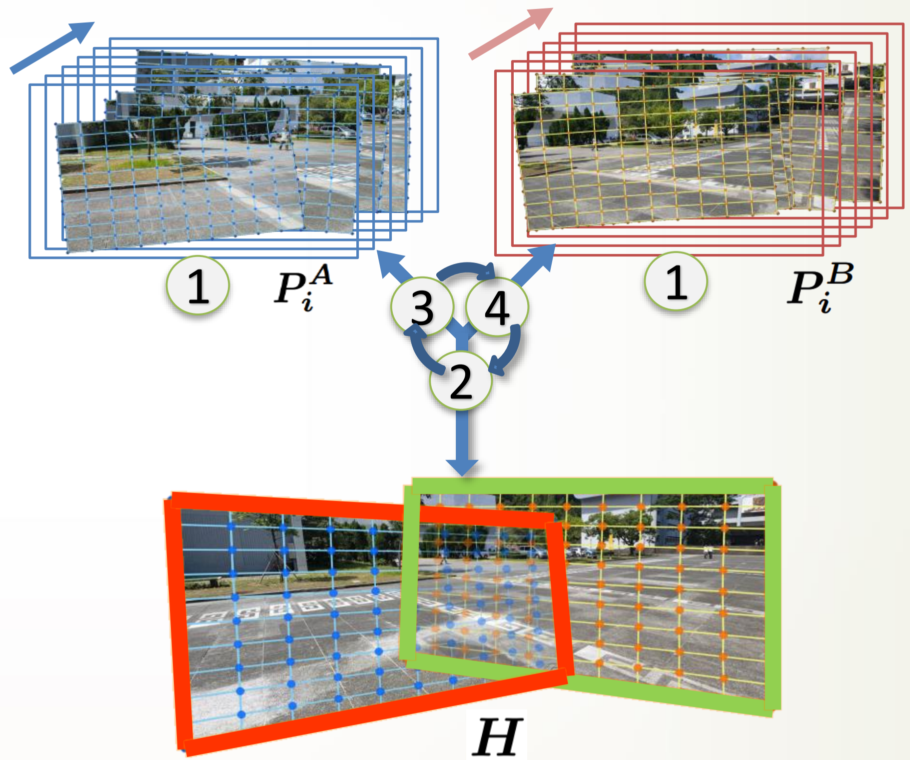


# ITERATIVE OPTIMIZATION STRATEGY

$$E\left(P_i^A, P_i^B, H\right) = E_A^{stable}(P_i^A) + E_B^{stable}(P_i^B) + \beta \cdot E^{stitch}(P_i^A, P_i^B, H)$$

1. Optimize $E_A^{stable}$ and $E_B^{stable}$ separately at very beginning, obtain initially optimized $P_i^A$ and $P_i^B$;
2. Optimize $H$ with temporally optimized $P_i^A$ and $P_i^B$;
3. Optimize $P_i^A$ with temporally optimized $H$ and $P_i^B$;
4. Optimize $P_i^B$ with temporally optimized $H$ and $P_i^A$;
5. Go to step 2 until convergence.

Time cost: ~20 minutes for a 15s video



# OUR METHOD

## Handling parallax

- Mesh-based camera path model
- Features that cannot be aligned are excluded after every iteration
  - Usually foreground objects
- Spatial-temporal Seam [4]
  - Eliminate ghosting effect

[4] Jiang, W. and Gu, J., 2015. Video stitching with spatial-temporal content-preserving warping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 42-48).

# Results and Comparison

# COMPARISON

Compare to the state-of-the-art image stitching method APAP [5]

[5] Zaragoza, J., Chin, T.J., Brown, M.S. and Suter, D., 2013. As-projective-as-possible image stitching with moving DLT. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2339-2346).

# COMPARISON

## Compare to LPVW with 3D reconstruction [6]

[6] Lin, K., Liu, S., Cheong, L.F. and Zeng, B., 2016, May. Seamless Video Stitching from Hand-held Camera Inputs. In *Computer Graphics Forum* (Vol. 35, No. 2, pp. 479-487).

# COMPARISON

## Compare to LPVW with 3D reconstruction [6]

[6] Lin, K., Liu, S., Cheong, L.F. and Zeng, B., 2016, May. Seamless Video Stitching from Hand-held Camera Inputs. In *Computer Graphics Forum* (Vol. 35, No. 2, pp. 479-487).

# CONCLUSION

- A lightweight video stitching framework for videos captured by handheld cameras
- Solve video stitching via a unified framework combined with stabilization
- No 3D reconstruction is needed

# THANK YOU!

Tan Su, Zhensong Zhang, Hanqiu Sun
– {tsu, zszhang, hanqiu}@cse.cuhk.edu.hk
Yongwei Nie, Guiqing Li
– {nieyongwei, ligq}@scut.edu.cn