ACCV '16 Tutorial: Large-scale 3D Reconstruction from Images

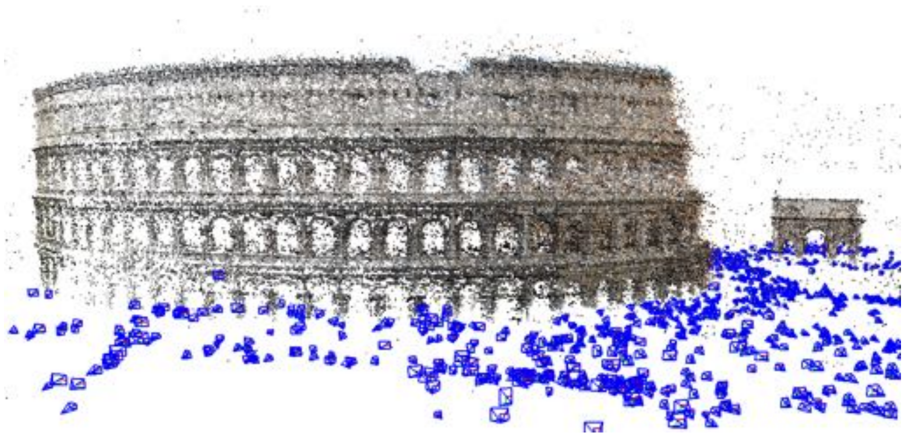
LONG QUAN, TIANWEI SHEN, JINGLU WANG

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Part I

Tianwei Shen

Large-scale Structure-from-Motion: A Modern Synthesis

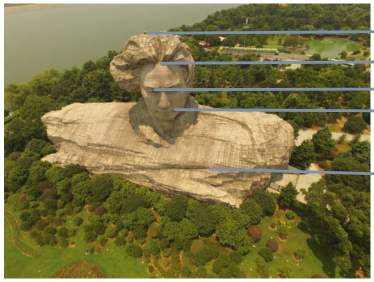


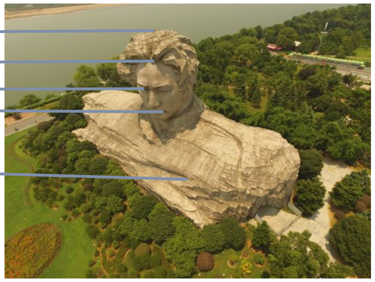
Outline

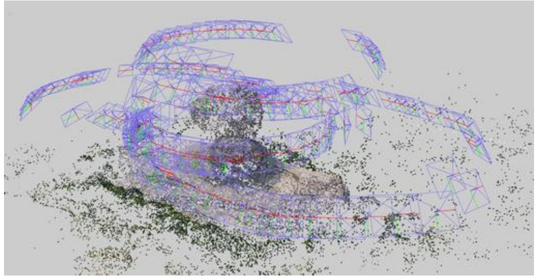
- Introduction to Structure-from-Motion (SfM)
- Component I: Feature Detection and Matching
- Component II: From Feature matches to 3D
- Component III: Large-scale Bundle Adjustment
- Applications and Future Directions

SfM - The entry point to 3D computer vision

- From pictures to 3D scenes







Pictures → Feature matching → Structure-from-Motion → Surface reconstrution

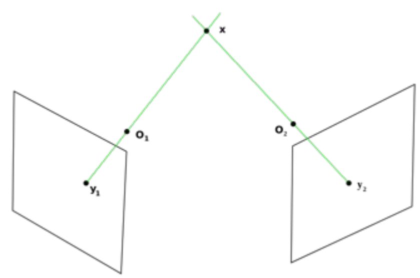
Not covered in this talk

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Imagery Credit: Hanyu@altizure.com

Notations

- Views/Frames/Images: {I_i}
- Features: 2D salient regions/blobs (edges, corners), e.g. SIFT
- Tracks: 3D point structures that correspond to 2D features in images
- Camera Intrinsic / Extrinsic: {P_i} => K [R T]
- Residual error: distance between 2D features and 3D projection
- Triangulation: the process of determining a point in 3D space given its projections onto two, or more images



A typical pipeline of SfM

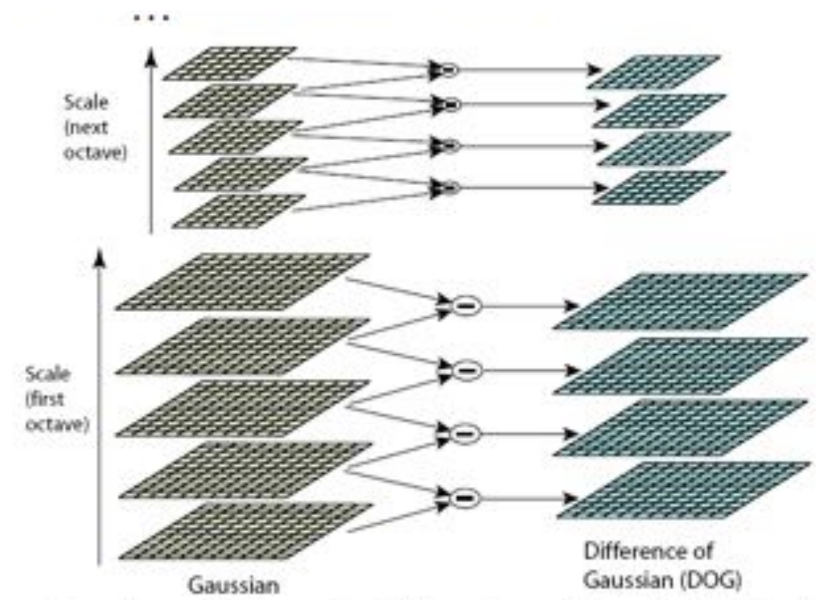
- Feature extraction: images $\{I_i\}$ → local feature collections $\{F_i\}$
- Feature matching: $\{F_i\}$ → match pairs $\{M_{ij}\}$, epipolar geometry $\{f, h, R_{ij}, t_{ij}\}$
- Match graph construction: $\{M_{ij}, R_{ij}, t_{ij}\}$ → camera poses $\{P_i\}$, tracks $\{p_k\}$
  - Graph initialization (select a robust initial match pair to build a metric reconstruction)
  - How we add edges to the match graph (global / incremental)
- Bundle adjustment: $\{P_i\}$, $\{p_k\}$ → optimized $\{P_i\}$, $\{p_k\}$
- *Building Rome in a day* (2009) – the first practical large-scale SfM system

SfM is just a large-scale optimization problem

- 2-view/3-view optimization (epipolar geometry)
- Match graph optimization
- Pose averaging
- Bundle adjustment (non-linear least squares)

Topic I: Local Features and Matching

- Local feature - the basis for SfM
- Scale Invariant Feature Transform (SIFT)
  - Scale-space extrema detection
  - Keypoint localization
  - Orientation assignment
  - Keypoint description
- Invariant to translation, scaling and rotation

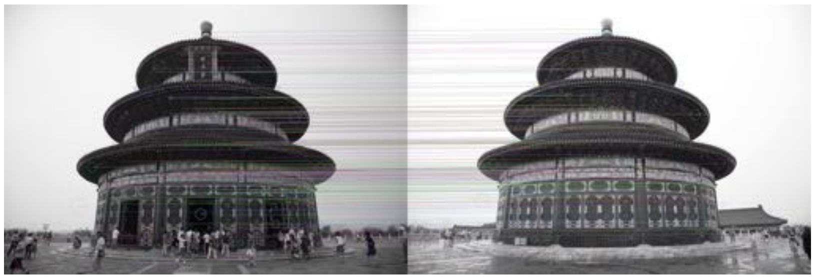


THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Problems with feature matching

- Tradeoff: SIFT is not invariant under geometric transformations
- Problem1: Pairwise feature matching is costly.
- Problem2: Erroneous matches is evitable, thus robust estimation is used.
  - An extreme case:

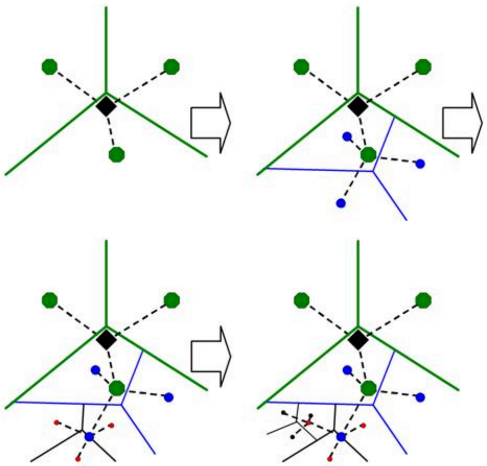
Front-front match



Erroneous front-back match

To tackle problem 1: matching efficiency

- Use image retrieval to compute a candidate match set
- Vocabulary tree: train -> build -> match
- Reduce cost from $O(n^2)$ to $O(kn)$, k decided by users
- Main problem with this approach:
  - k is not known beforehand
  - Too small k is not sufficient
  - Too large k slows down the process



[1] Nister, David, and Henrik Stewenius. "Scalable recognition with a vocabulary tree." *CVPR.* Vol. 2. IEEE, 2006.

To tackle problem 1: matching efficiency

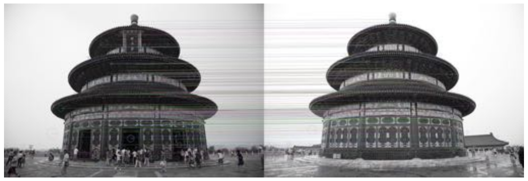
- Other approaches:
  - Relevance feedback and entropy minimization ([1] Lou et al.)
  - Match features in larger pyramid scale ([2] Wu)
  - Learning-based method to predict overlaps ([3] Scho¨nberger et al.)
  - A hashing-based cascading matching ([4] Cheng et al.)

[1] Lou, Y., Snavely, N., Gehrke, J.: Matchminer: Efficient spanning structure mining in large image collections. In: ECCV, pp. 45–58 (2012)
[2] Wu,C.:Towardslinear-timeincrementalstructurefrommotion.In:3DV.pp.127–134(2013)
[3] Scho¨nberger, J.L., Berg, A.C., Frahm, J.M.: Paige: Pairwise image geometry encoding for improved efficiency in structure-from-motion. In: CVPR. pp. 1009–1018 (2015)
[4] Cheng, Jian, et al. "Fast and accurate image matching with cascade hashing for 3d reconstruction." *CVPR.* 2014.

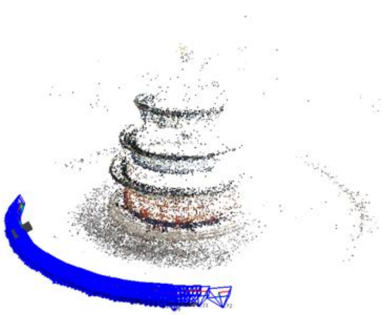
To tackle problem 2: erroneous matches

- Identification and removal of erroneous epipolar geometry is a recent research focus for SfM.
- Can lead to catastrophic results for SfM.

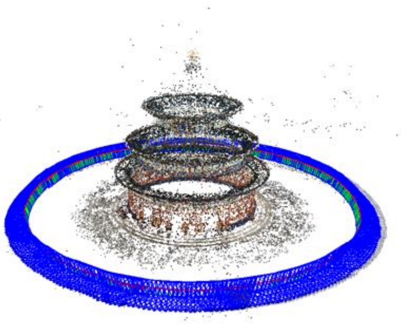
Front-front match



Erroneous front-back match



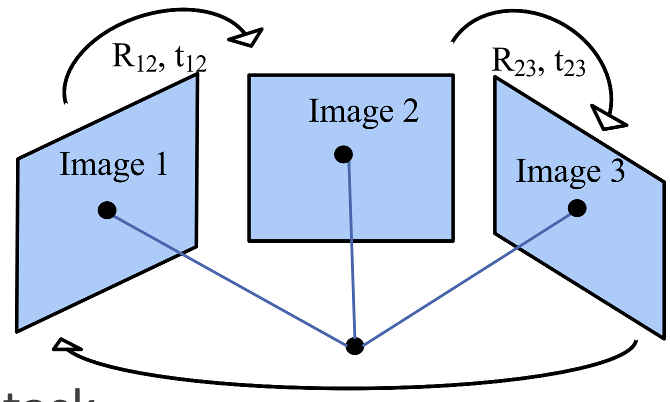
Wrong model



Correct model

To tackle problem 2: erroneous matches

- Loop consistency [1]:
  - Chained relative motion should be an identity map: $R_{12}R_{23}R_{31} = I$
  - Start from a full match graph
  - Sample cycles from the full graph
  - The problem is casted as a Bayesian inference task
  - Strong assumption on variable independence



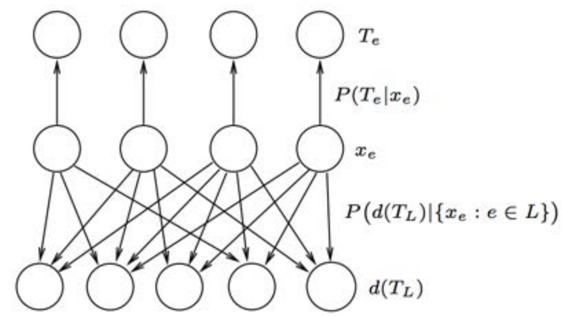


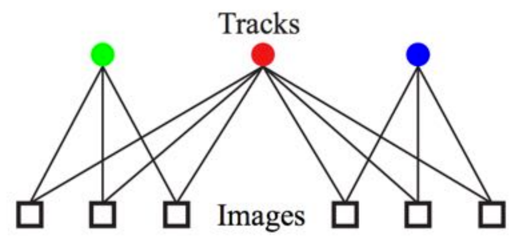
Figure 3. The Bayesian network for cycle inference.

[1] Zach, Christopher, Manfred Klopschitz, and Manfred Pollefeys. "Disambiguating visual relations using loop constraints." *CVPR*. Vol. 2. 2010.

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

To tackle problem 2: erroneous matches

- Other works:
  - Sampling match graph based on missing correspondences and time stamp cue. [1]
  - Analysis of visibility graph. [2]



  - Splits the camera graph and then leverages conflicting observations. [3]

[1] Roberts, Richard, et al. "Structure from motion for scenes with large duplicate structures." *CVPR,* IEEE, 2011.
[2] Wilson, Kyle, and Noah Snavely. "Network principles for sfm: Disambiguating repeated structures with local context." CVPR, 2013.
[3] J. Heinly, E. Dunn, and J.-M. Frahm, "Correcting for duplicate scene structure in sparse 3d reconstruction," in *ECCV*, pp. 780–795, 2014.

Motivation: Solve two problems together

- All disambiguation methods start from a relatively full match graph

- Construct an error-free match graph in a bottom-up fashion

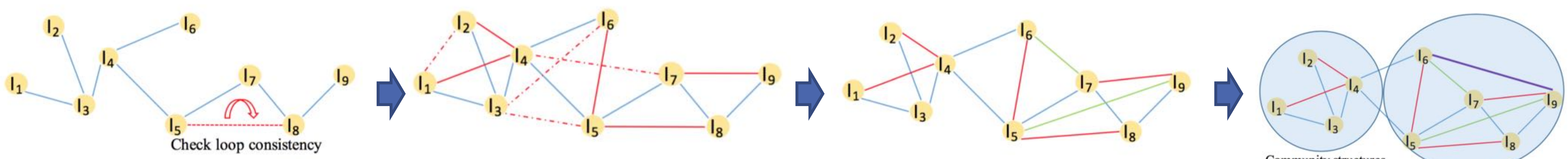
- Select a sufficient match set that can guarantee a reconstruction

- Prevent additions of erroneous pairs

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Graph-based consistent matching for SfM

- Multi-stage matching process:
  - Stage 1: Starts from a minimal spanning tree based on vocabulary tree ranks
  - Stage 2: Expand the spanning tree with loop consistency guaranteed
  - Stage 3: Find loop closures by community detection



(a) MST (b) First-order strong triplets (c) Second-order strong triplets (d) Third-order strong triplets

T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, "Graph-based consistent matching for structure-from-motion," in *ECCV*, 2016.

Graph-based consistent matching for SfM

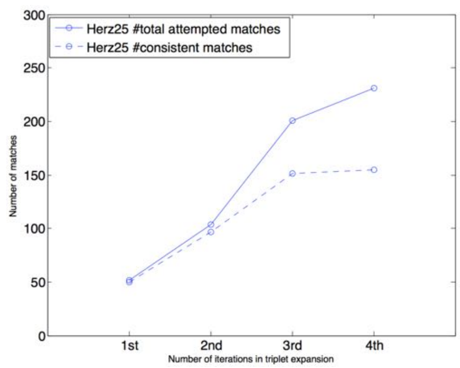
- Stage 1: start from a minimal spanning tree
  - The purpose is to quickly chain the views
  - A modified Kruskal's algorithm (online version): reject outliers
  - Edge weight paran $w(e_{ij}) = \sqrt{\frac{Rank_i^2(j)+Rank_j^2(i)}{2}}$ ation given by vocabulary tree:

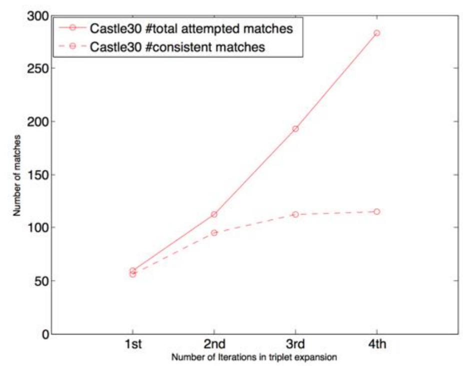
T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, "Graph-based consistent matching for structure-from-motion," in *ECCV*, 2016.

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Graph-based consistent matching for SfM

- Stage 2: Graph Expansion by Strong Triplets
  - Verifying all loops is hard to achieve, even verifying all triplets is $O(n^3)$
  - Generate a consistent match graph in a bottom-up way
  - A empirical choice: traversing two steps starting from each node



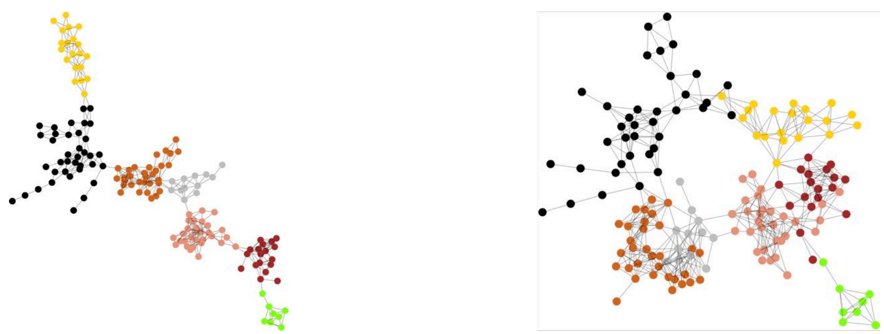


T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, “Graph-based consistent matching for structure-from-motion,” in *ECCV*, 2016.

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Graph-based consistent matching for SfM

- Stage 3:Community-Based Graph Reinforcement
  - Too sparse connection after triplet expansion
  - Longer loops are not verified
  - Community detection: divide a graph into groups with denser connections inside and sparser connections outside.

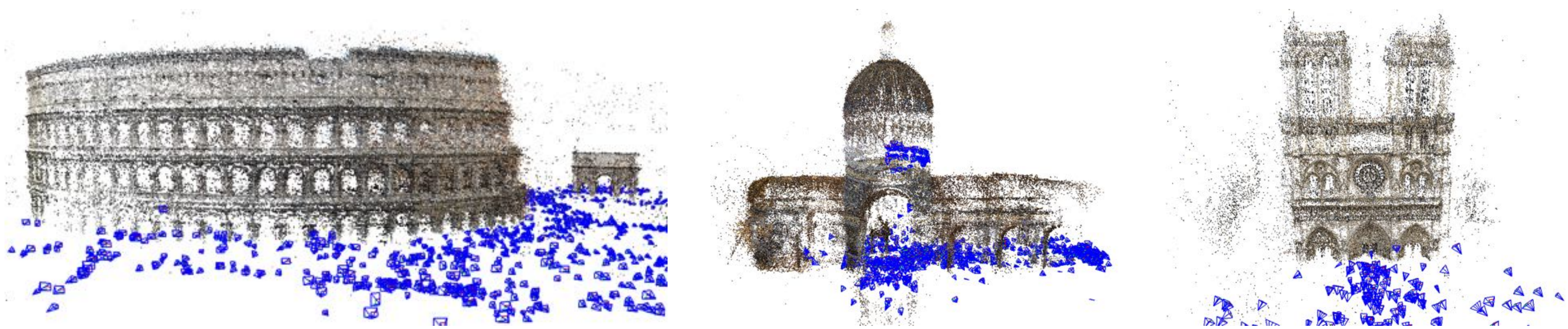


T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, "Graph-based consistent matching for structure-from-motion," in *ECCV*, 2016.

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Graph-based consistent matching for SfM

- Results – Internet data



T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, "Graph-based consistent matching for structure-from-motion," in *ECCV*, 2016.

Graph-based consistent matching for SfM

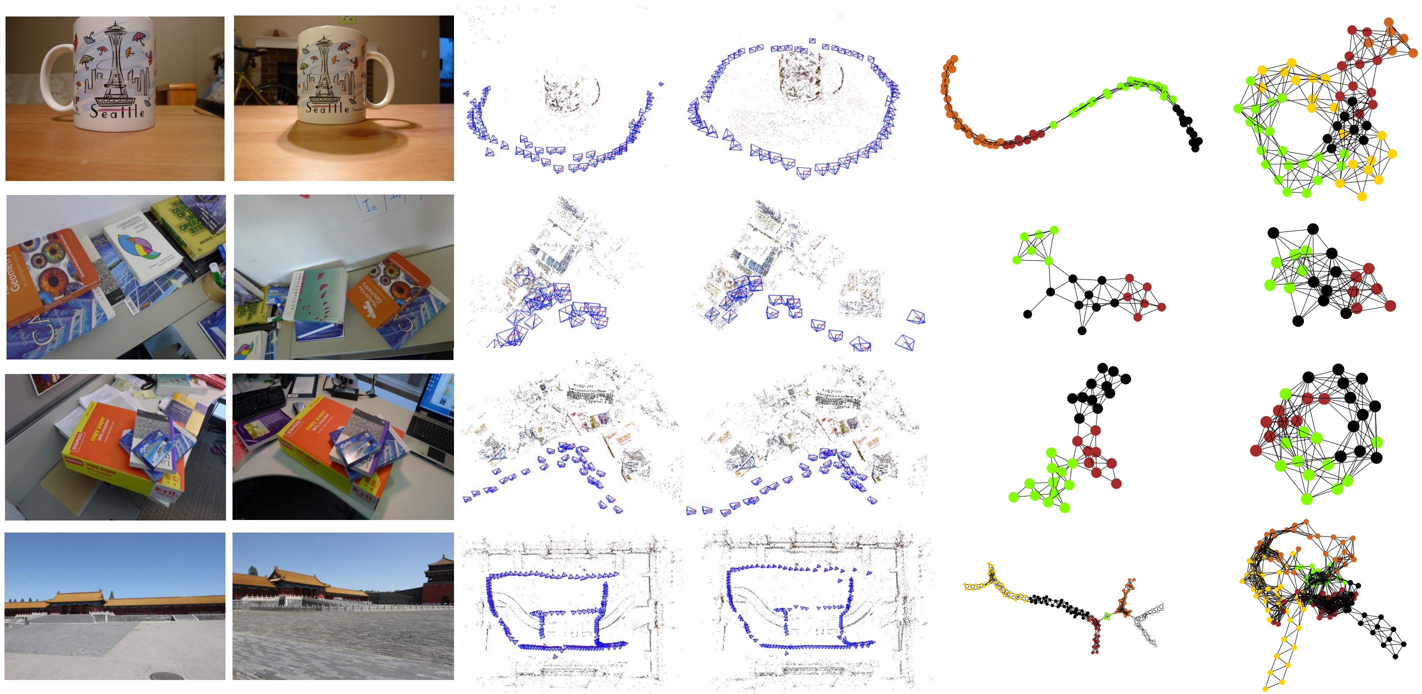
Results – ambiguity data

(a)

(b)

(c)

(d)



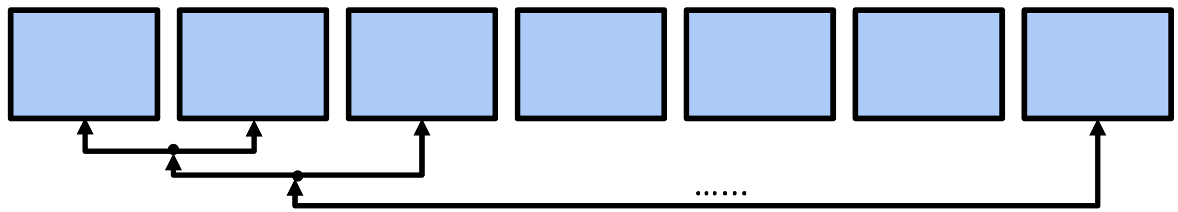
T. Shen, S. Zhu, T. Fang, R. Zhang, and L. Quan, “Graph-based consistent matching for structure-from-motion,” in *ECCV*, 2016.

Future direction: learning local features

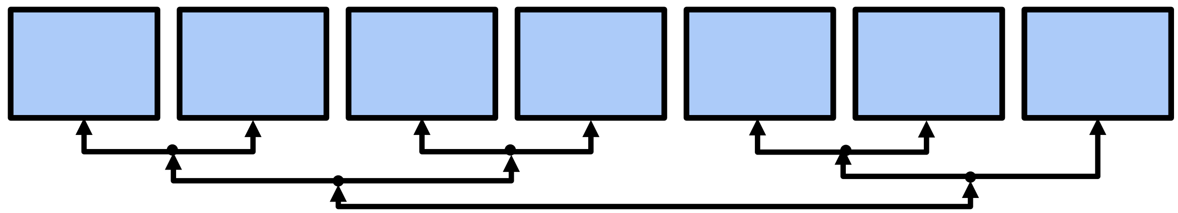
- Feature is the most important factor in SfM accuracy
- Deep learning approaches: learning local feature descriptors
- Speed up matching and improve matching accuracy

Topic II: From Feature matches to 3D

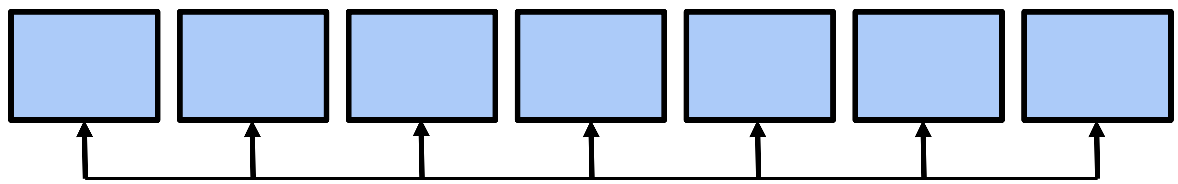
- Incremental



- Hierarchical



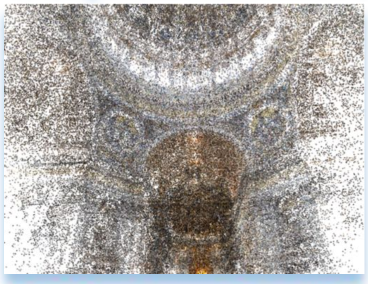
- Global



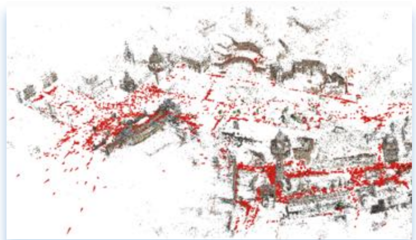
THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Some Recent Representative Architectures

Sequential/Incremental Approaches



Building Rome in a day

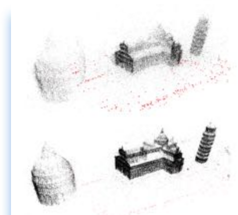


Colmap: SfM Revisited

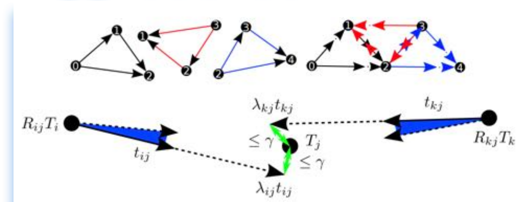
Hierarchical

Randomized structure from motion based on atomic 3d models from camera triplets

Global Approaches



Optimizing the Viewing Graph for Structure-from-Motion



Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion.

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Three SfM Paradigms

	Incremental	Hierarchical	Global
Feature extraction and matching	-	-	-
Match graph initialization	Initialized by carefully selected two-view	Atomic models	All views are treated equally
Image Registration	Perspective-n-Point (PnP), 2D-3D correspondences	3D-3D fusion	Rotation and translation averaging
Bundle adjustment	Iterative, many times	BA when merging	one time
Advantages	Robust	Fewer BA steps	Evenly-distributed errors
Disadvantages	Prune to drifting errors	Model merging, graph partition	Prune to noisy pairwise matches
Softwares	Bundler, openMVG, VisualSfM, MVE	Research papers	openMVG, Theia

Key technique: motion averaging

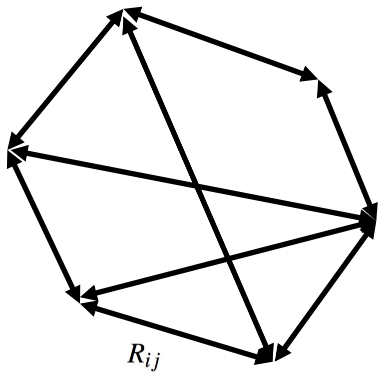
- Correct accumulating errors in chained pose estimation
- First rotation averaging, then translation averaging

Images $\{I_i\}$ → relative motions $\{R_{ij}, t_{ij}\}$ → rotation averaging $\{R_i\}$ → translation averaging $\{T_i\}$ → 3D structures $\{X_k\}$

Convex optimization in SfM

- Convex optimization becomes popular because of its elegant mathematical forms and the existence of global minimum
- First investigated by Hartley et al. on triangulation
- Quasi-convex optimization by Ke et al. and Kahl, casted as an Second-Order Cone Programming (SOCP)
- Not practical due to its sensitivity to noises, but theoretically interesting

Rotation averaging on a graph



R_{ij}

$t_{ij} = (T_j - T_i)/||T_j - T_i||$

Viewing Graph: $G = (\mathscr{V}, \mathscr{E})$

Globally consistent rotation: $R_{ij} = R_j R_i^{-1}, \quad \forall (i,j) \in \mathscr{E}$

Minimize Riemannian distance: $d(\mathbf{X}, \mathbf{Y}) = ||\log(\mathbf{Y}\mathbf{X}^{-1})||$

Rotation average is non-convex

V. M. Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *CVPR*, vol. 1, pp. I–684, IEEE, 2004.

Rotation averaging: other approaches

- Quaternions parameterization (Martinec et al. [1])

- L1 norm based on Weiszfeld algorithm (Hartley et al. [2])

[1] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *CVPR*, pp. 1–8, 2007.
[2] R. Hartley, K. Aftab, and J. Trumpf, "L1 rotation averaging using the weiszfeld algorithm," in *CVPR*, pp. 3041–3048, IEEE, 2011.

Translation averaging

- Long been characterized as a convex optimization problem
- Min-max formualation, SOCP
- Same L-infinity drawbacks: prune to outliers

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Translation averaging

- Considering observed points together (triplet bundle)

- Re-projection error: $\rho(t_i, X_j) = \left\| \left(\hat{x_{ij}}^{(1)} - \frac{R_i^{(1)T} X_j + t_i^{(1)}}{R_i^{(3)T} X_j + t_i^{(3)}}, \hat{x_{ij}}^{(2)} - \frac{R_i^{(2)T} X_j + t_i^{(2)}}{R_i^{(3)T} X_j + t_i^{(3)}} \right) \right\|_\infty$

- Linear program minimal case with RANSAC

$$\begin{aligned} \underset{\{t_i\},\{X_j\},\gamma}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \rho(t_i, X_j) \leq \gamma, \\ & R_i^{(3)} X_j + t_i^{(3)} \geq 1, \\ & t_i = (0,0,0) \; \forall i, j. \end{aligned}$$

P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *ICCV*, pp. 3248–3255, 2013

Translation averaging

- Then global translation averaging

- Formulation under L-infinity:

$$\begin{aligned} &\underset{\{T_i\},\{\lambda_{ij}\},\gamma}{\text{minimize}} && \gamma \\ &\text{subject to} && \left\|T_j - R_{ij}T_i - \lambda_{ij}t_{ij}\right\|_\infty \leq \gamma, \\ &&& \lambda_{ij} \geq 1, \ \forall i,j \\ &&& T_1 = (0,0,0). \end{aligned}$$

- Minimizing two-side of $\lambda_{ij}t_{ij} = T_j - R_{ij}T_i$

P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *ICCV*, pp. 3248–3255, 2013

Translation averaging: robust formulation

- A small robust L1 formulation improvement: consider L1 norm of the re-projection error vector:

$$(\cdots, \rho(t_i, X_j), \cdots)$$

$$\begin{aligned} \underset{\{t_i\},\{X_j\},\gamma}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \rho(t_i, X_j) \le \gamma, \\ & R_i^{(3)} X_j + t_i^{(3)} \ge 1, \\ & t_i = (0,0,0)\ \forall i, j. \end{aligned} \quad \Rightarrow \quad \begin{aligned} \underset{\{t_i\},\{X_j\},\{\gamma_i\}}{\text{minimize}} \quad & \sum_i \gamma_i \\ \text{subject to} \quad & \rho(t_i, X_j) \le \gamma_i, \\ & R_i^{(3)} X_j + t_i^{(3)} \ge 1, \\ & t_i = (0,0,0)\ \forall i, j. \end{aligned}$$

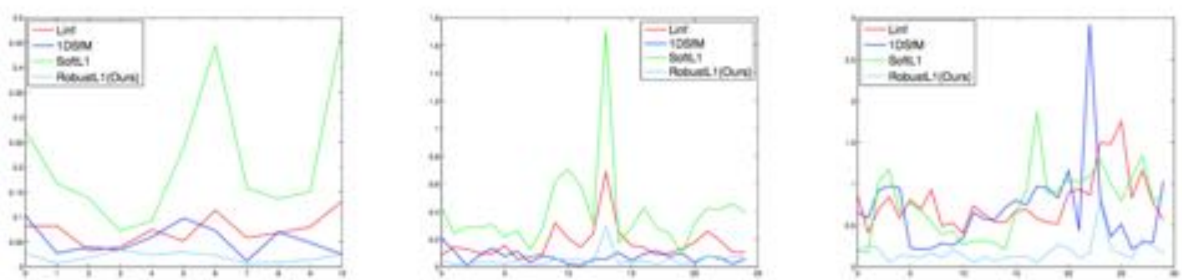
$$\begin{aligned} \underset{\{T_i\},\{\lambda_{ij}\},\gamma}{\text{minimize}} \quad & \gamma \\ \text{subject to} \quad & \left\| T_j - R_{ij} T_i - \lambda_{ij} t_{ij} \right\|_\infty \le \gamma, \\ & \lambda_{ij} \ge 1,\ \forall i, j \\ & T_1 = (0,0,0). \end{aligned} \quad \Rightarrow \quad \begin{aligned} \underset{\{T_i\},\{\lambda_{ij}\},\{\gamma_{(i,j)}\}}{\text{minimize}} \quad & \sum_{(i,j)} \gamma_{(i,j)} \\ \text{subject to} \quad & \left\| T_j - R_{ij} T_i - \lambda_{ij} t_{ij} \right\|_\infty \le \gamma_{(i,j)}, \\ & \lambda_{ij} \ge 1,\ \forall i, j \\ & T_1 = (0,0,0). \end{aligned}$$

Tianwei Shen. Convex Modelling of Motion Estimation in Structure-from-Motion. (unpublished report)

Translation averaging: comparison

- RobustL1 outperforms other methods

(a) fountain-P11 dataset (b) Herz-Jesu-P25 dataset (c) castle-P30 dataset



(d) fountain-P11 per camera error (e) Herz-Jesu-P25 per camera error (f) castle-P30 per camera error

Translation averaging: comparison

- However, advantage is not evident after bundle adjustment (BA)
- Also, the problem scale is larger
- Future direction: no large-scale benchmark datasets for testing
- A potential useful settings is SLAM, where BA is costly

Topic 3: Bundle adjustment

- Joint optimization of camera poses and 3D tracks

$$\min_{P_i \in \mathscr{Q}} \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij} f(u_{ij} - \Pi(P_i, X_j))$$

- Error model: $f(\Delta z_{ij}) = \frac{1}{2} \Delta {z_{ij}}^T W_{ij} \Delta z_{ij}$

$$\Delta z_{ij} = u_{ij} - \Pi(P_i, X_j)$$

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING 計算機科學及工程學系

Bundle adjustment

- Levenberg-Marquardt algorithm
  - Taylor expansion: $f(x+\delta x) \approx f(x) + g^T \delta x + \frac{1}{2}\delta x^T H \delta x, g \equiv \frac{df}{dx}(x), H \equiv \frac{d^2 f}{dx^2}(x)$
  - Newton step: $\frac{d}{dx} f(x+\delta x) \approx H\delta x + g = 0 \implies \delta x = -H^{-1} g$
  - New value: $f(x+\delta x) \approx f(x) - \frac{1}{2} g^T H^{-1} g$
  - Damped Newton's methods: $(H + \lambda W)\delta x = -g$

Bundle adjustment

- Large-scale endeavors
  - Multi-core bundle adjustment [1]
  - Distributed settings [2]
- Essentially a non-linear least square problem, thus generally useful for other vision problems.

[1] Wu, Changchang, et al. "Multicore bundle adjustment." *CVPR, 2011*.
[2] Eriksson, Anders, et al. "A Consensus-Based Framework for Distributed Bundle Adjustment." CVPR, 2016.

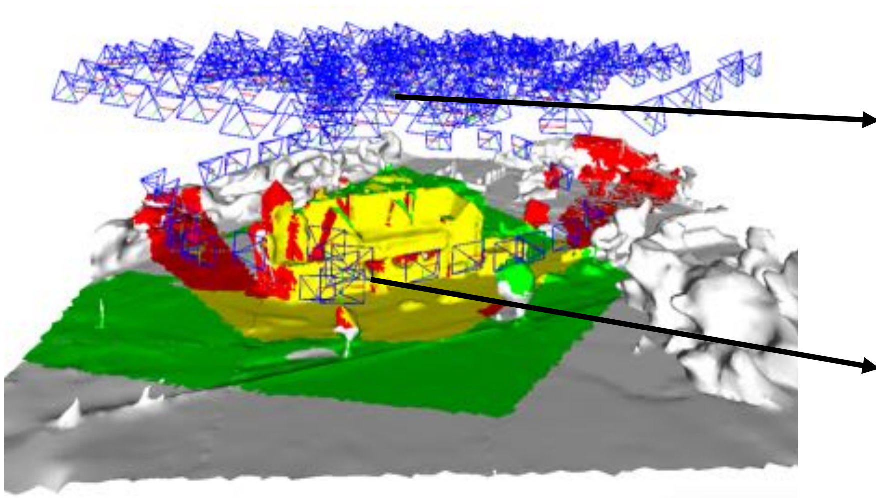
What can we do with SfM?

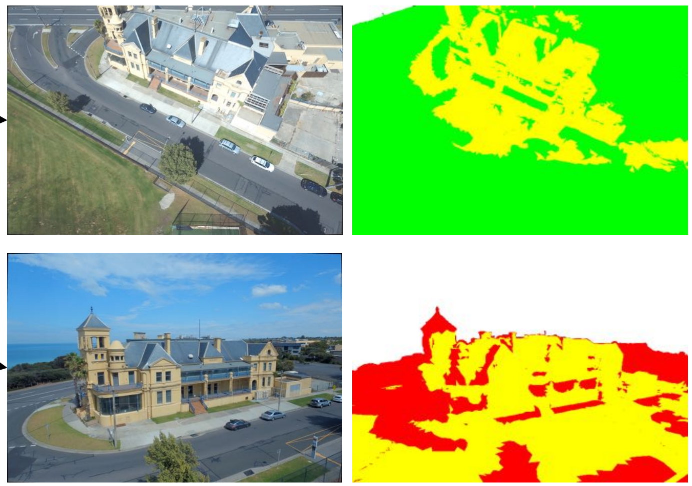
- 3D reconstruction
- Simultaneous localization and mapping (SLAM)
- Test base for local features (distinctiveness, efficiency, matchability)
- Color correction for image collections
- Visual effects
- …

Application: Large-Scale Color Correction

- Motivation: Images captured for 3D reconstruction are color-inconsistent

- Optimize color of image collections, based on geometric information





THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Tianwei Shen, Jinglu Wang, Tian Fang, Siyu Zhu, Long Quan. Color Correction for Image-Based Modelling in the Large. In ACCV 2016.

Application: Large-Scale Color Correction

- Non-linear optimization on color histogram:

$$\underset{\{s_i\},\{o_i\}}{\text{minimize}} \quad \sum_{i,j,k} \rho\left(\frac{(s_i Q_{ij}^{(k)} + o_i) - (s_j Q_{ji}^{(k)} + o_j)}{s_i + s_j}\right)^2$$

$$\text{subject to} \quad 1 - \delta_s \leq s_i \leq 1 + \delta_s, -\delta_o \leq o_i \leq \delta_o, \ \forall i.$$

$$\rho(x) = \delta^2(\sqrt{1 + (x/\delta)^2} - 1)$$

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

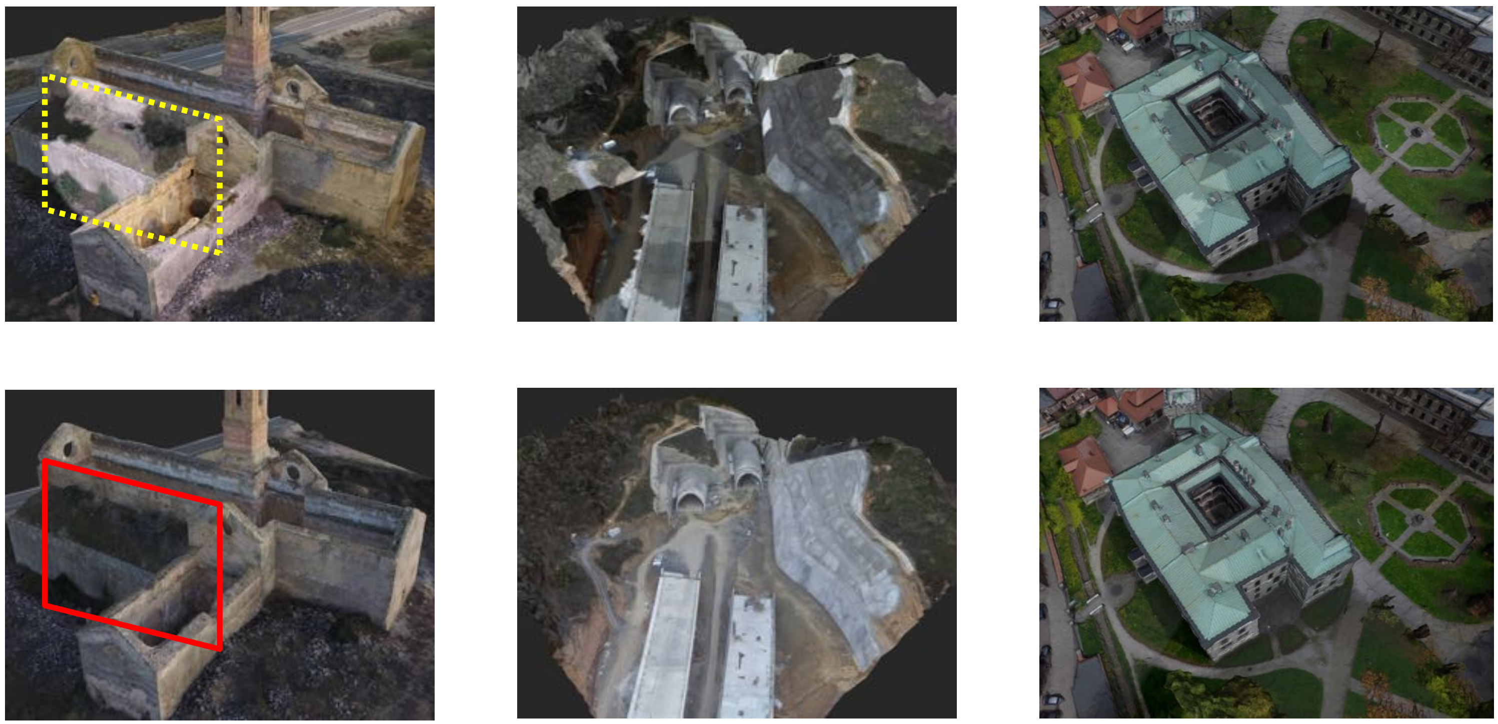
Tianwei Shen, Jinglu Wang, Tian Fang, Siyu Zhu, Long Quan. Color Correction for Image-Based Modelling in the Large. In ACCV 2016.

Application: Large-Scale Color Correction

- Consistent texturing:

Before

After



Tianwei Shen, Jinglu Wang, Tian Fang, Siyu Zhu, Long Quan. Color Correction for Image-Based Modelling in the Large. In ACCV 2016.

Final Remarks

- ❑Merge ground-level street-view images with aerial images
- ❑Better local invariant features and efficient matching
- ❑Distributed everything in SfM

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系