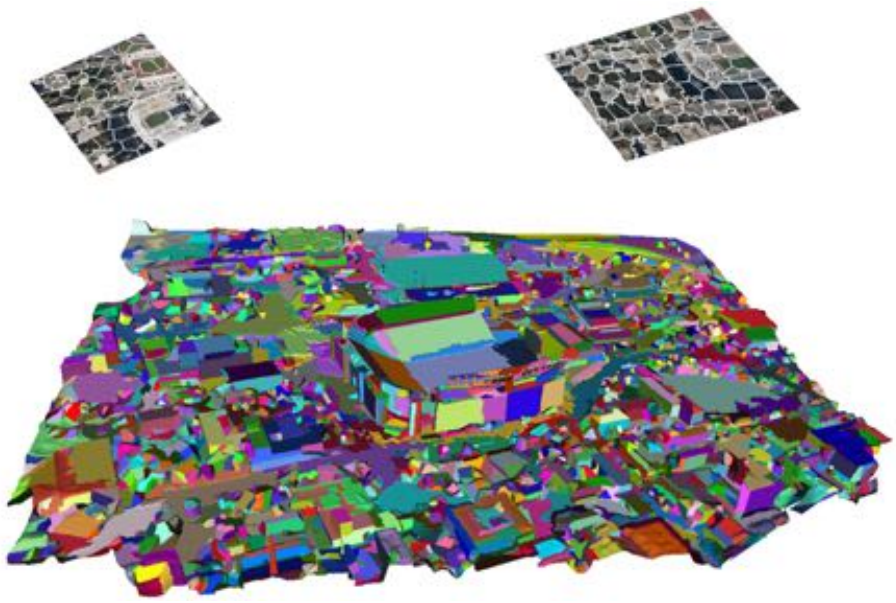
Part III

Jinglu Wang

Urban Scene Segmentation, Recognition and Remodeling

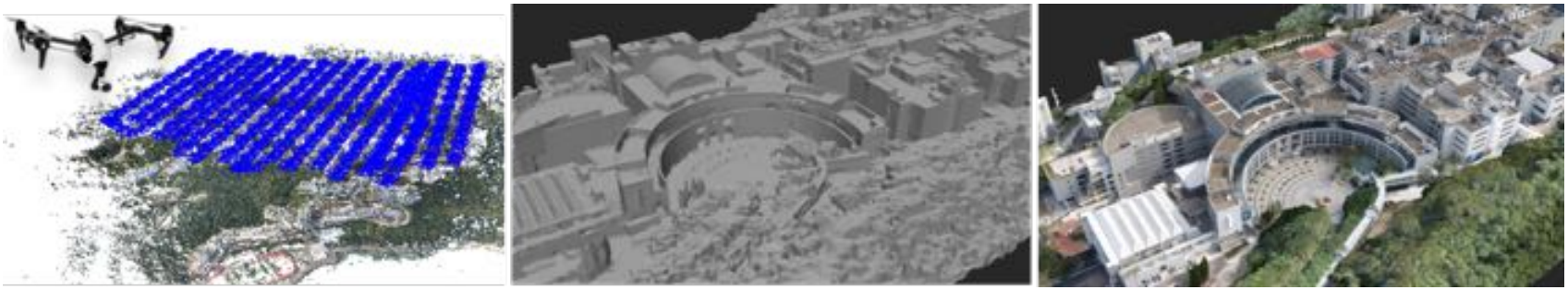


Outline

- Introduction
  - Motivation
  - Overview
- Related work
- Approaches
- Conclusion and future work

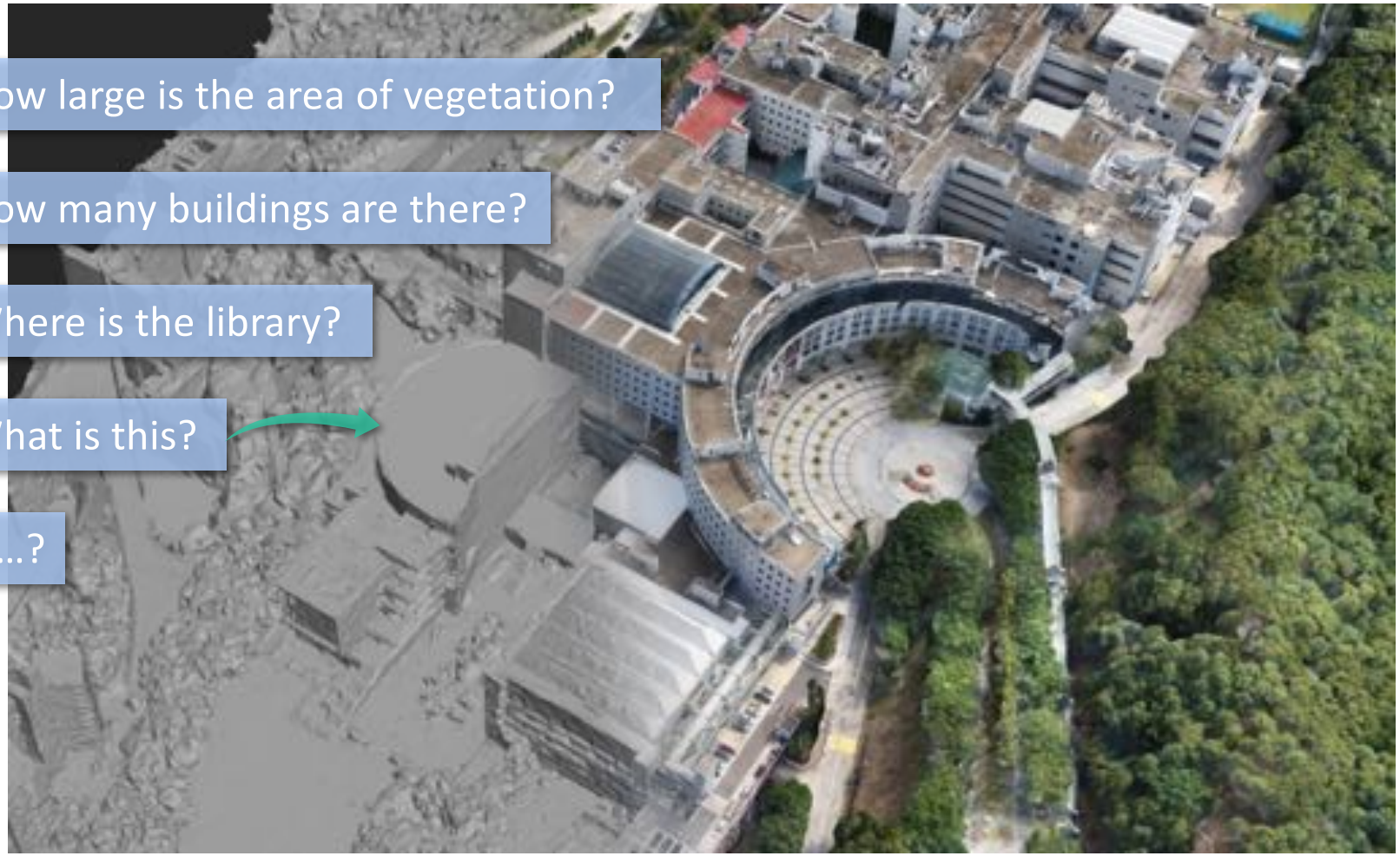
Motivation

- Large-scale multi-view reconstruction

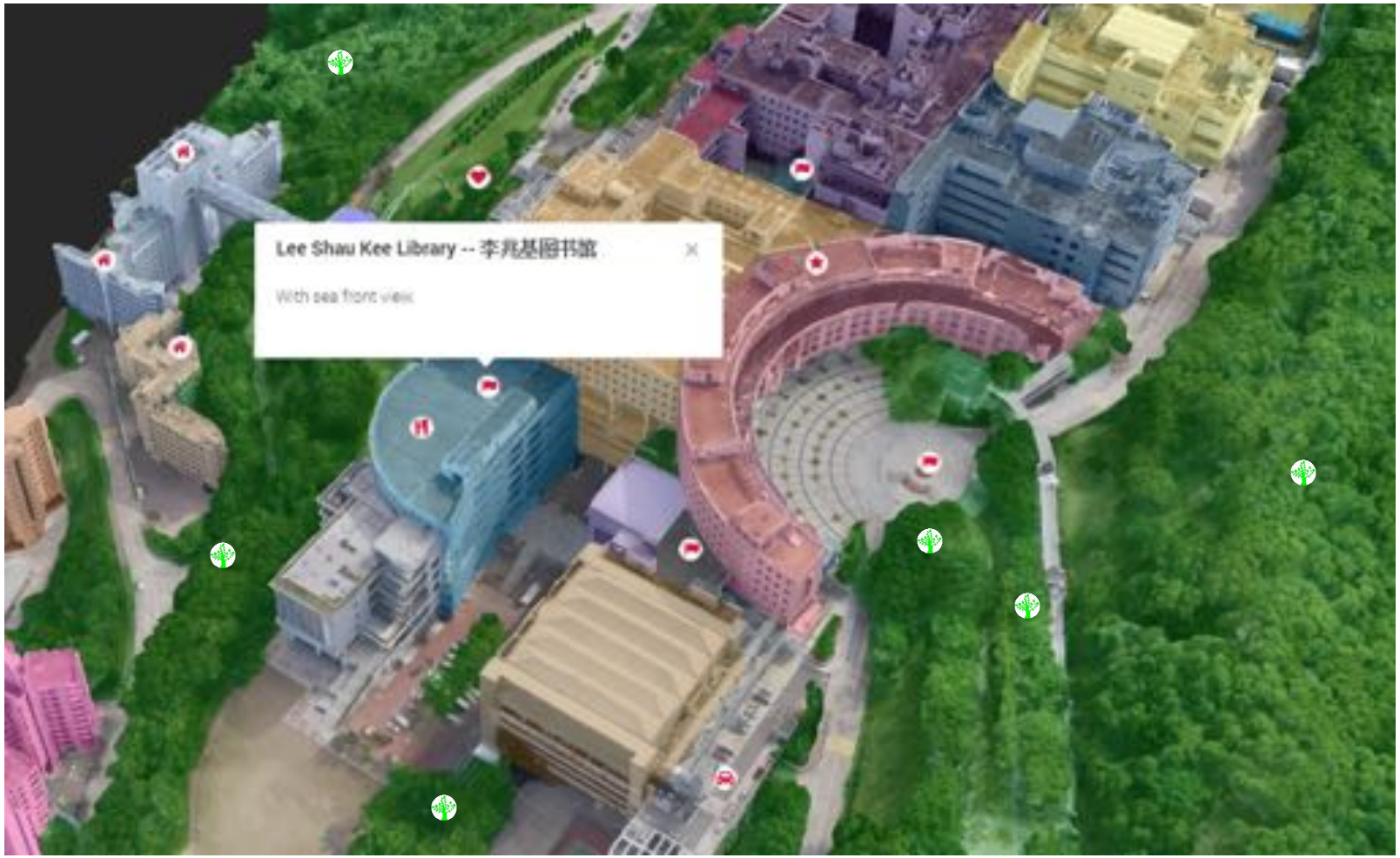


- 3D city models are available

Motivation


How large is the area of vegetation?
How many buildings are there?
Where is the library?
What is this?
……?

Motivation



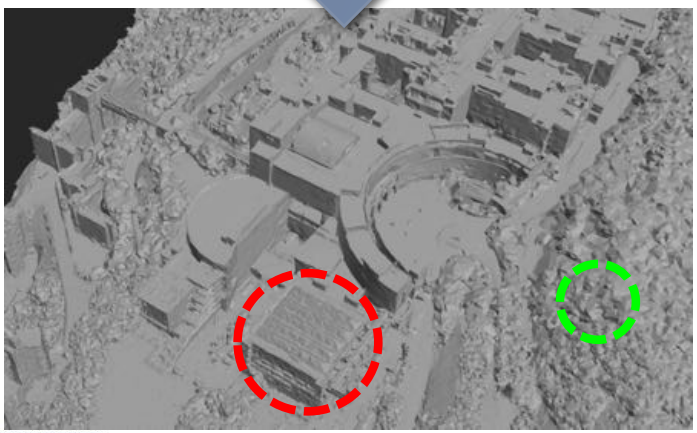
Semantic meaning of objects (concept model)

Motivation

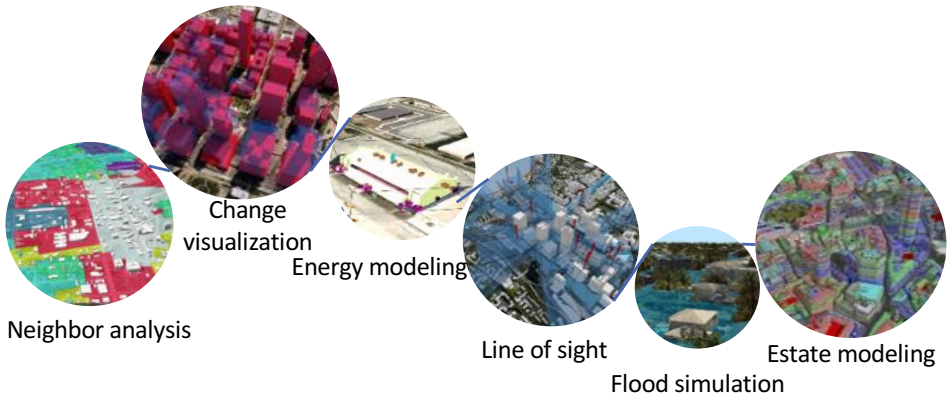
Semantic meaning

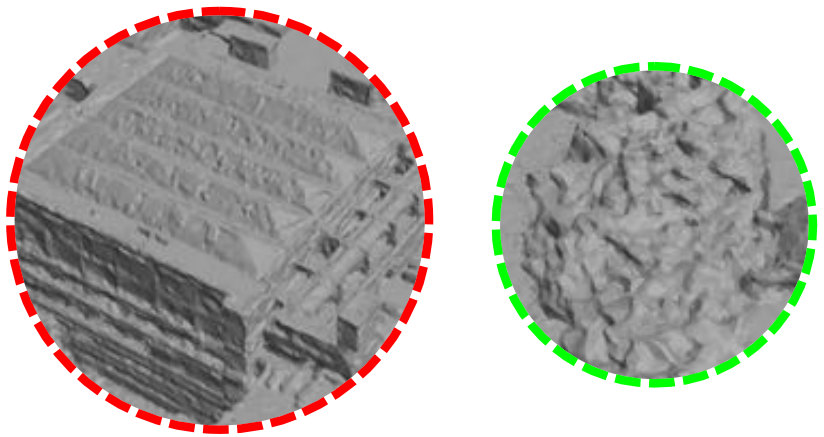
Refine

Geometry



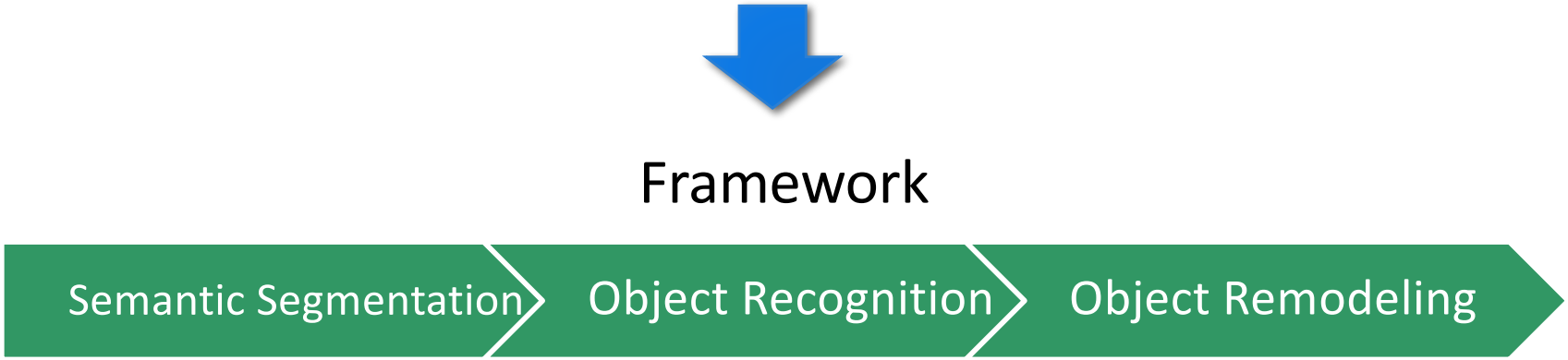
Practical applications



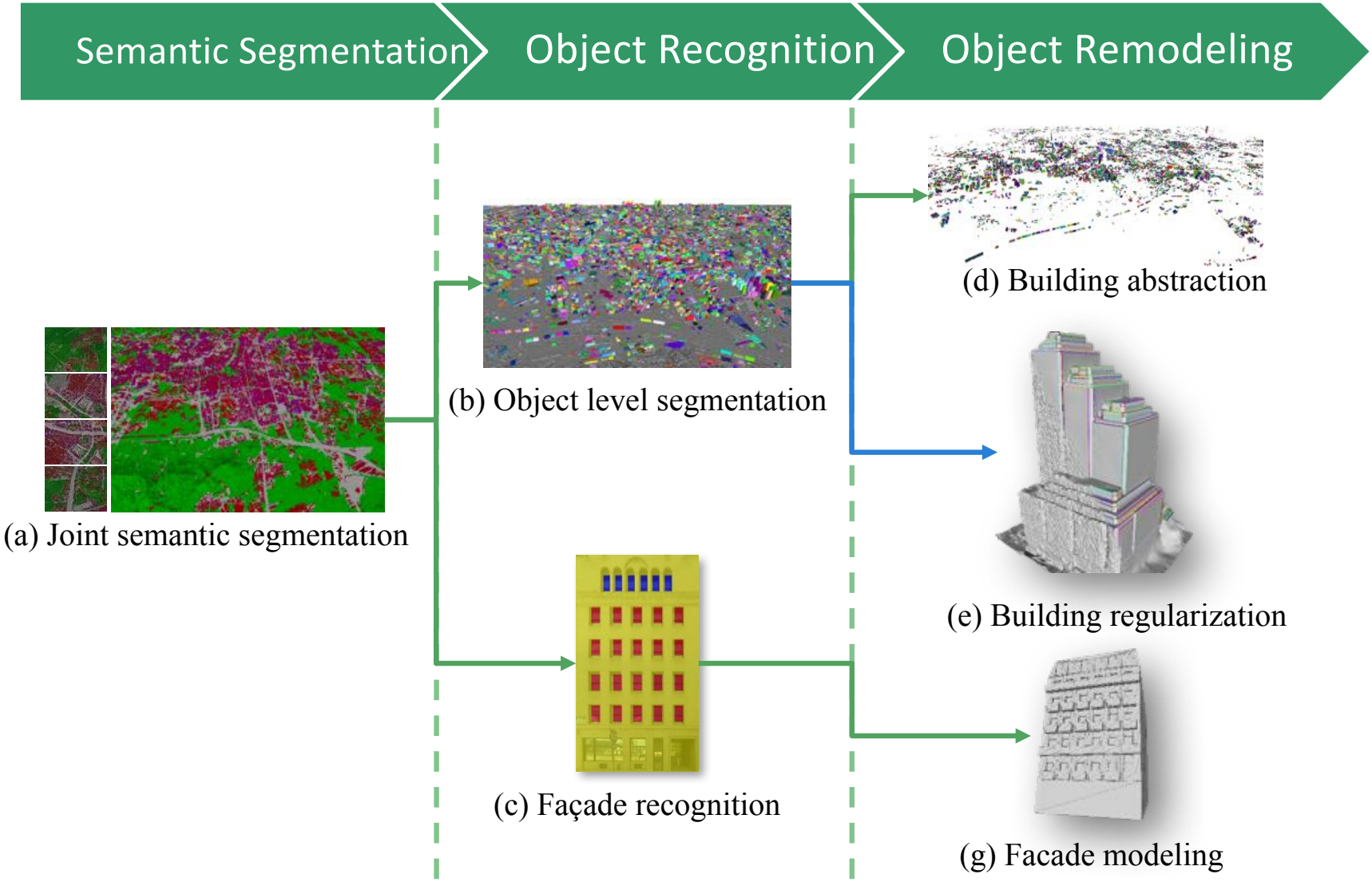


Findings

- Semantic meanings
- Object level
- Refine models



Framework


Semantic Segmentation
Object Recognition
Object Remodeling
(a) Joint semantic segmentation
(b) Object level segmentation
(c) Façade recognition
(d) Building abstraction
(e) Building regularization
(g) Facade modeling

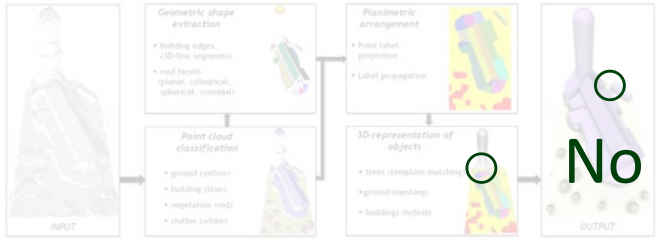
Outline

- Introduction
- Related work
- Approaches
- Conclusion

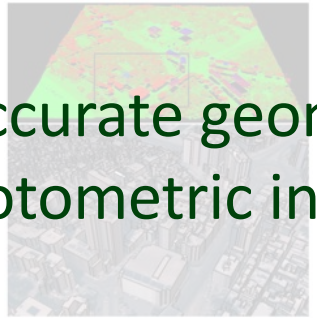
- Scanned data
- (Multi-view stereo) MVS data
-

Related Work

- With Laser scanned data
  - Accurate geometry
  - No photometric information



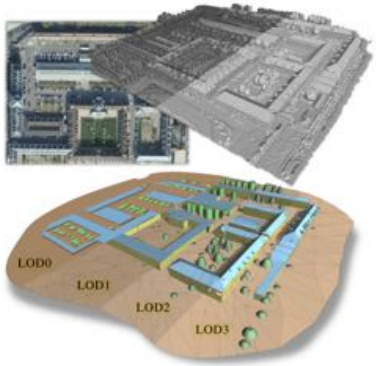
[Lafarge et al. 2012]



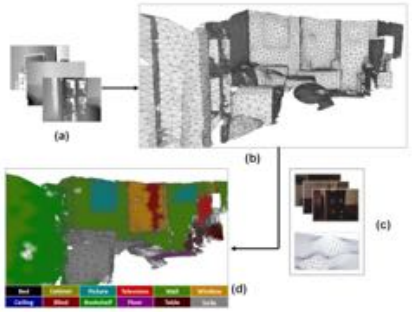
[Matei et al. 2008]

[Lin et al. 2013]

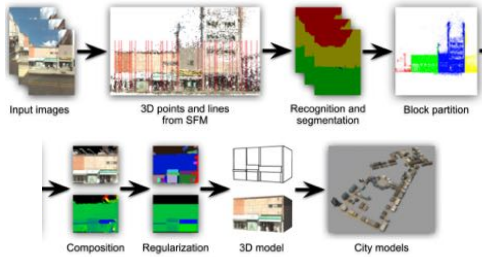
- With multi-view reconstructed data



[Yannick et al., 2015]



[Julien et al. 2013]

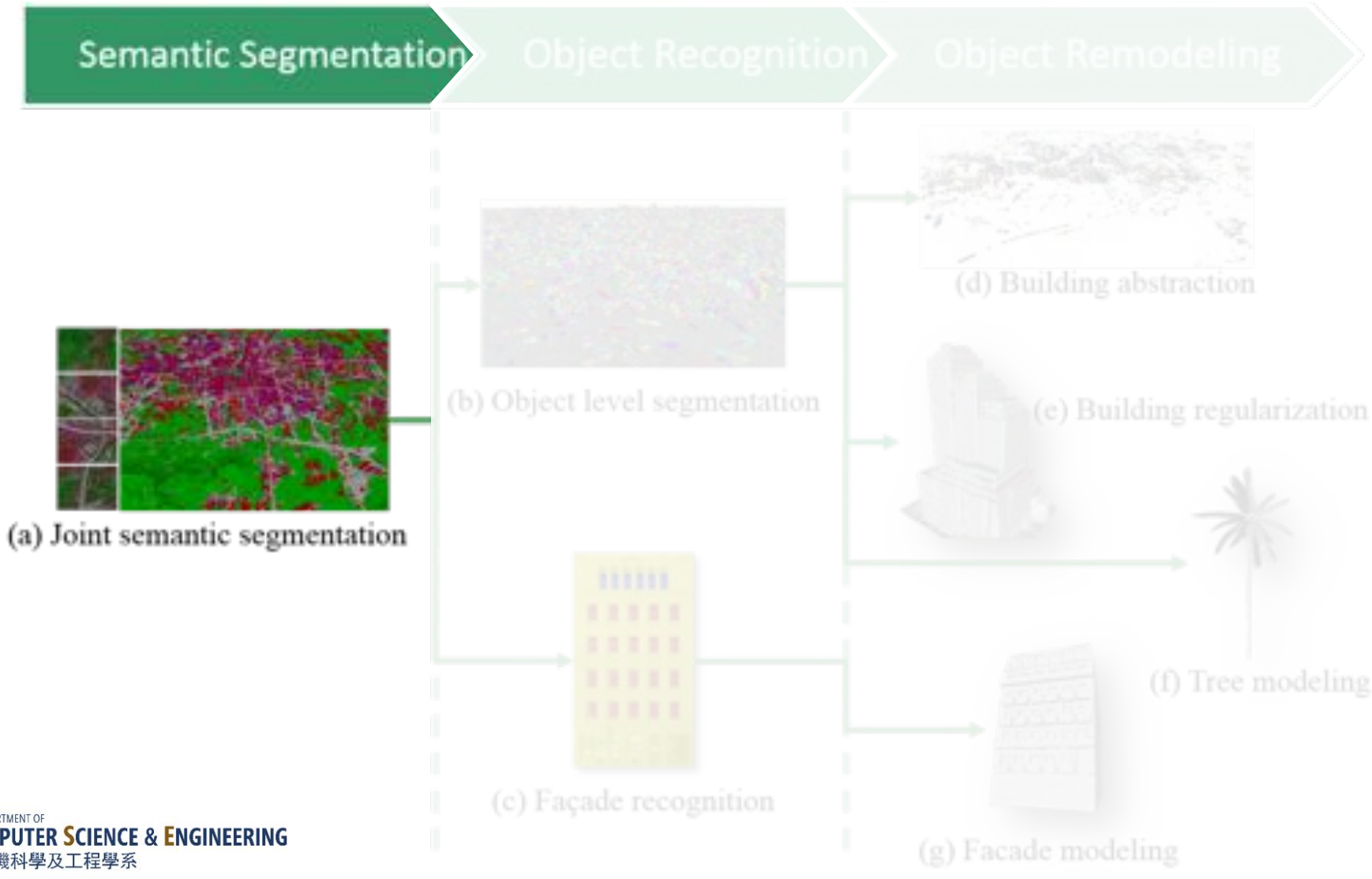


[Xiao et al. 2009]

Outline

- Introduction
- Related work
- Approaches
- Conclusion

Approaches

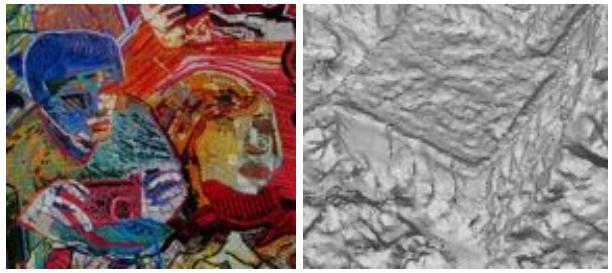

Semantic Segmentation
Object Recognition
Object Remodeling
(a) Joint semantic segmentation
(b) Object level segmentation
(c) Façade recognition
(d) Building abstraction
(e) Building regularization
(f) Tree modeling
(g) Facade modeling

Joint Semantic Segmentation

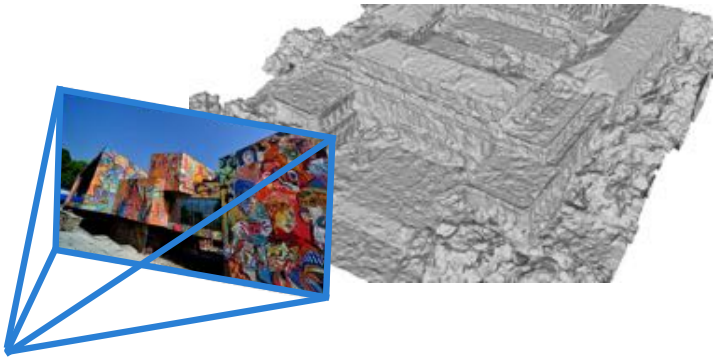
- Challenge
- Representation
- Workflow
- Result
- Evaluation

Challenges

- Joint

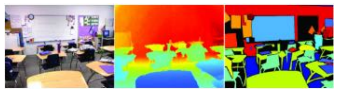


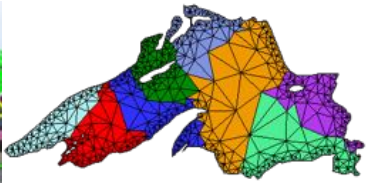
2D 3D



2D-3D consistency

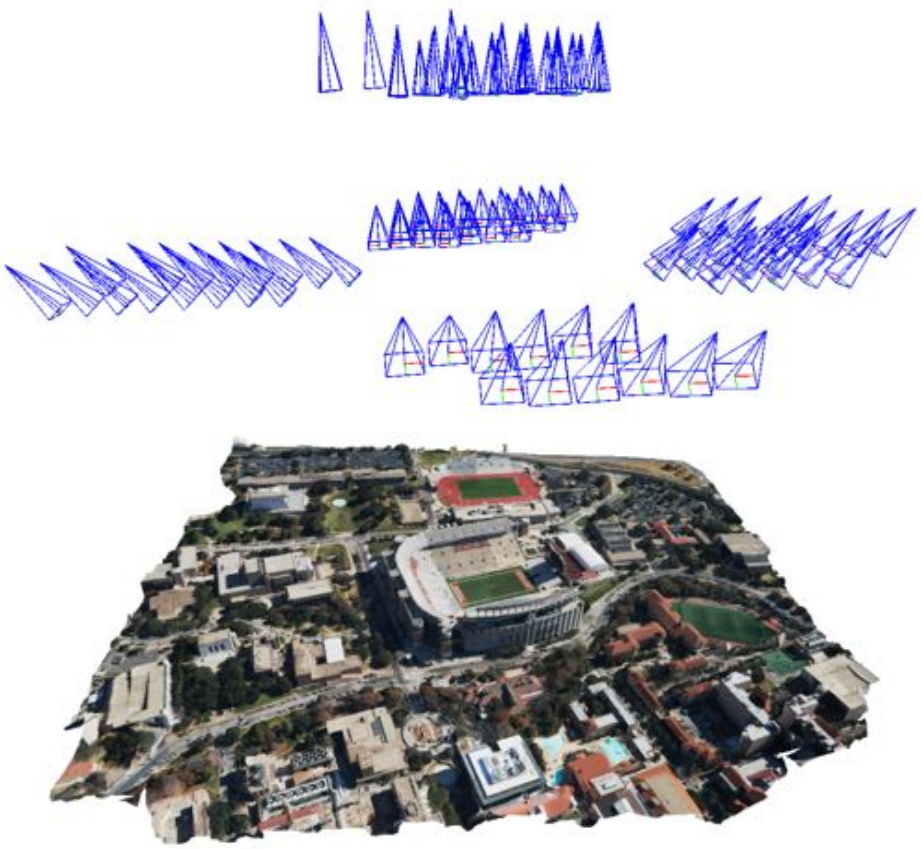
- Efficient labeling





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

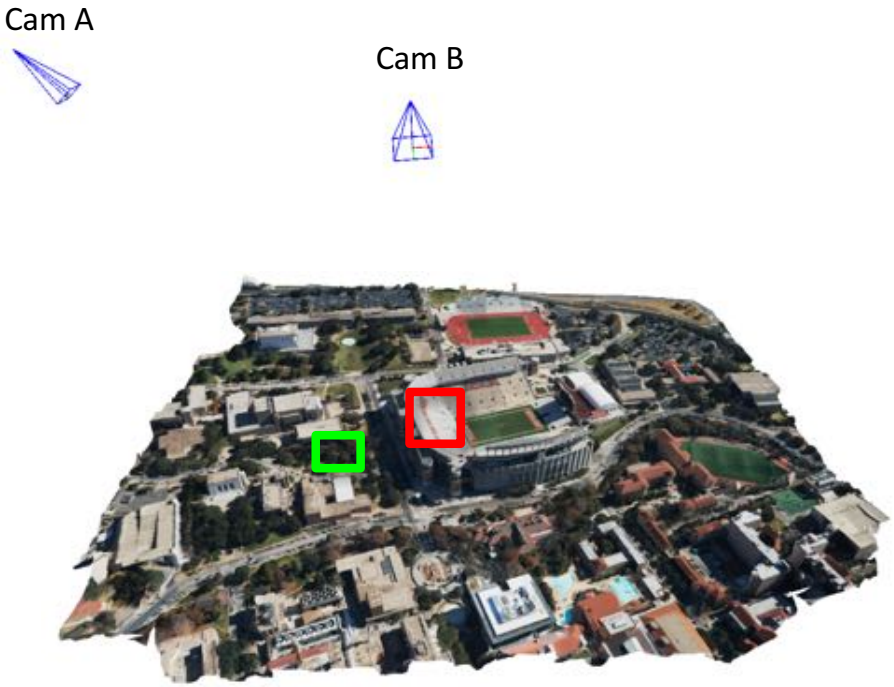
Input



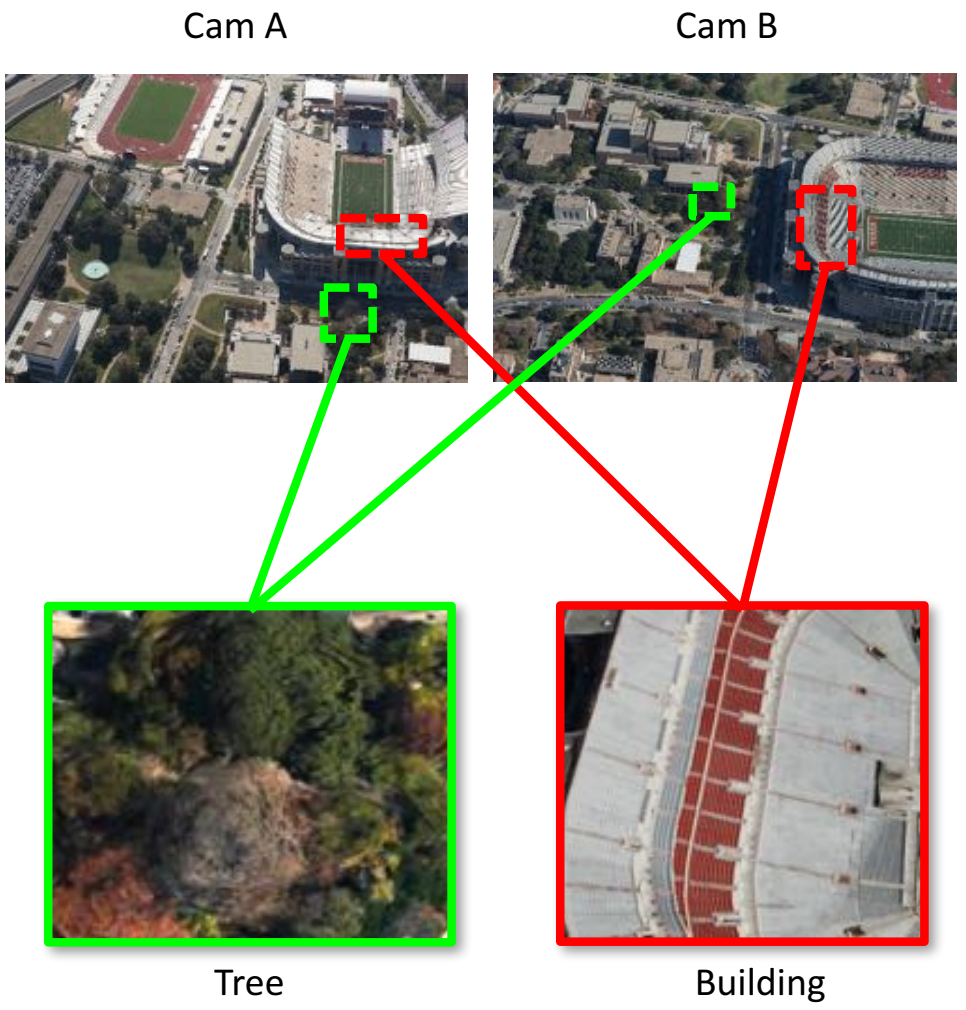
Multi-view reconstruction

Multi-view images

Label Consistency for Joint Segmentation



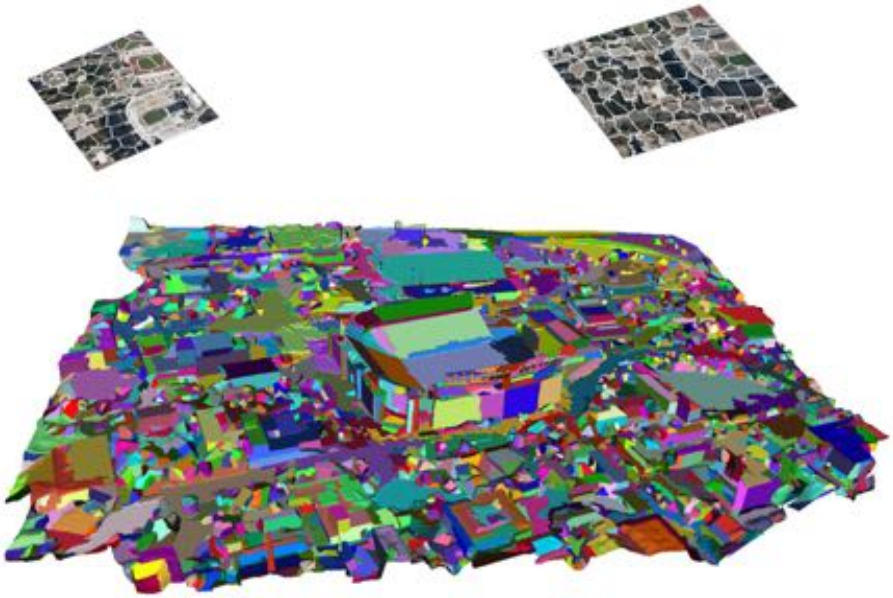
Multi-view reconstruction



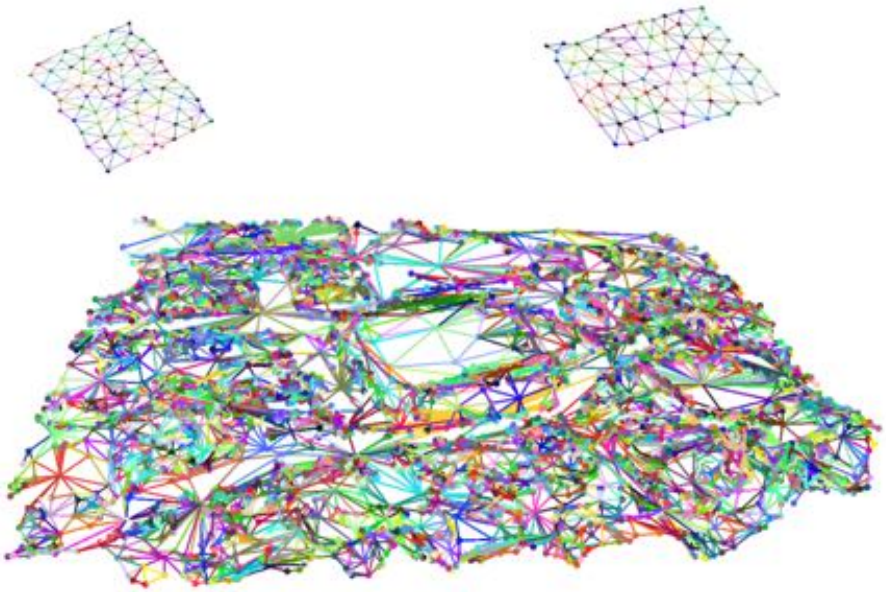
Multi-view label consistency

Correspondence Graph

Patch segmentation

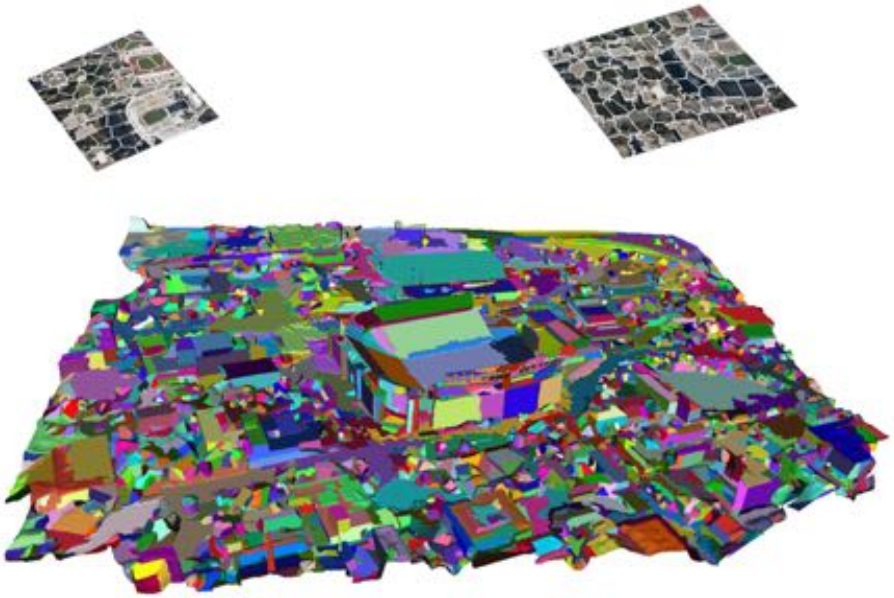


Graph visualization

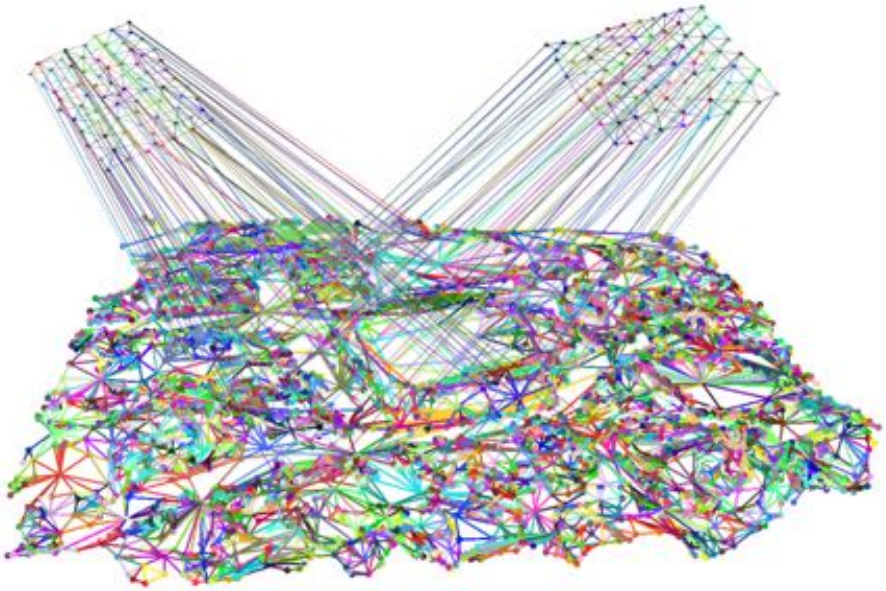


Correspondence Graph

Patch segmentation

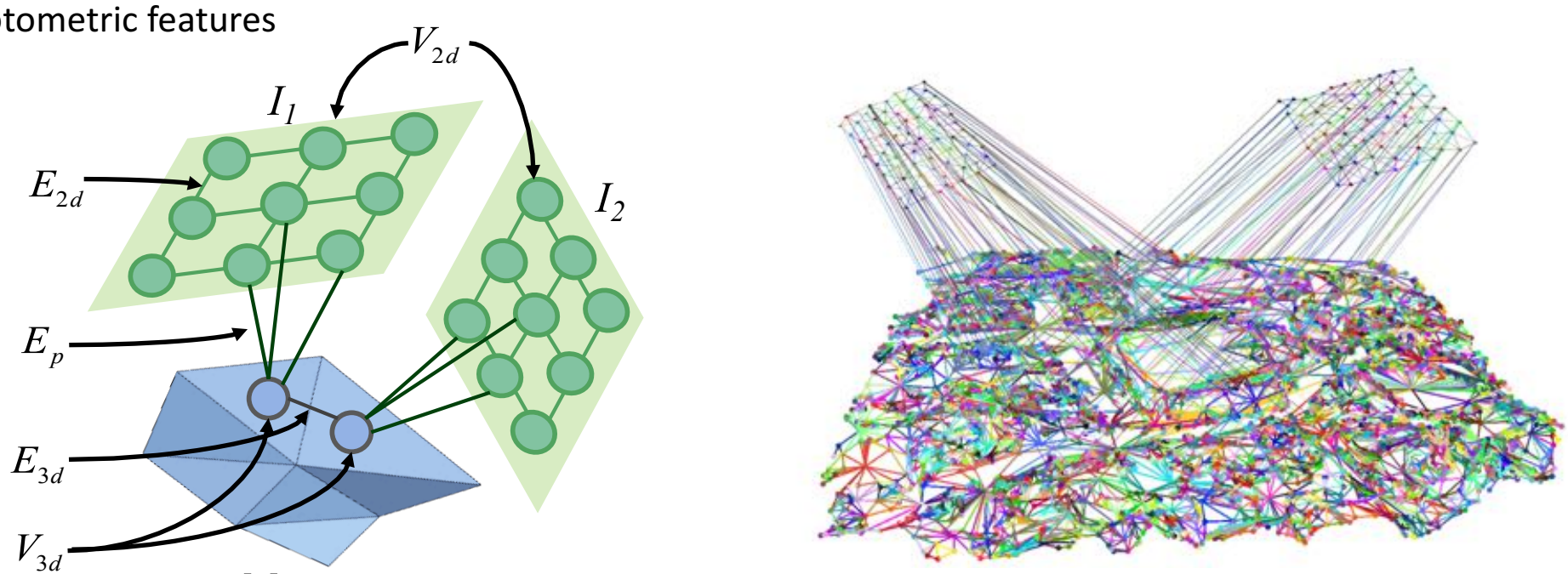


Graph visualization

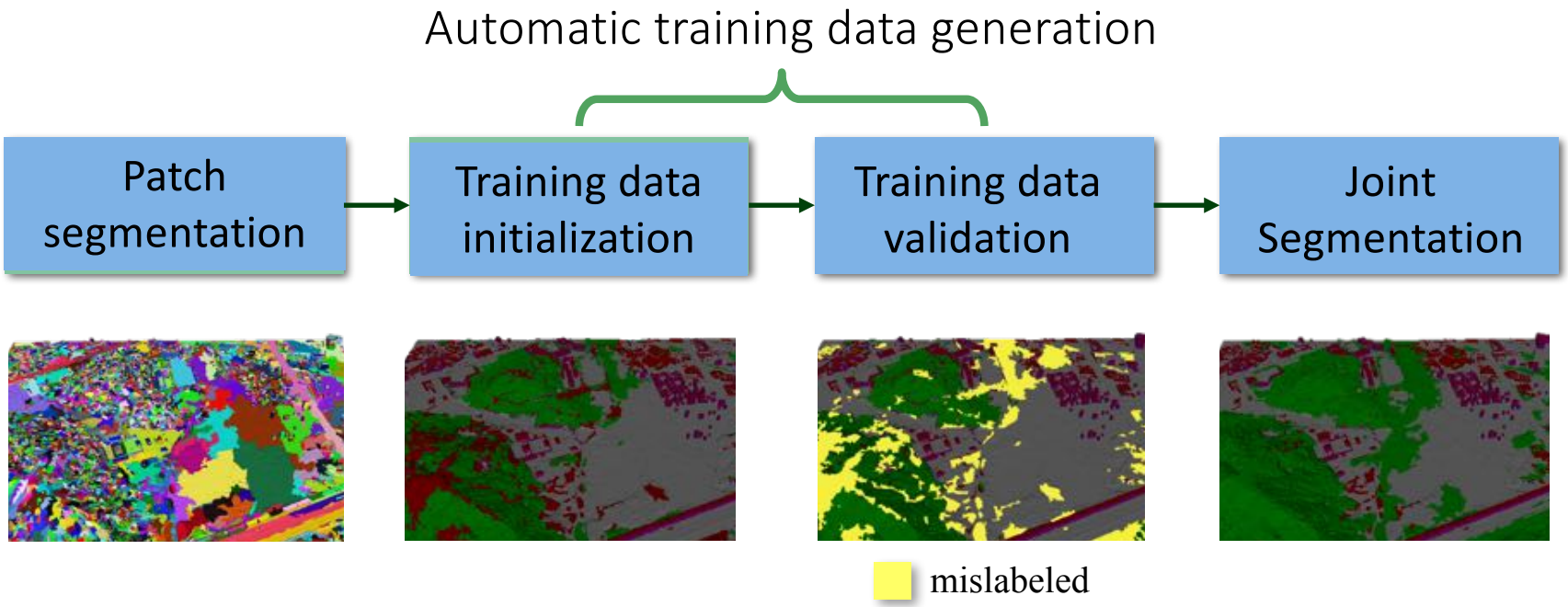


Correspondence Graph

Graph visualization



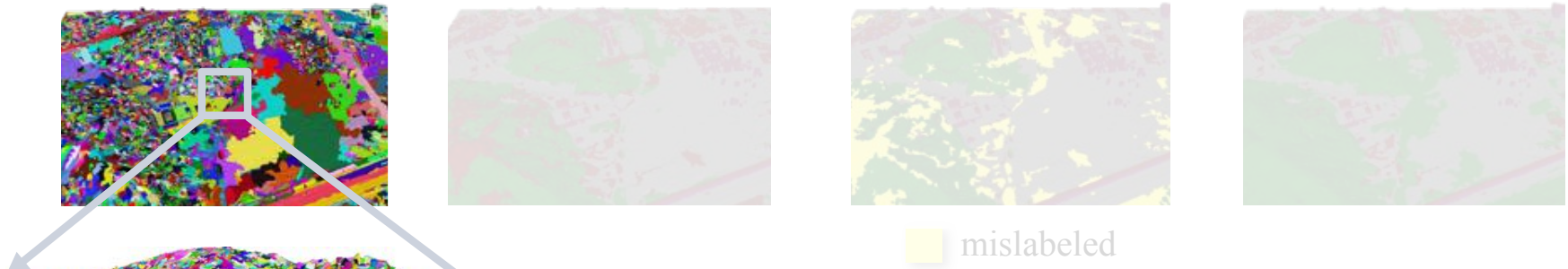
Workflow


Automatic training data generation
Patch segmentation
Training data initialization
Training data validation
Joint Segmentation
mislabeled

Workflow

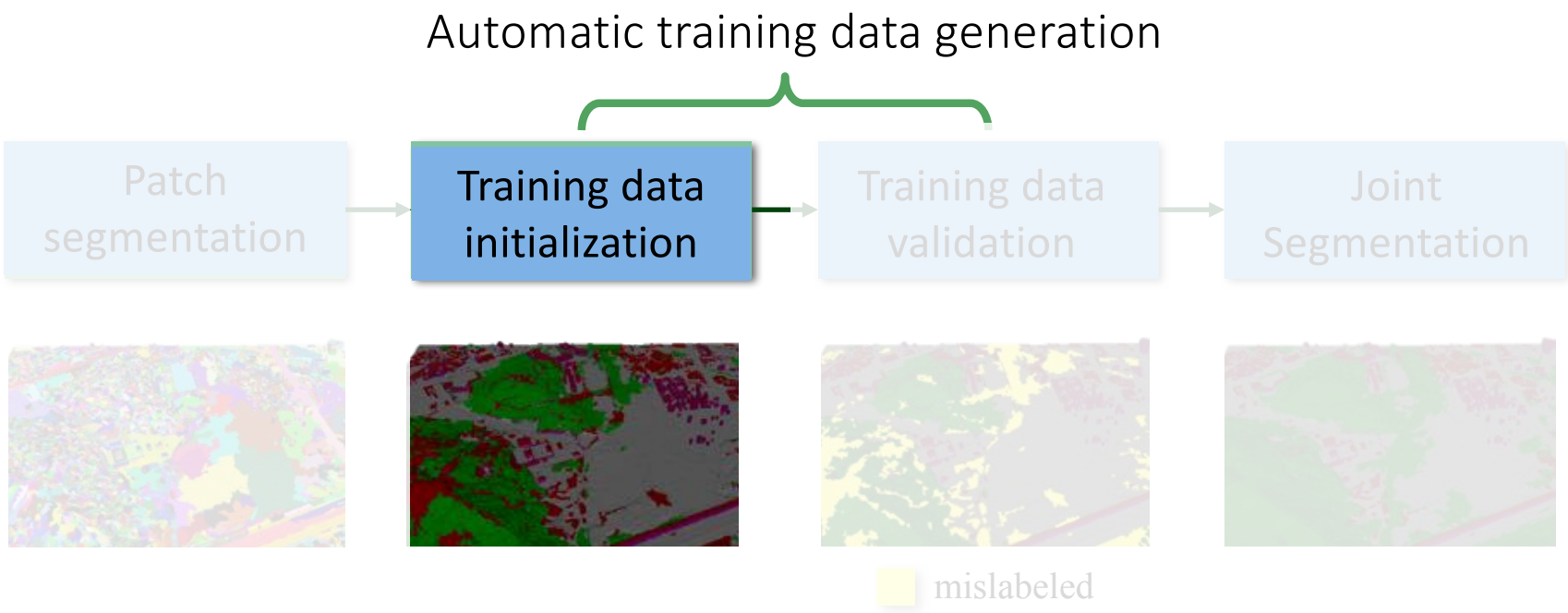
Automatic training data generation

Patch segmentation → Training data initialization → Training data validation → Joint Segmentation

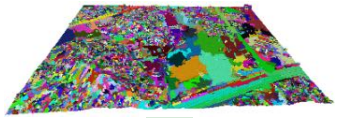


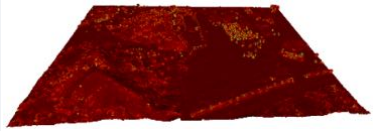
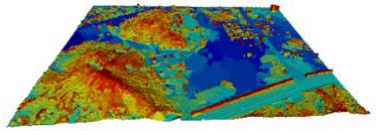
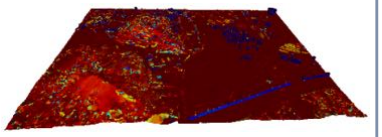
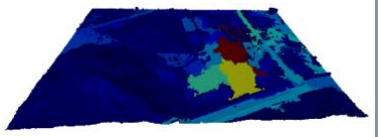
Patch-based representation
[Liu & Wang 2015]

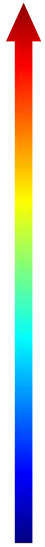
Workflow


Automatic training data generation
Patch segmentation
Training data initialization
Training data validation
Joint Segmentation
mislabeled

Training Data Initialization

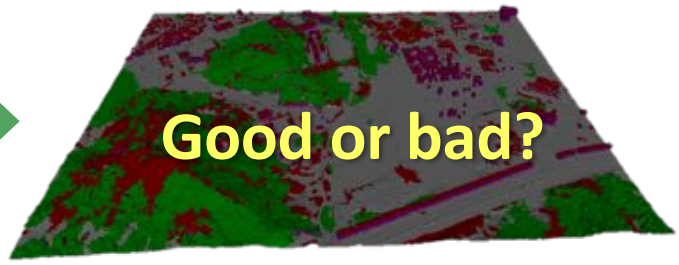


Feature / Category	Planarity	Relative altitude	Horizontality	Area
				
Ground				
Roof				
Facade				
Vegetation				



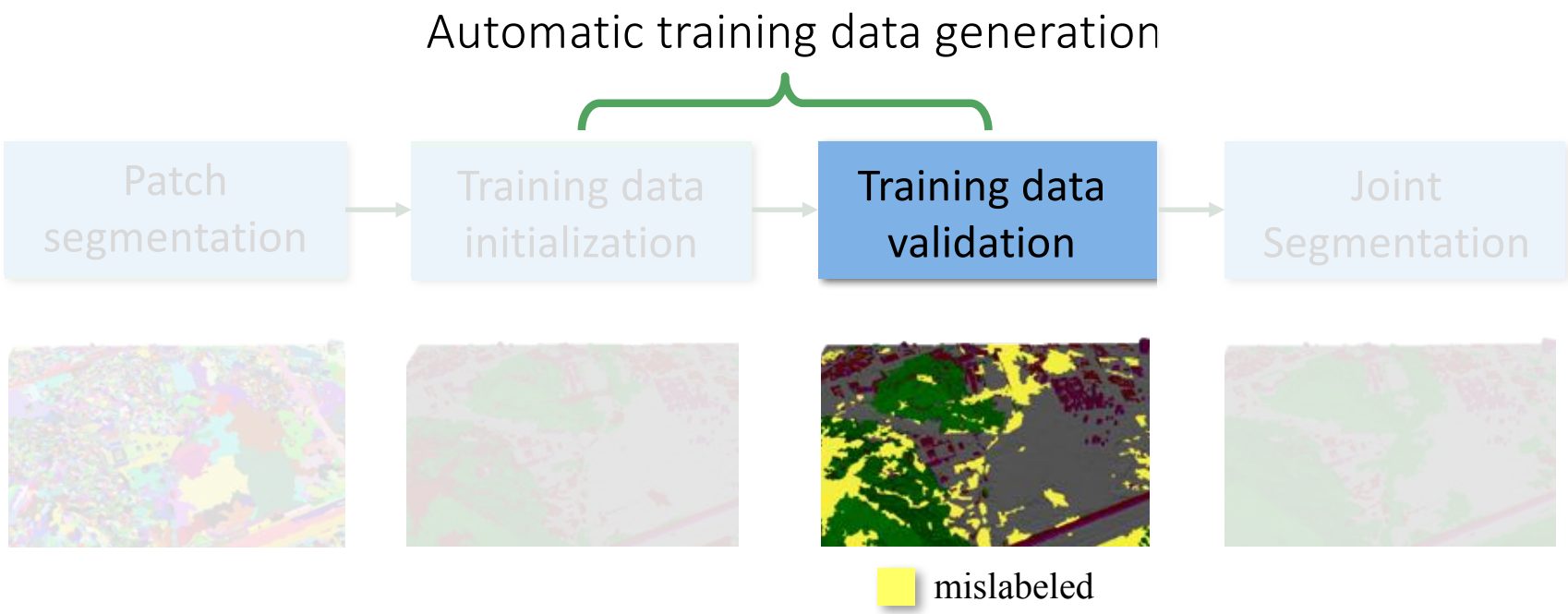
Training data initialization result

$$E_p(y_i) = \begin{cases} 1 - f_p \cdot f_h \cdot \overline{f_a}, & y_i = ground \\ 1 - f_p \cdot f_h \cdot f_a, & y_i = roof \\ 1 - f_p \cdot \overline{f_h}, & y_i = facade \\ 1 - \overline{f_p} \cdot f_h, & y_i = tree \end{cases}$$



Roof
Facade
Ground
vegetation

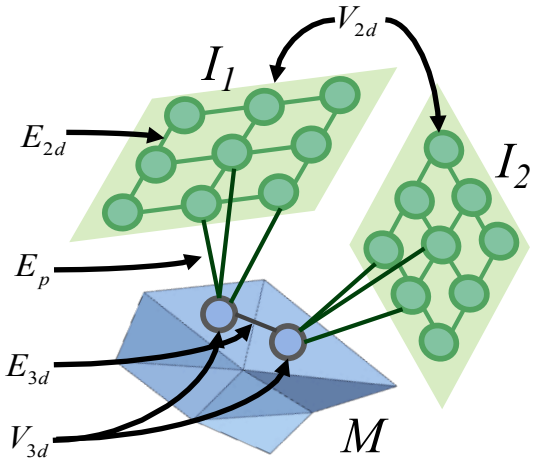
Workflow


Automatic training data generation
Patch segmentation
Training data initialization
Training data validation
Joint Segmentation
mislabeled

Training Data Validation

- Validation Conditional Random Field (CRF)

$$E(\mathbf{y},\mathbf{Y}) = \sum_{x_i \in \mathcal{V}_{2d}} \varphi_i(y_i) + \sum_{X_i \in \mathcal{V}_{3d}} \phi(Y_i) + \sum_{(x_i,x_j)\in\mathcal{E}_{2d}} \varphi_{ij}(y_i,y_j) + \sum_{(X_i,X_j)\in\mathcal{E}_{3d}} \phi_{ij}(Y_i,Y_j) + \sum_{(x_i,X_j)\in\mathcal{E}_{\mathcal{P}}} \Phi_{ij}(y_i,Y_j)$$



- Data term $\varphi_i(y_i) = \begin{cases} w_i(1-P_i) & y_i = l_{good} \\ w_i P_i & y_i = l_{bad} \end{cases}$

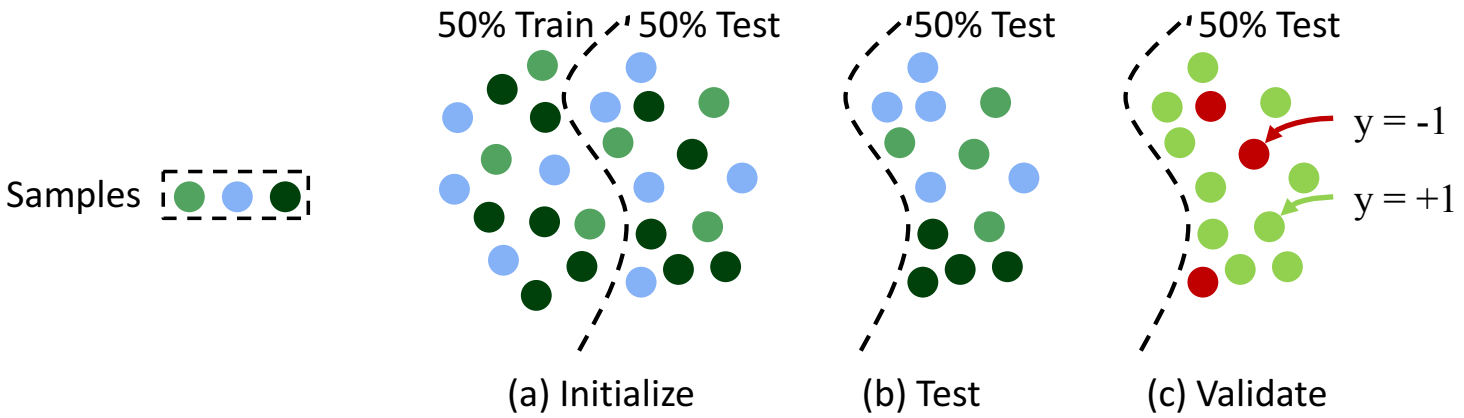
$$L = \{l_{good}, l_{bad}\}$$

- Smoothness term $\varphi_{ij}(y_i,y_j) = \begin{cases} 0 & \text{if } y_i = y_j \\ \lambda & \text{otherwise} \end{cases}$

- Correspondence term $\phi_{ij}(y_i,Y_j) = \begin{cases} \infty & \text{if } Y_j = l_{bad}, y_i \neq Y_j \\ (1-2P_j)(0.1|\mathcal{V}_{2d}|) & \text{if } Y_j = l_{good}, y_i \neq Y_j \\ 0 & \text{if } Y_j = l_{good}, y_i = Y_j \end{cases}$

Training Data Validation

- Data term: labeling confidence based on cross-validation

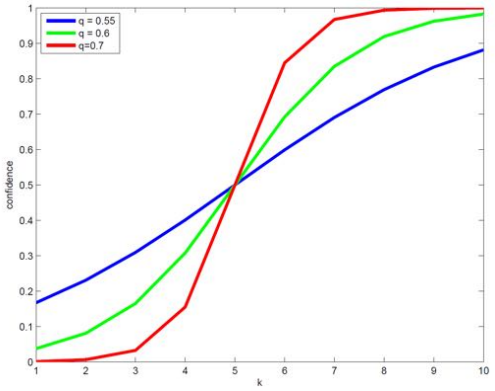


k: the times of being validated

$$\mathbf{P}(k|y=-1) = \binom{N}{k}(1-q)^k q^{N-k}$$

$$\mathbf{P}(k|y=+1) = \binom{N}{k} q^k (1-q)^{N-k}$$

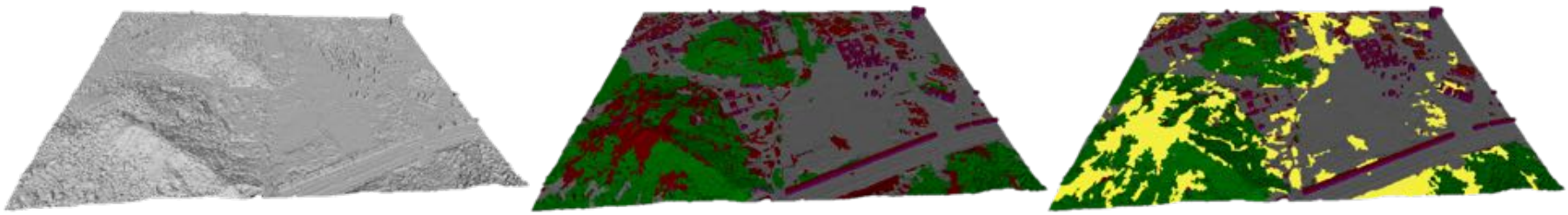
$$P = f(q,k,N) = \frac{q^k(1-q)^{N-k}}{q^k(1-q)^{N-k} + (1-q)^k q^{N-k}}$$



Training Data Validation Result

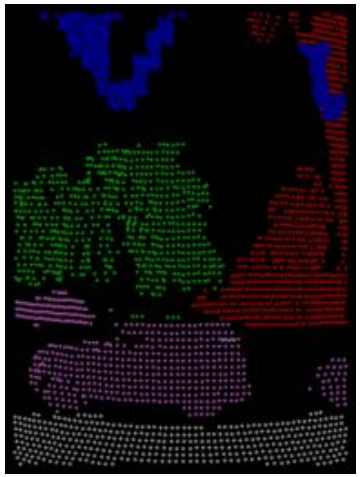
Identified mislabeled sample

MVS data

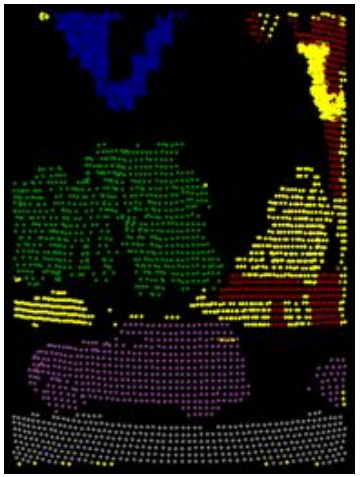


Scanned data

(a) Input



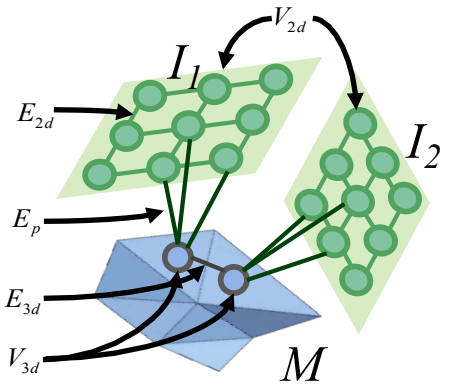
(b) Initial



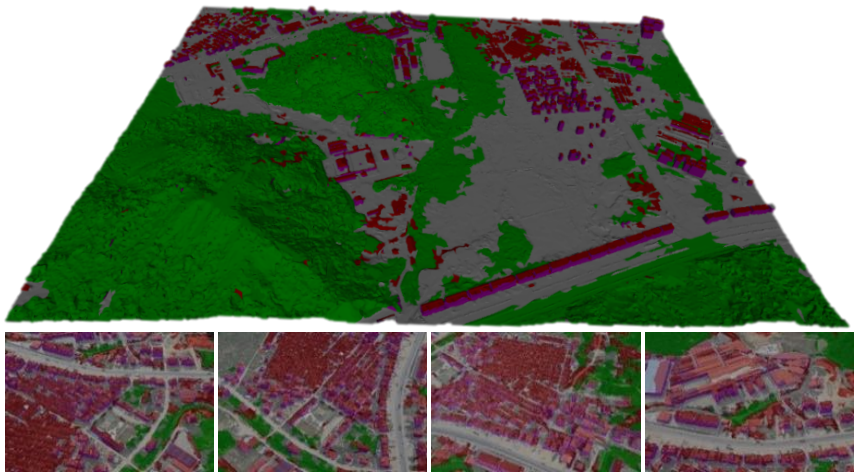
(c) Validated

Joint Segmentation Results

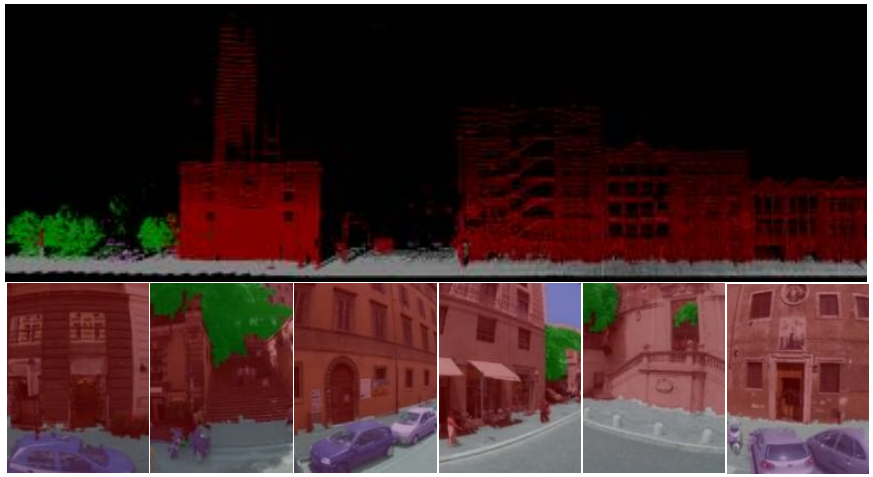
- Apply similar CRF over the correspondence graph



$$L = \{categories\}$$



(a) Joint segmentation result with MVS data



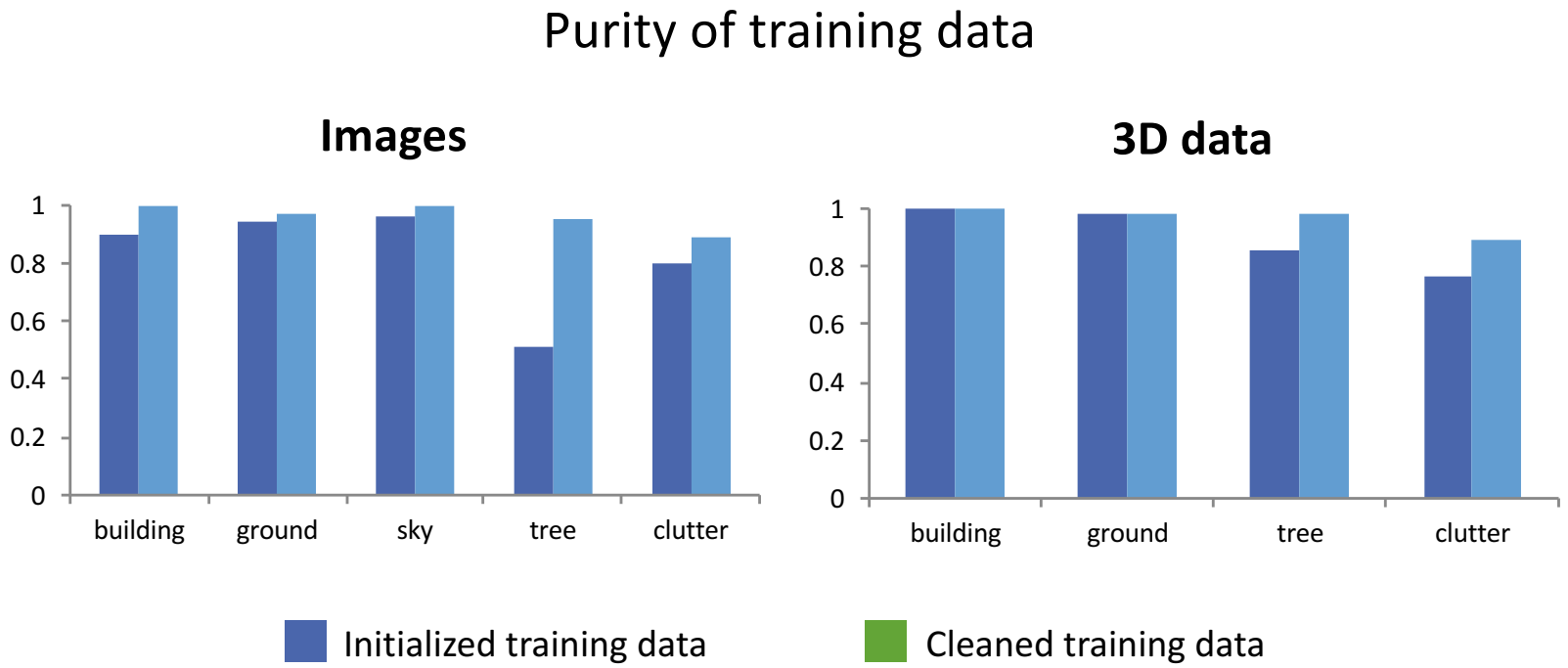
(b) Joint segmentation result with scanned data

Evaluation

- Evaluation of the automatically generated training data
- Evaluation of the parsing performance achieved by the automatically generated training data

Evaluation

- Evaluation of the automatically generated training data
- Evaluation of the parsing performance achieved by the automatically generated training data



Evaluation

- Evaluation of the automatically generated training data
- Evaluation of the parsing performance achieved by the automatically generated training data

Table 3.4: Comparison with manually labeled data

Training data	Image segmentation		3D data segmentation	
	GA	CAA	GA	CAA
automatically generated	0.839	0.82	**0.911**	**0.832**
manually labeled	**0.843**	**0.822**	0.891	0.805

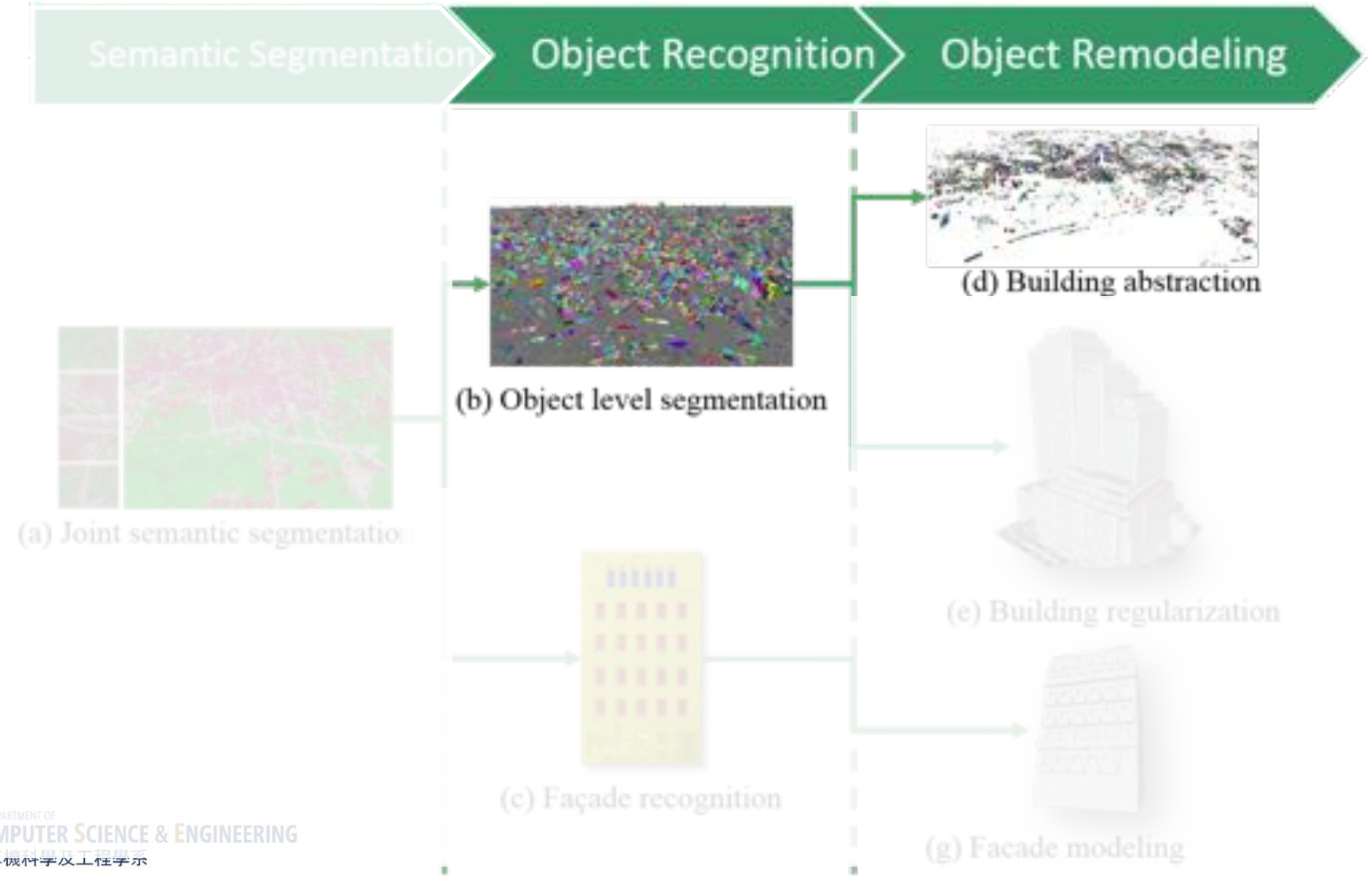
Table 3.5: Comparison with previous method

Method	Image segmentation		3D data segmentation	
	GA	CAA	GA	CAA
Our method	**0.839**	**0.82**	**0.911**	**0.832**
(Ladicky et al., 2009)	0.821	0.809	-	-
(Zhang et al., 2010)	0.8	0.792	-	-
(Anguelov et al., 2005)	-	-	0.89	0.75
(Munoz et al., 2009)	-	-	0.91	0.787

Contributions

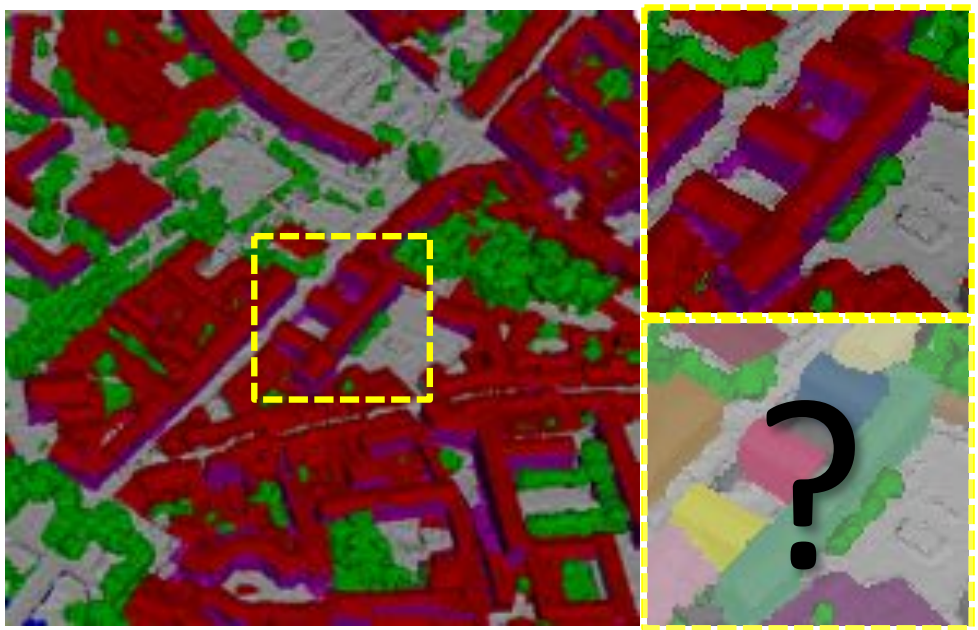
- Patch based representation
- Automatic training data generation
- Model multi-view label consistency

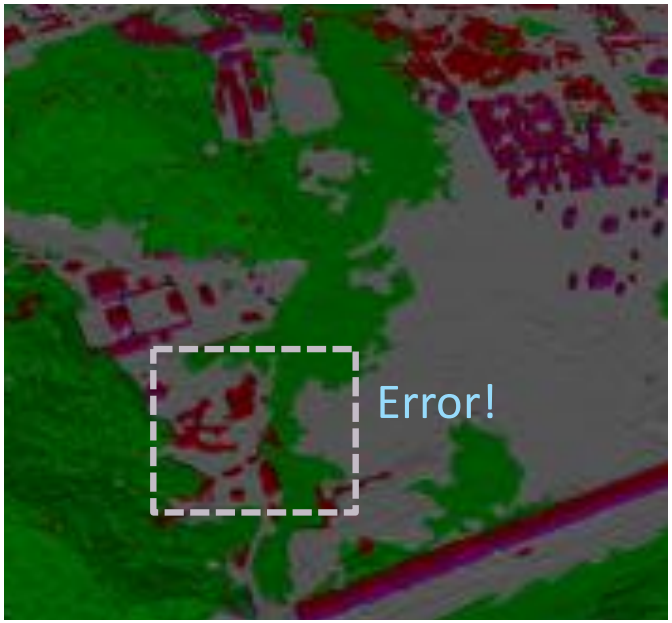
Framework


Semantic Segmentation
Object Recognition
Object Remodeling
(d) Building abstraction
(b) Object level segmentation
(a) Joint semantic segmentation
(e) Building regularization
(c) Façade recognition
(g) Facade modeling

Object Level Segmentation Motivation

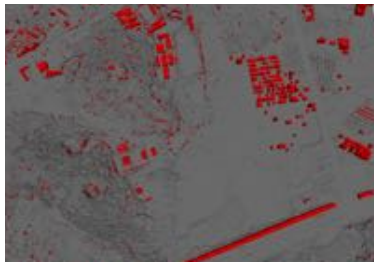
- Objects are not separated
- Category-level segmentation are not precise enough

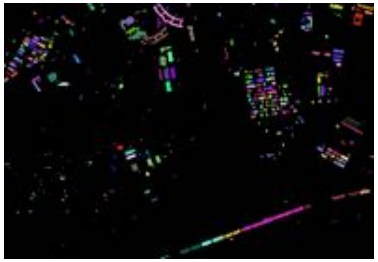
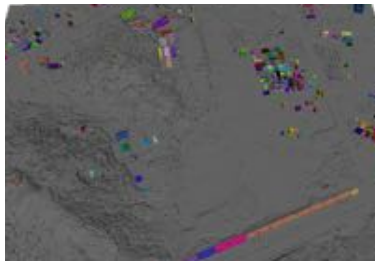




Object Level Segmentation

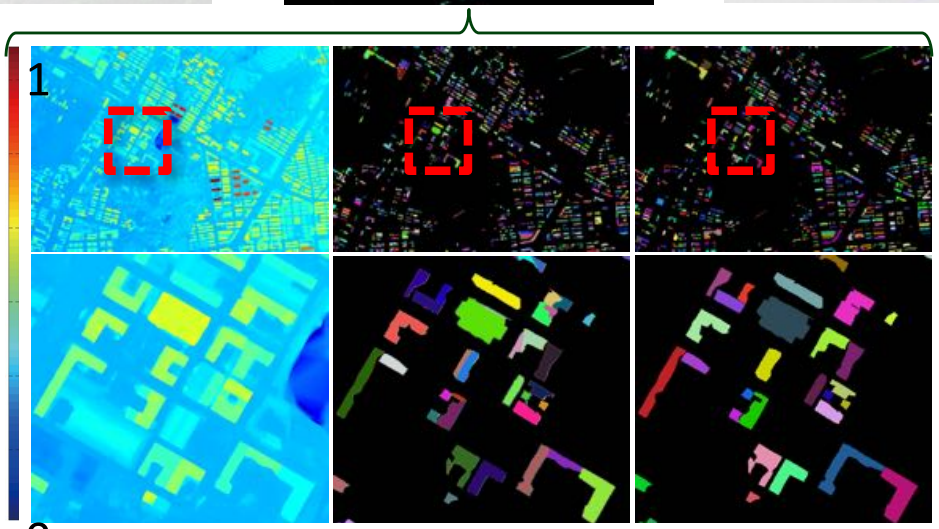
Terrain extraction → Urban object classification → 2D map consolidation → segmentation 2D to 3D



Object Level Segmentation

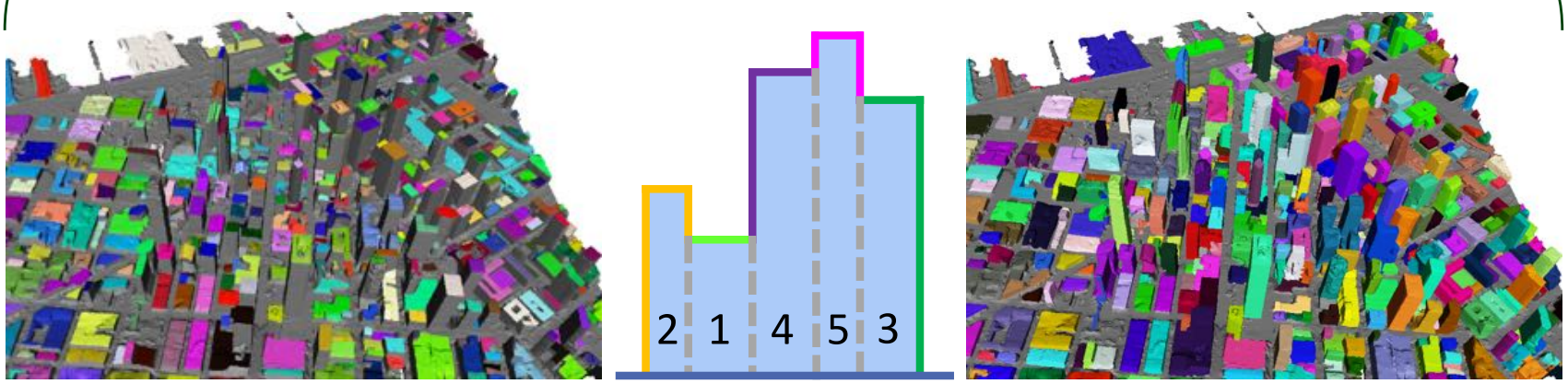
Terrain extraction → Urban object classification → 2D map consolidation → segmentation 2D to 3D



(a) Height map (b) Over-segmented map (c) Final segmented map

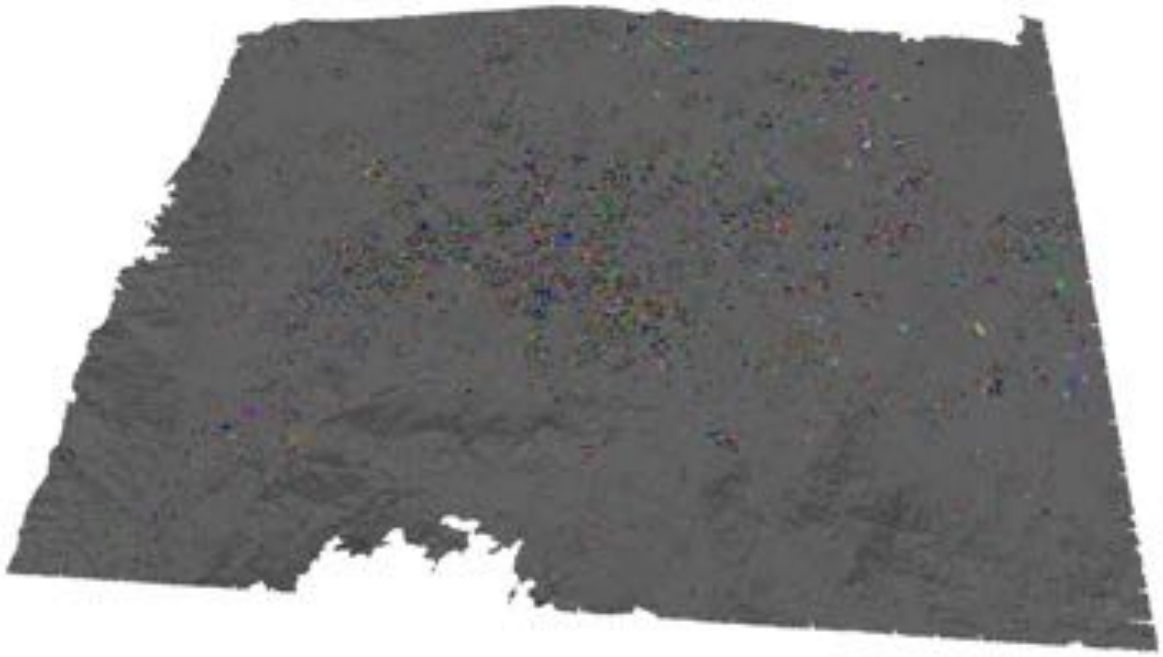
Object Level Segmentation

Terrain extraction → Urban object classification → 2D map consolidation → segmentation 2D to 3D

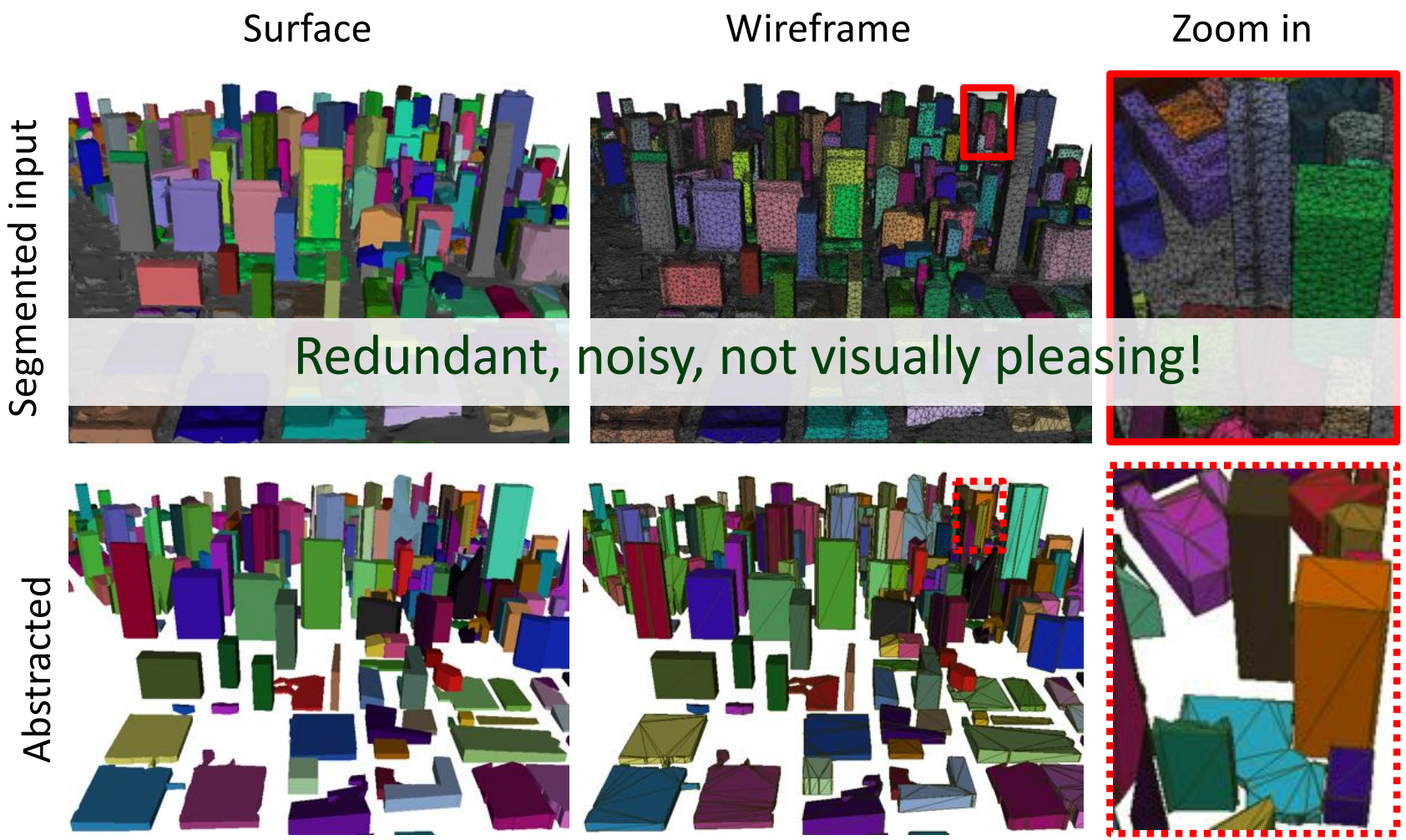


(a) Segmented roofs (b) Flooding illustration (c) Segmented objects

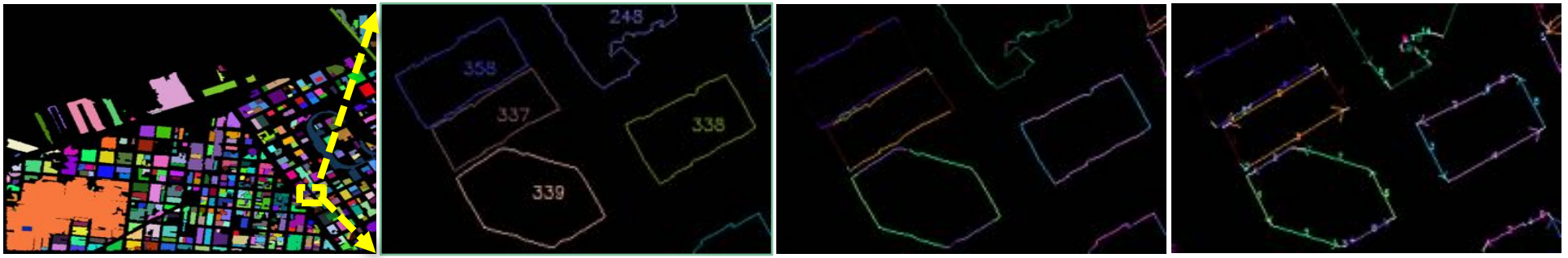
Object Level Segmentation Results



Building Abstraction Motivation



Contour Abstraction



(a) Input contours (b) Labeled contour points (c) Abstracted contours

- Higher order CRF formulation

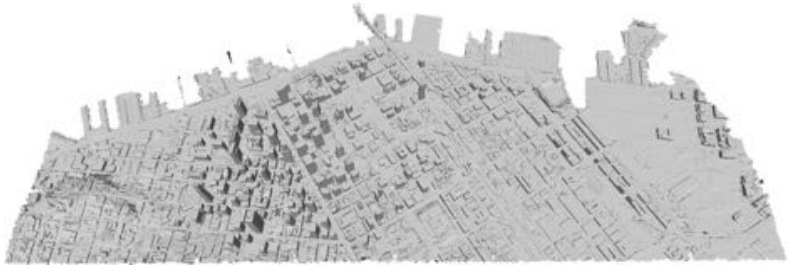
$$E = \sum_{x_i \in \mathcal{S}} \phi_i(y_i) + \sum_{x_j \in \mathcal{N}(x_i)} \psi_{ij}(y_i, y_j) + \sum_{c \in \mathcal{C}} \psi_R(\mathbf{y}_c)$$

Higher order regularity term

- Higher order regularity

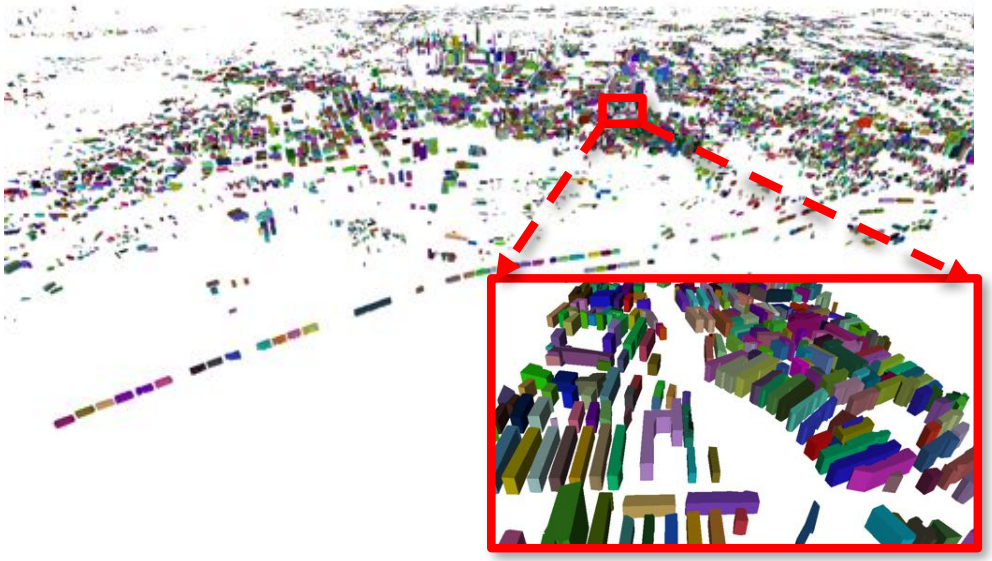
$$\psi_c(\mathbf{y}_c) = \begin{cases} 0 & \text{if } \mathbf{y}_c = \mathscr{R}(\mathbf{x}_c) \\ \theta_p^h |c|^{\theta_\alpha} & \text{otherwise} \end{cases}$$

Results and Evaluation



(a) Input of "SF"

(b) Abstraction of "SF"



(c) Abstraction of "CityA"

Save 94% storage

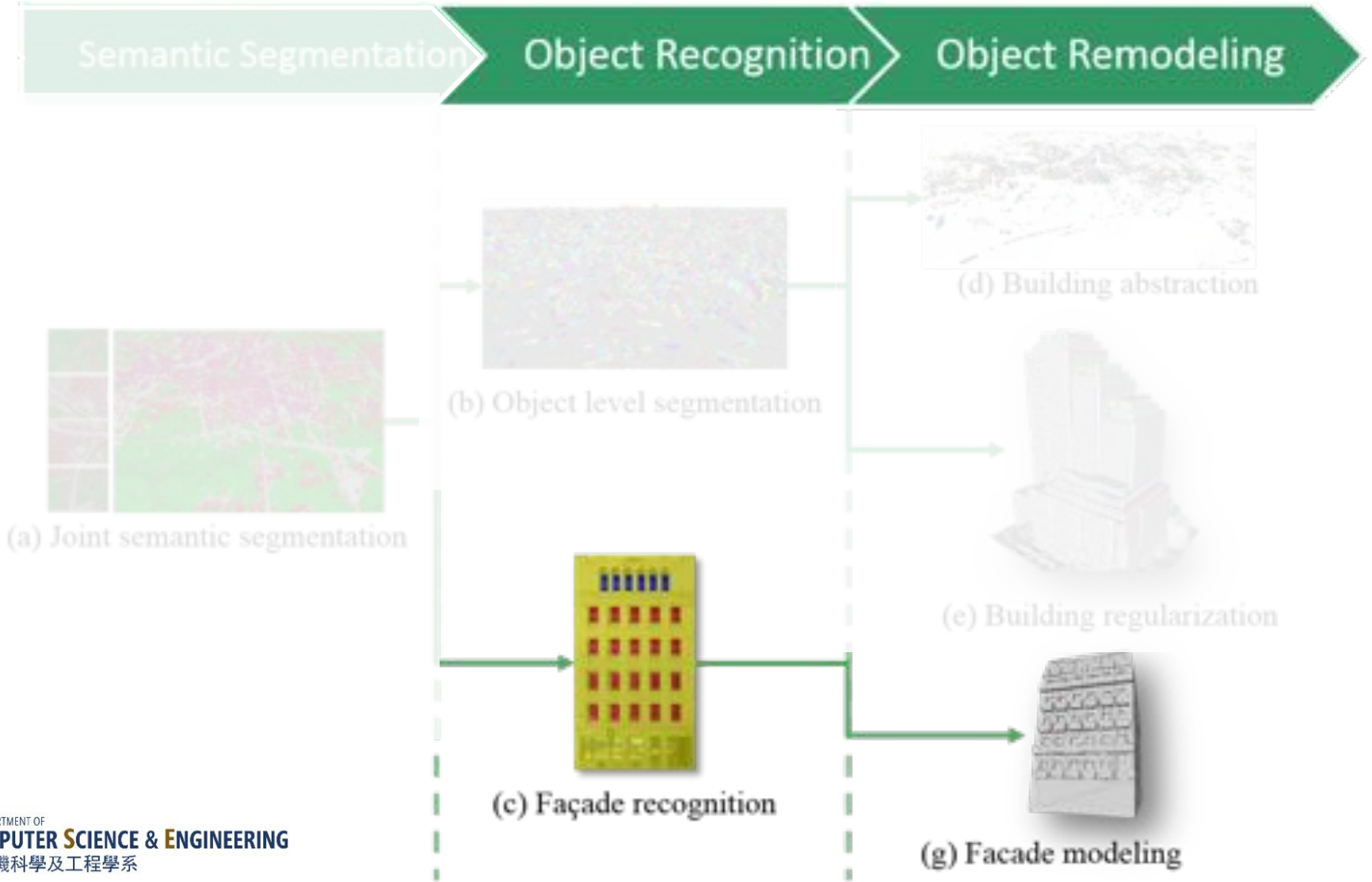
Dataset	Area (km^2)	#F (input)	#F (abstract)	CR	#object	Avg #line	precision	recall
SF	6	746012	50472	6.77%	893	9	0.947	0.897
CityA	50	8550173	508259	5.94%	13014	5	0.948	0.888

Table 5.1: Building object abstraction statistics

Contributions

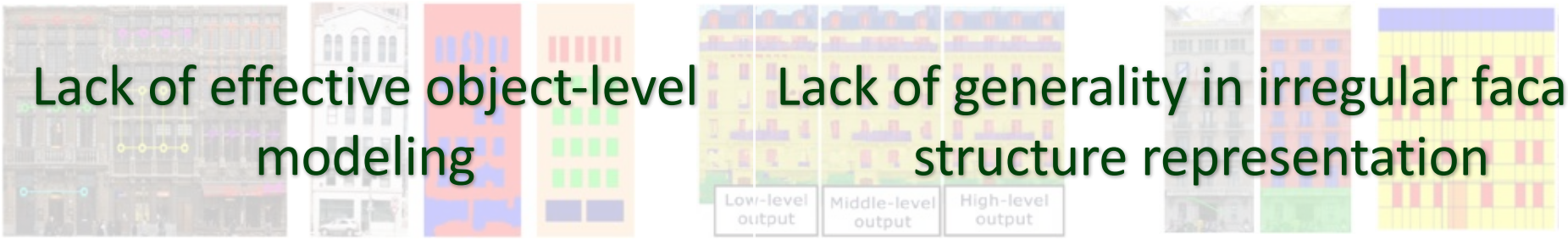
- Segmentation at the object level
- Higher order CRF encodes global regularities

Framework


Semantic Segmentation
Object Recognition
Object Remodeling
(a) Joint semantic segmentation
(b) Object level segmentation
(c) Façade recognition
(d) Building abstraction
(e) Building regularization
(g) Facade modeling

Existing Methods

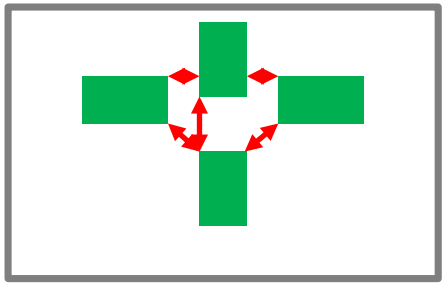
Bottom-up | Combination | Top-down



Translation symmetry [Zhao et al. 2011] | Rank-one approximation [Yang et al. 2012] | Three layers [Martinovic et al. 2012] | Shape grammar [Teboul et al. 2011] | Irregular lattice [Riemensheneider et al. 2012]

Lack of effective object-level modeling

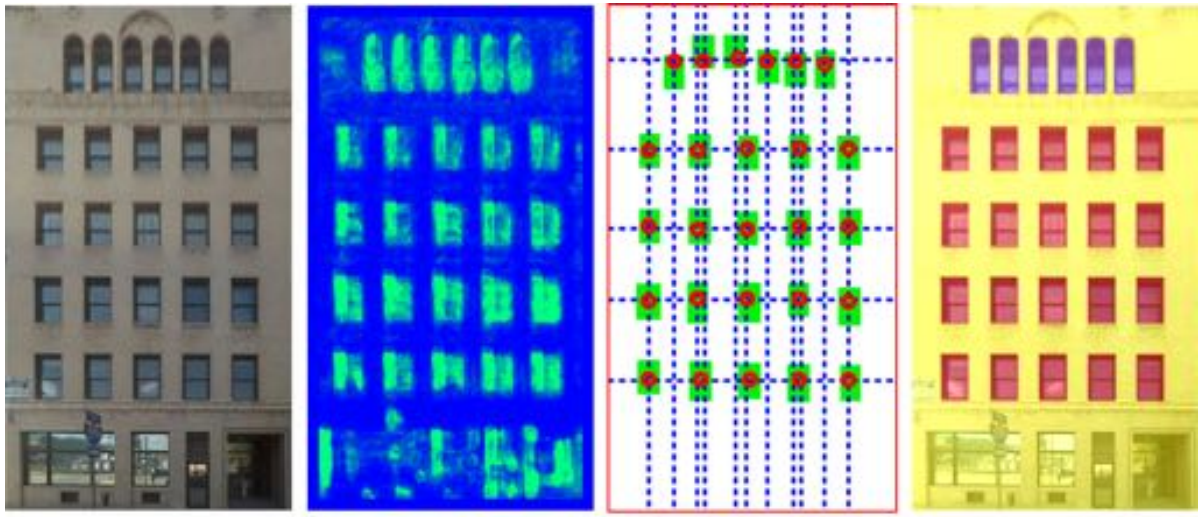
Lack of generality in irregular facade structure representation



Interactions between objects? Noise?

Irregular structure?

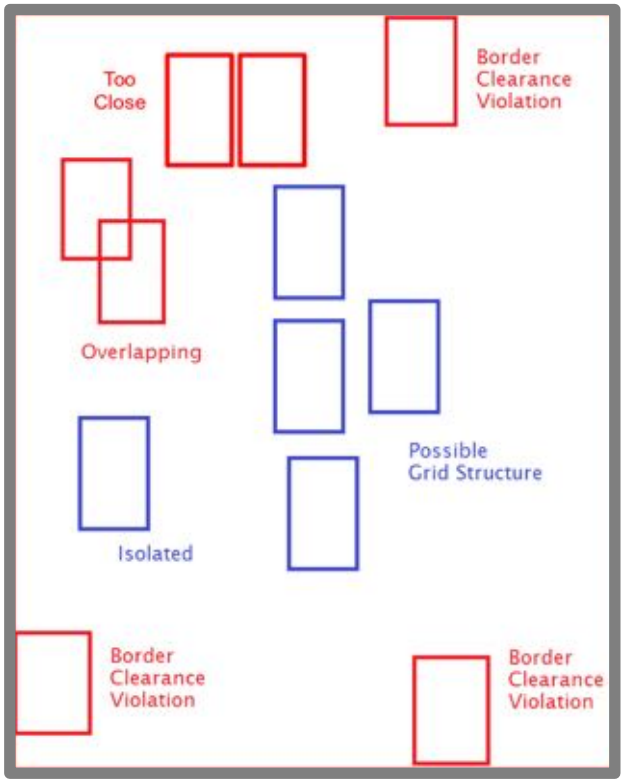
Our Approach



(a) input image;
(b) object (window) probability map;
(c) 2D incomplete grid facade structure representation.
(d) regulation result using a multi-lattice structure.

An Object-level Model

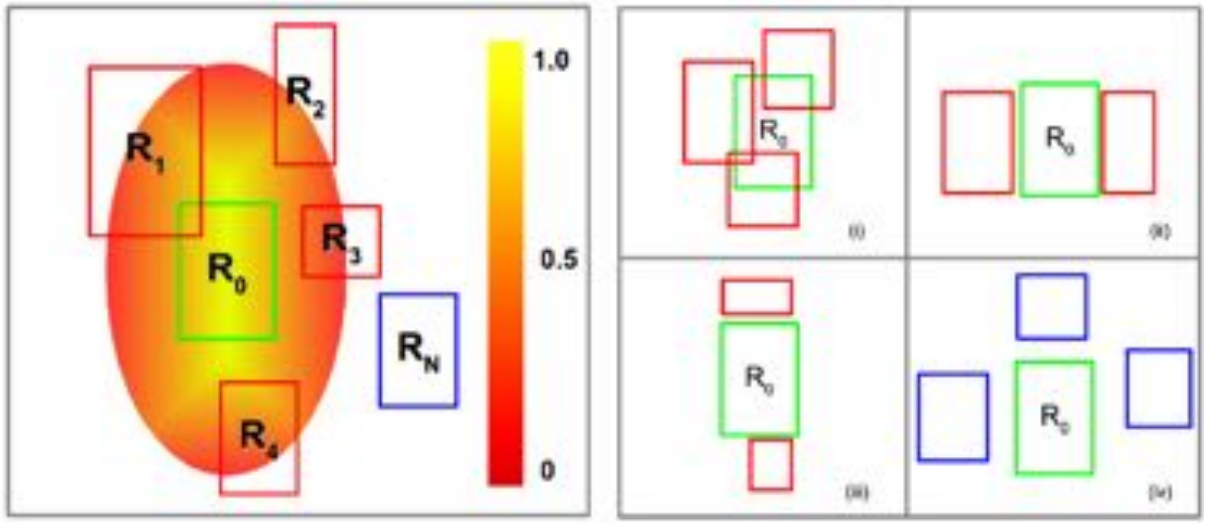
- A marked point process model (MPP)



Sparsely populated objects

Energy Function

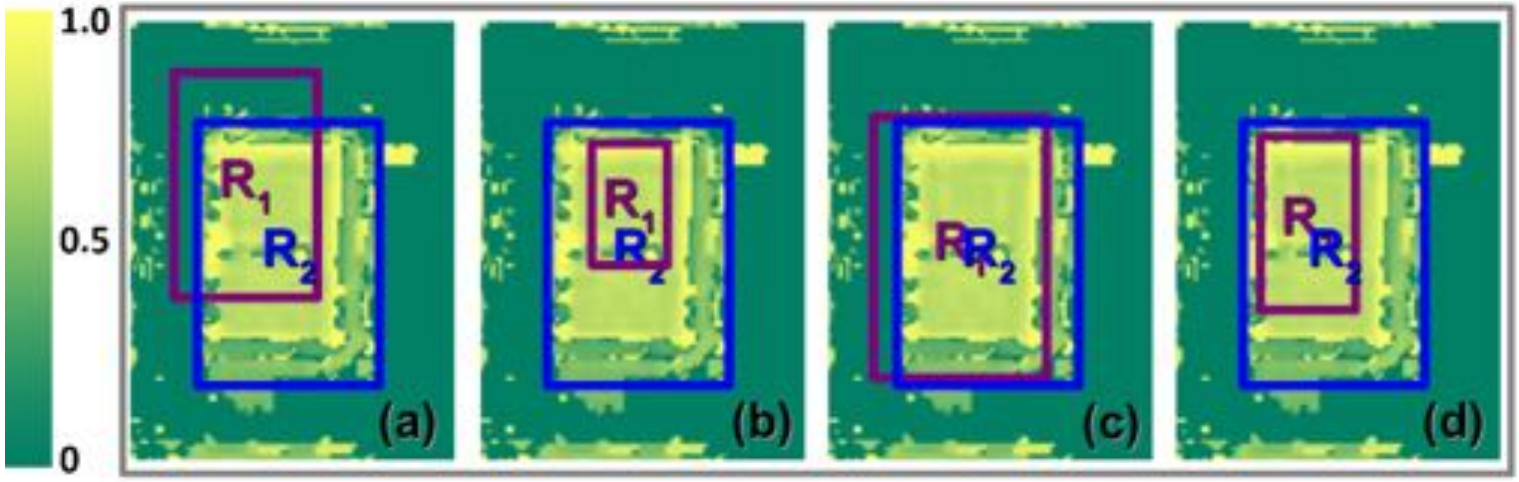
$$U(X) = U_{prior}(X) + \alpha U_{data}(X), \quad \alpha \in (0,1)$$



$$U_{R_{soft}}(x_i, x_j) = \begin{cases} 1.0 - \frac{t_x^2}{\delta_x^2} - \frac{t_y^2}{\delta_y^2} & \text{if } t_x < \delta_x \text{ and } t_y < \delta_y, \\ 0 & \text{otherwise} \end{cases}$$

Energy Function

$$U(X) = U_{prior}(X) + \alpha U_{data}(X), \quad \alpha \in (0,1)$$



Rectangle Rate Contrast Rate Homogeneity Rate Area Rate

$$U_d(x_i) = \max[R(x_i), C(x_i)] * H(x_i) * A(x_i)$$

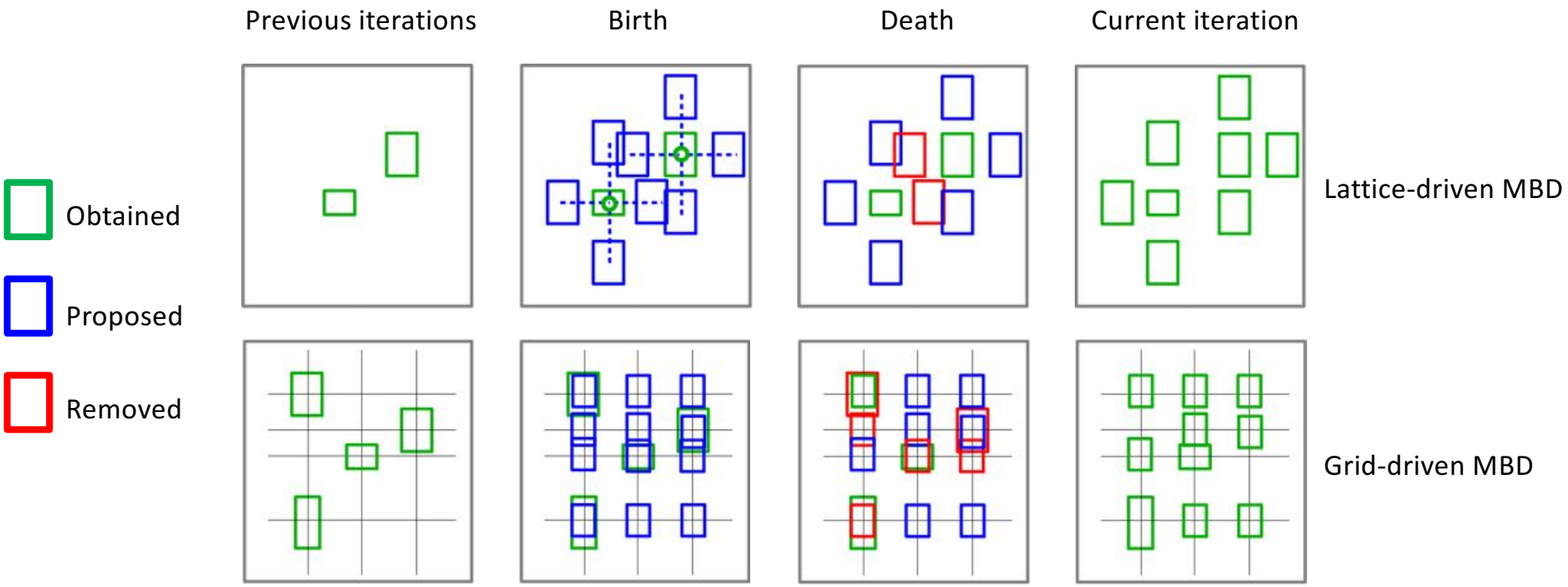
Structure-driven MCMC Optimization

$$U(X) = U_{prior}(X) + \alpha U_{data}(X), \quad \alpha \in (0,1)$$

- The energy function is not convex
- Large solution space

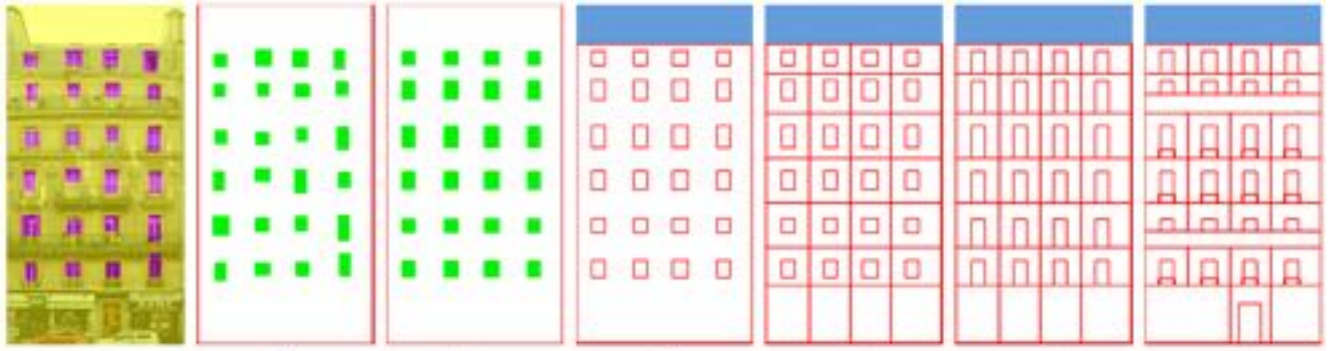
Multiple-birth-and-death (MBD) algorithm

Structure-driven Multiple-birth-and-death (MBD)

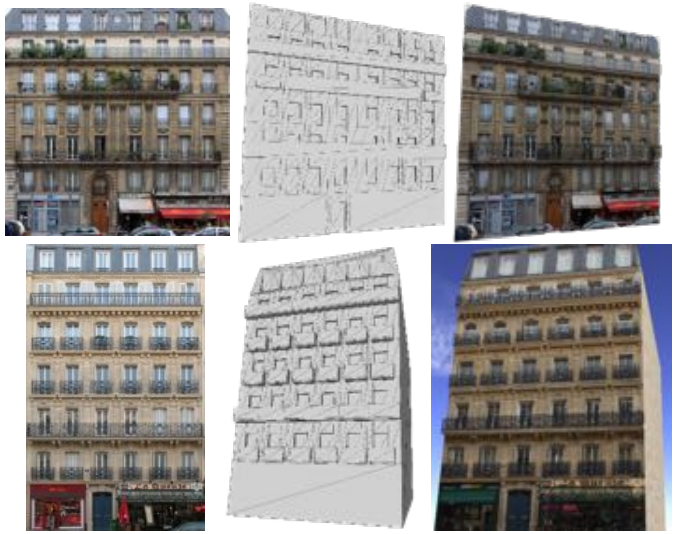


Results

Application - Grammar Factorization



Grammar factorization using detected windows



Input image Remodeled Textured

Rendering result of block data

Evaluation

Table 4.1: Comparison with other methods

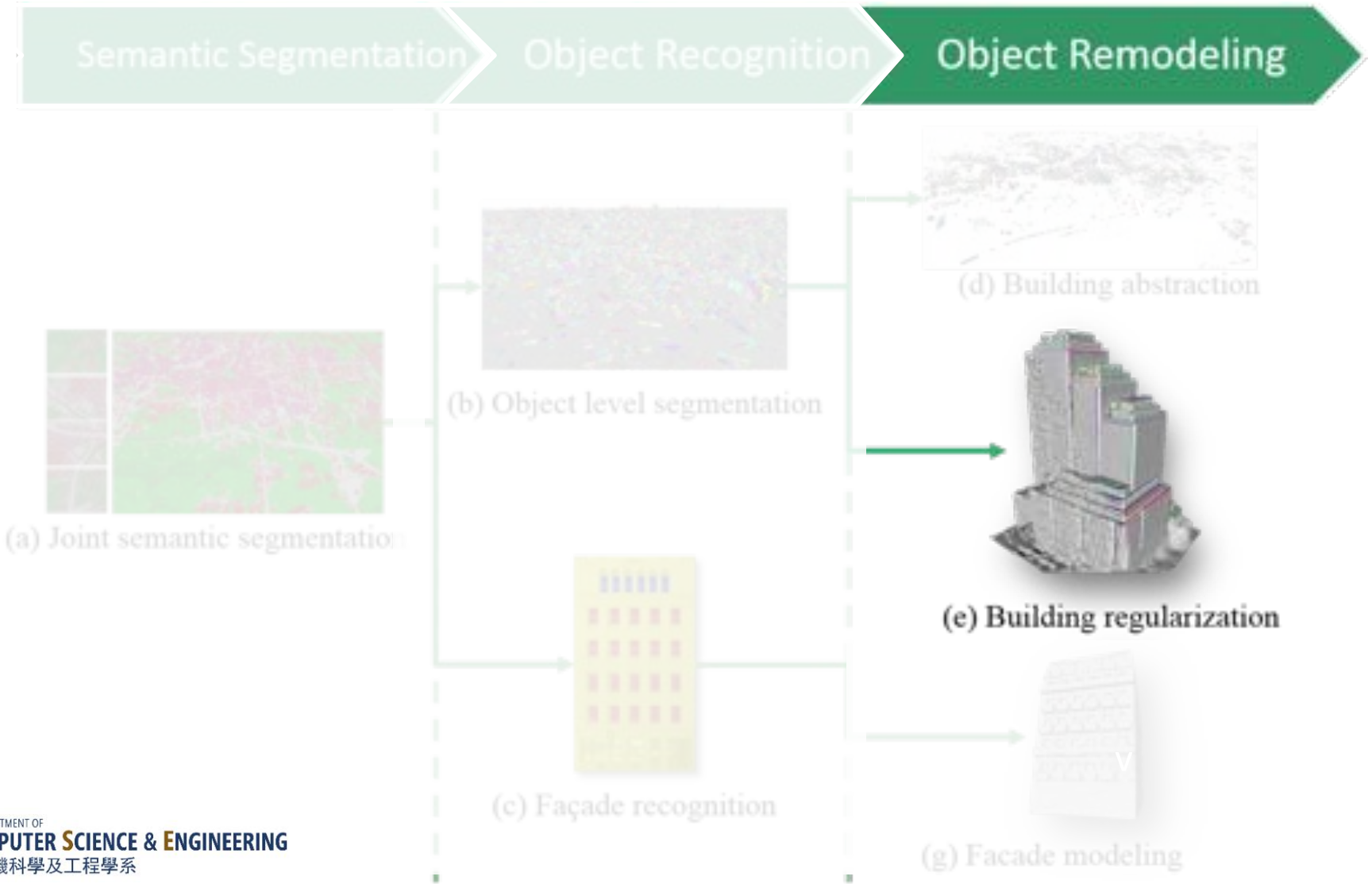
	RL (Teboul et al., 2011)	IL (Riemenschneider et al., 2012)	TL (Martinović et al., 2012)	ours	ours+LR	ours+Gr
ECP	81	68	75	80	**87**	**85**
Boulevard	65	N/A	N/A	**76**	**80**	**78**

"LR": short for with low rank constraint.
"Gr": is short for using grammar constraints

Contributions

- A general facade structure representation
- A marked point process model to model sparsely populated objects in man-made scenes
- A structure-driven Markov Cain Mote Carlo (MCMC) optimization

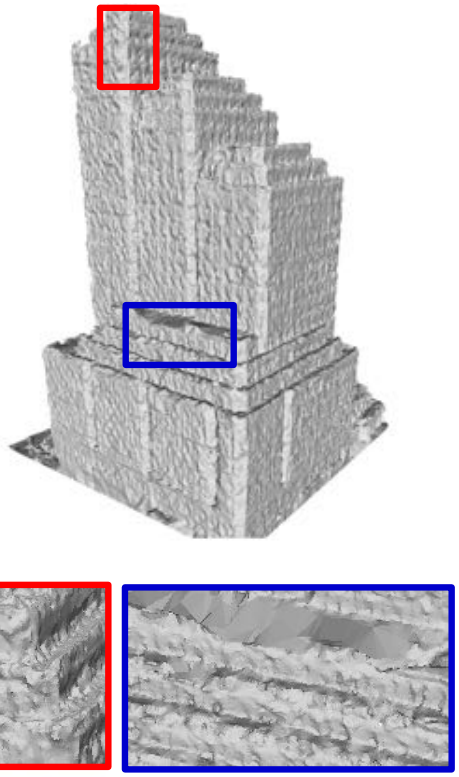
Framework


Semantic Segmentation
Object Recognition
Object Remodeling
(a) Joint semantic segmentation
(b) Object level segmentation
(c) Façade recognition
(d) Building abstraction
(e) Building regularization
(g) Facade modeling

Problem

Noisy, incomplete, distorted structure!

Input MVS model



Textured MVS model

Simplified model

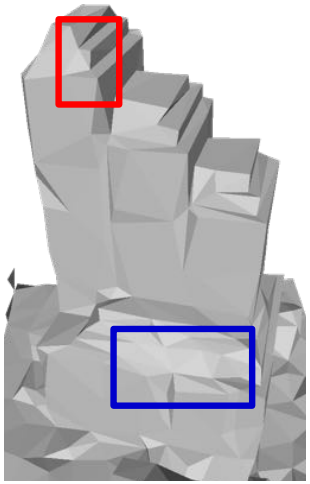


Image-based Building Scaffolding

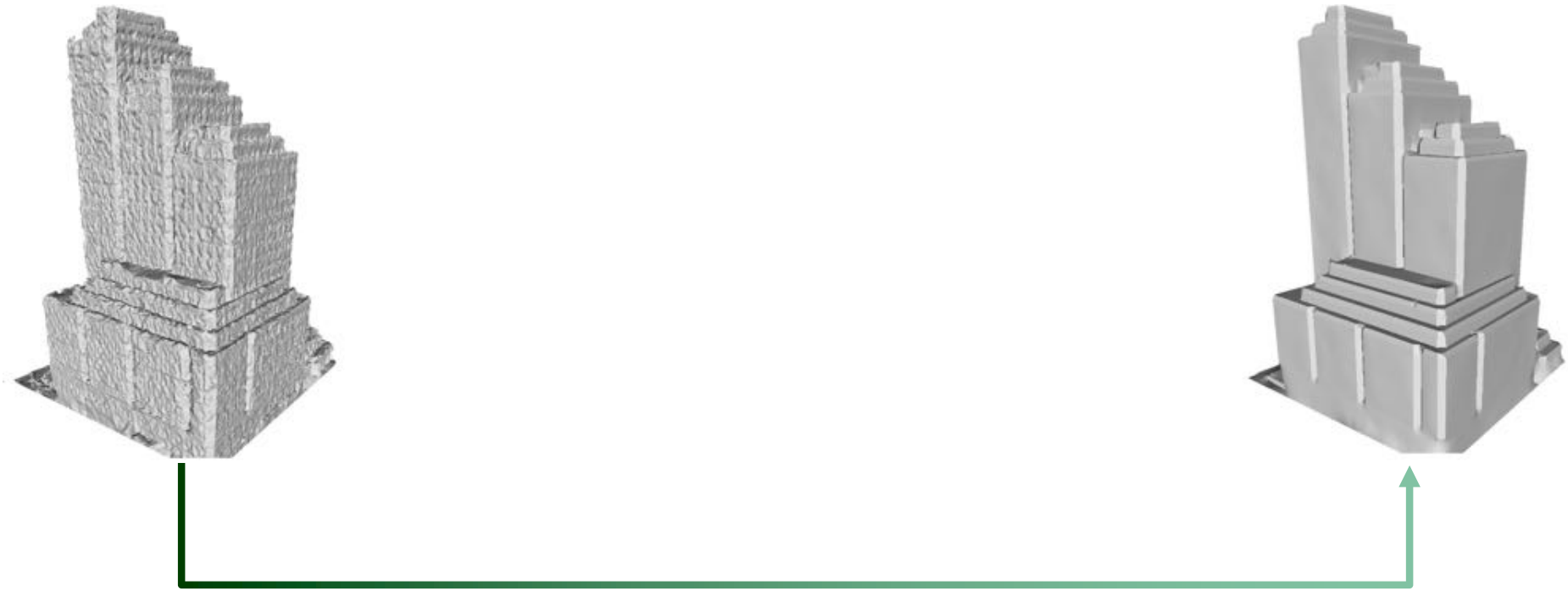
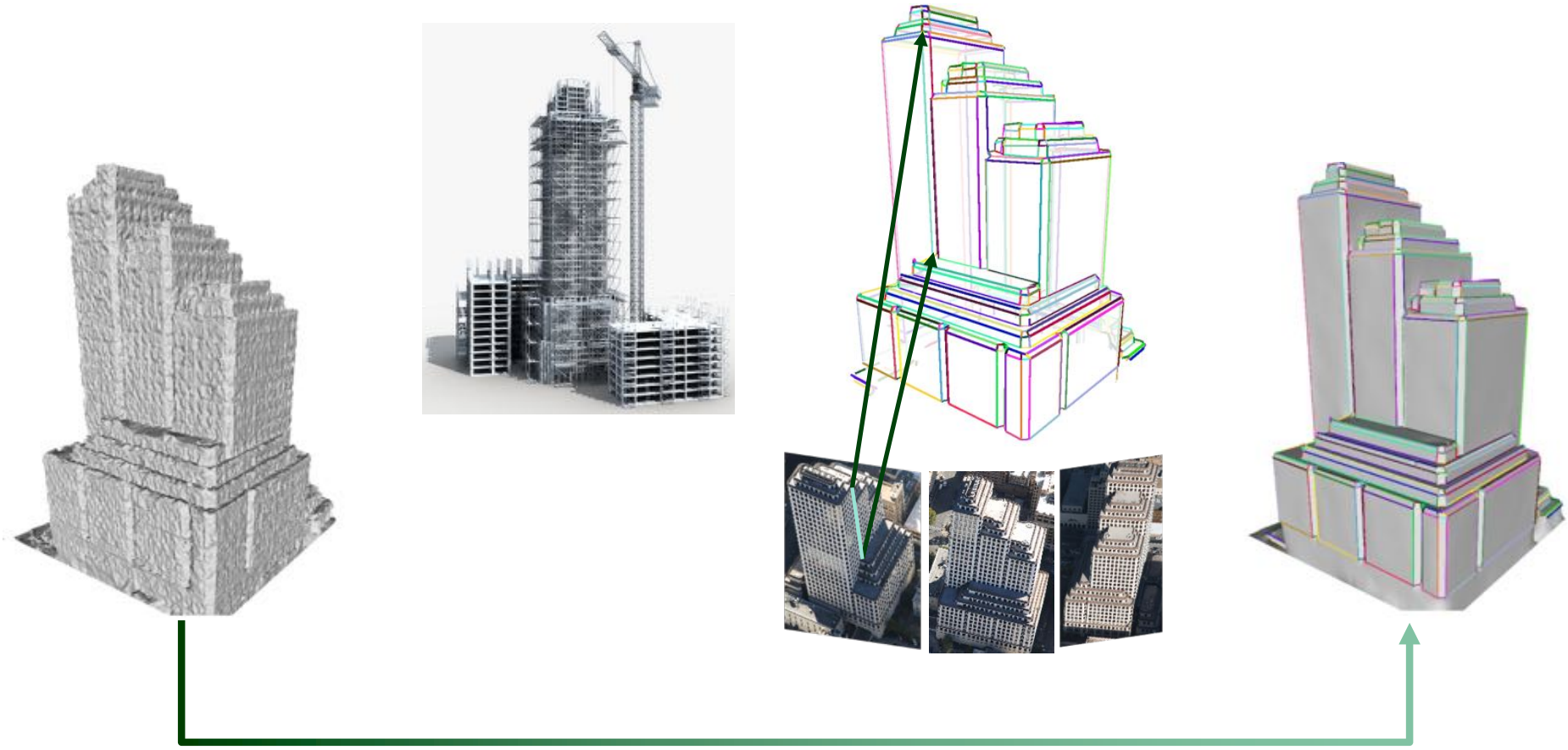
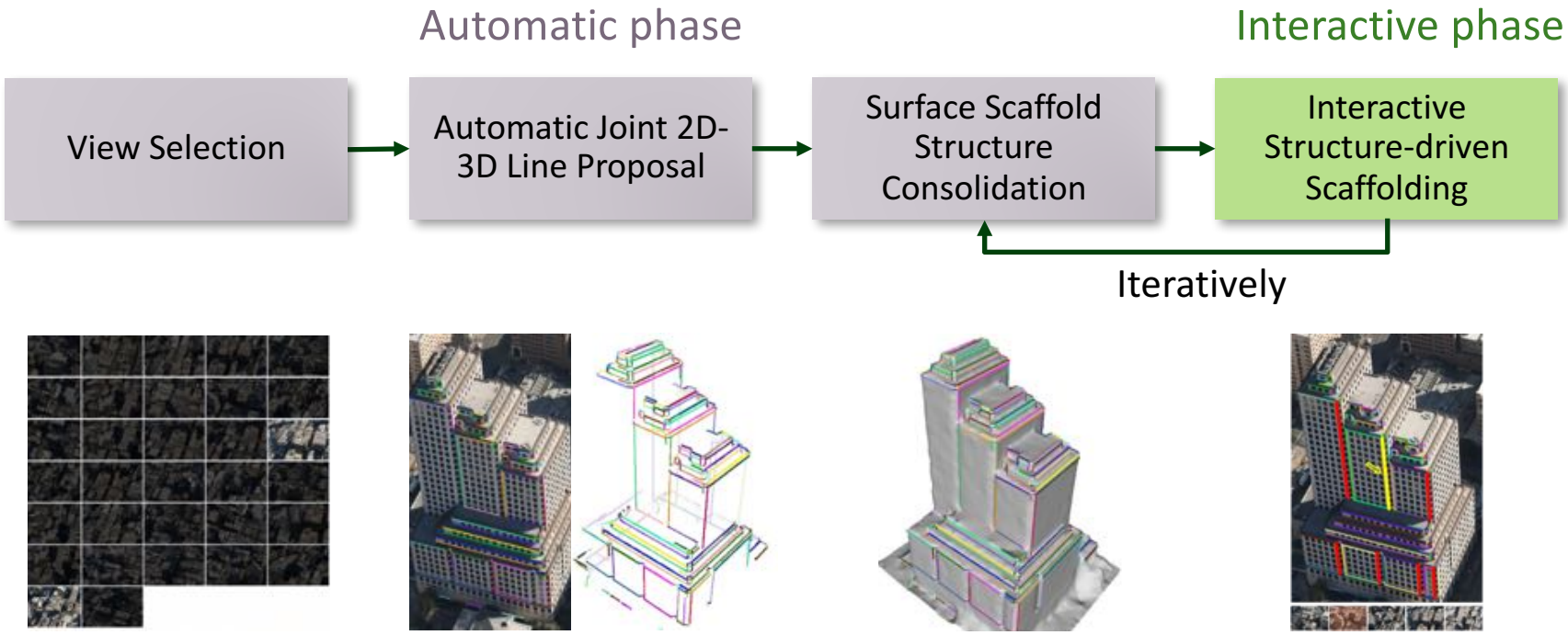


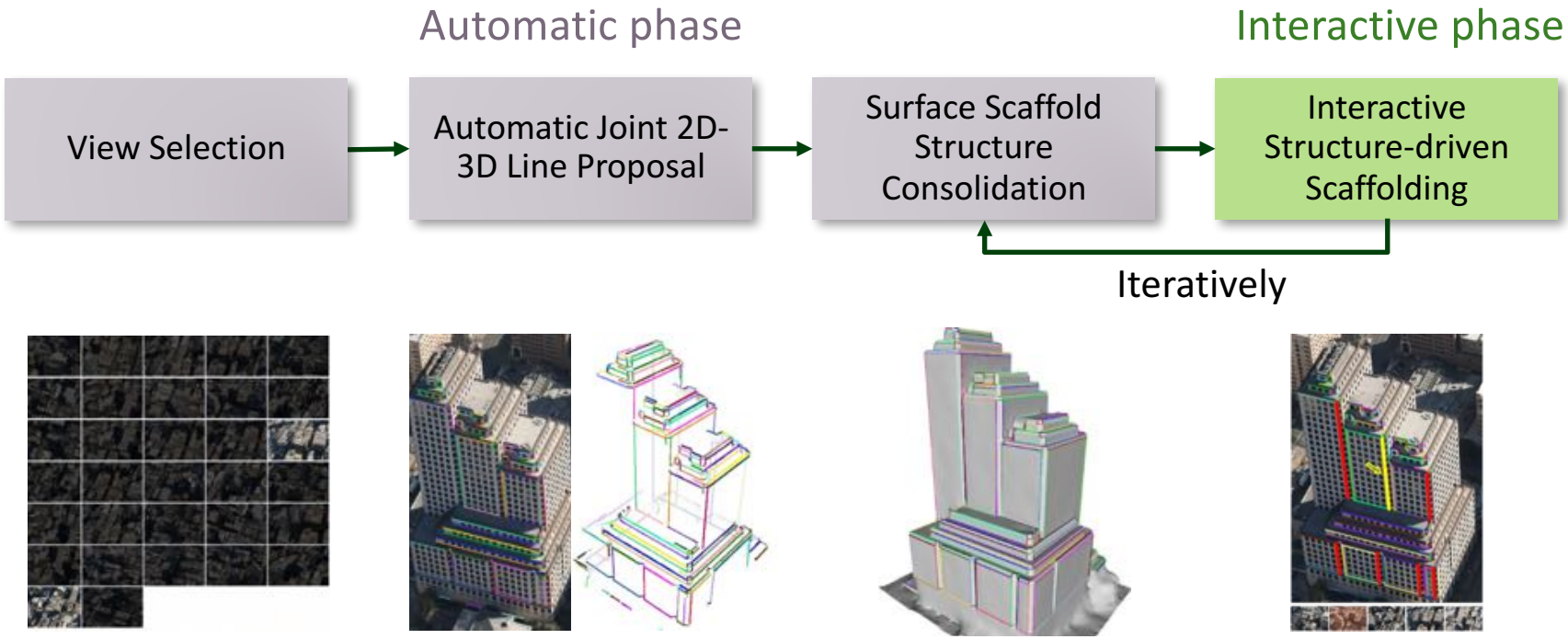
Image-based Building Scaffolding



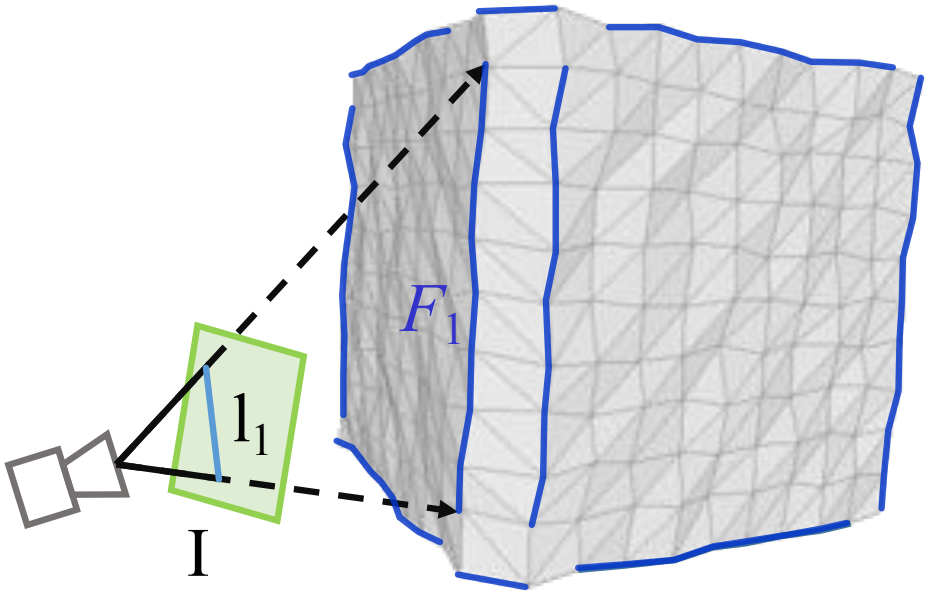
Workflow


Automatic phase
Interactive phase
View Selection
Automatic Joint 2D-3D Line Proposal
Surface Scaffold Structure Consolidation
Interactive Structure-driven Scaffolding
Iteratively

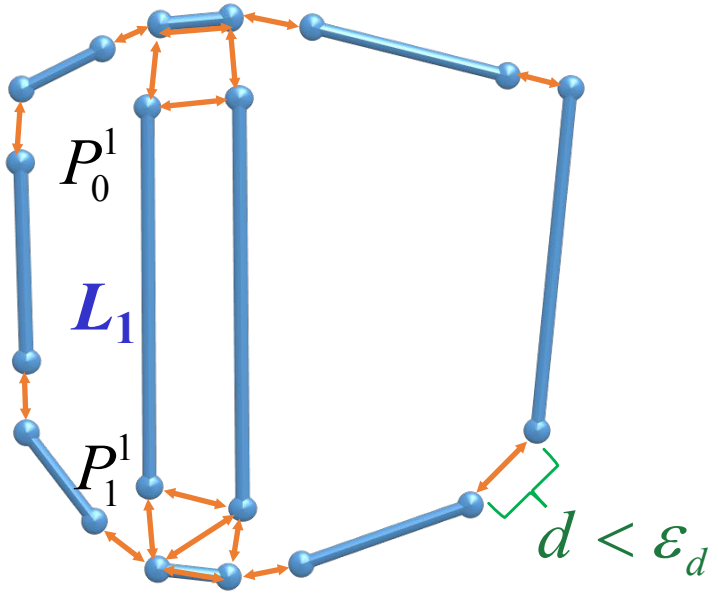
Workflow


Automatic phase
Interactive phase
View Selection
Automatic Joint 2D-3D Line Proposal
Surface Scaffold Structure Consolidation
Interactive Structure-driven Scaffolding
Iteratively

Surface Scaffold Structure



(a) Input mesh M and 3D linear feature set F

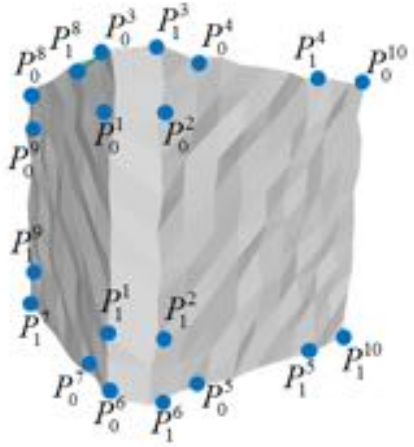


(b) Scaffold S

Scaffold Topology Representation

LPW relation types

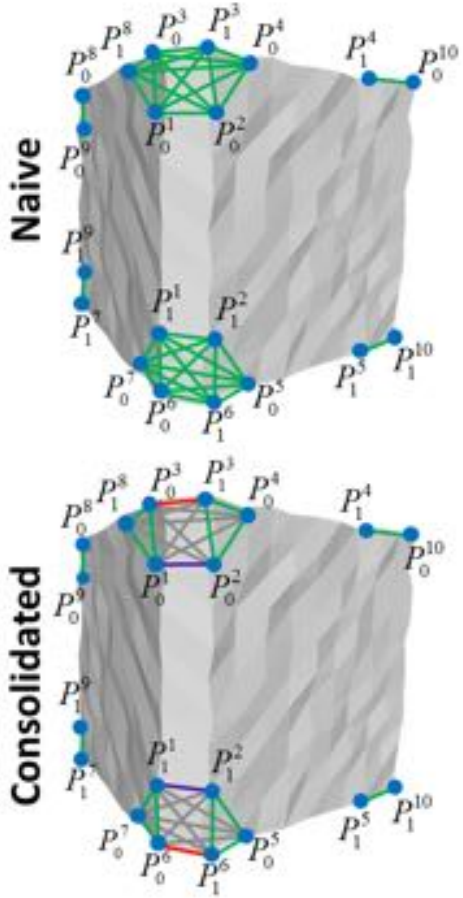
- *Contiguous* relation
- *Incident* relation
- *Parallel* relation
- *Other* relation



Scaffold Topology Representation

LPW relation types

- *Contiguous* relation
- *Incident* relation
- *Parallel* relation
- *Other* relation



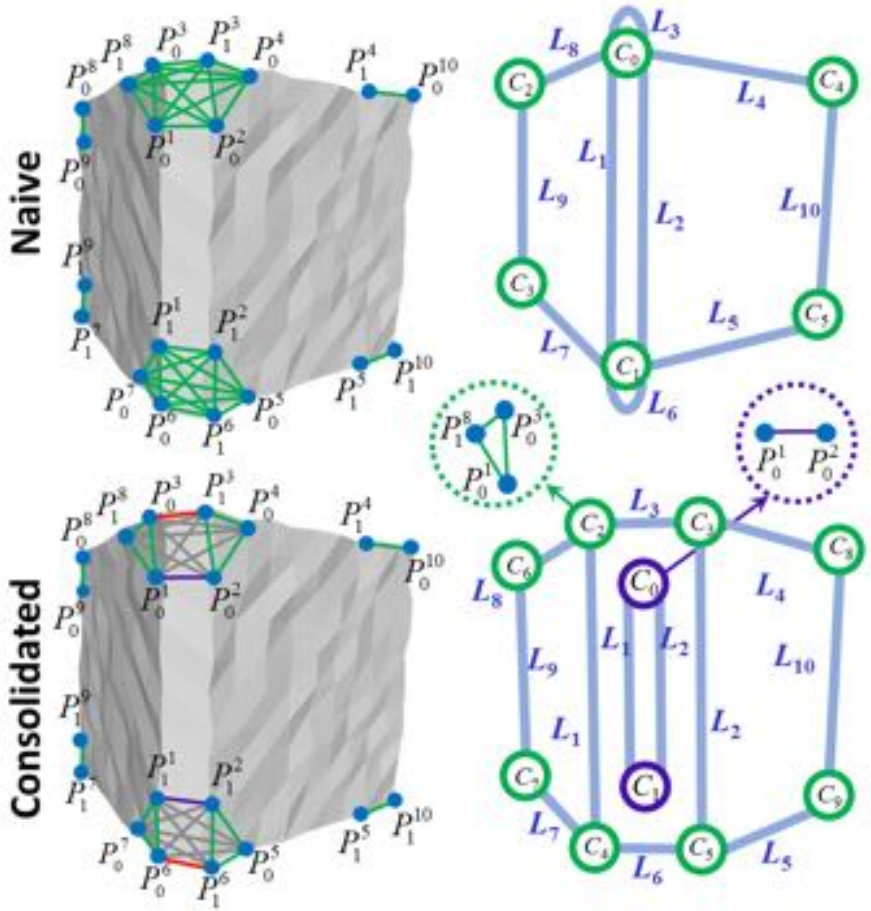
(a) Scaffold graph

Scaffold Topology Consolidation

LPW relation types

- *Contiguous* relation
- *Incident* relation
- *Parallel* relation
- *Other* relation

- *Incident* clique
- *Parallel* clique
- Interaction between relation cliques



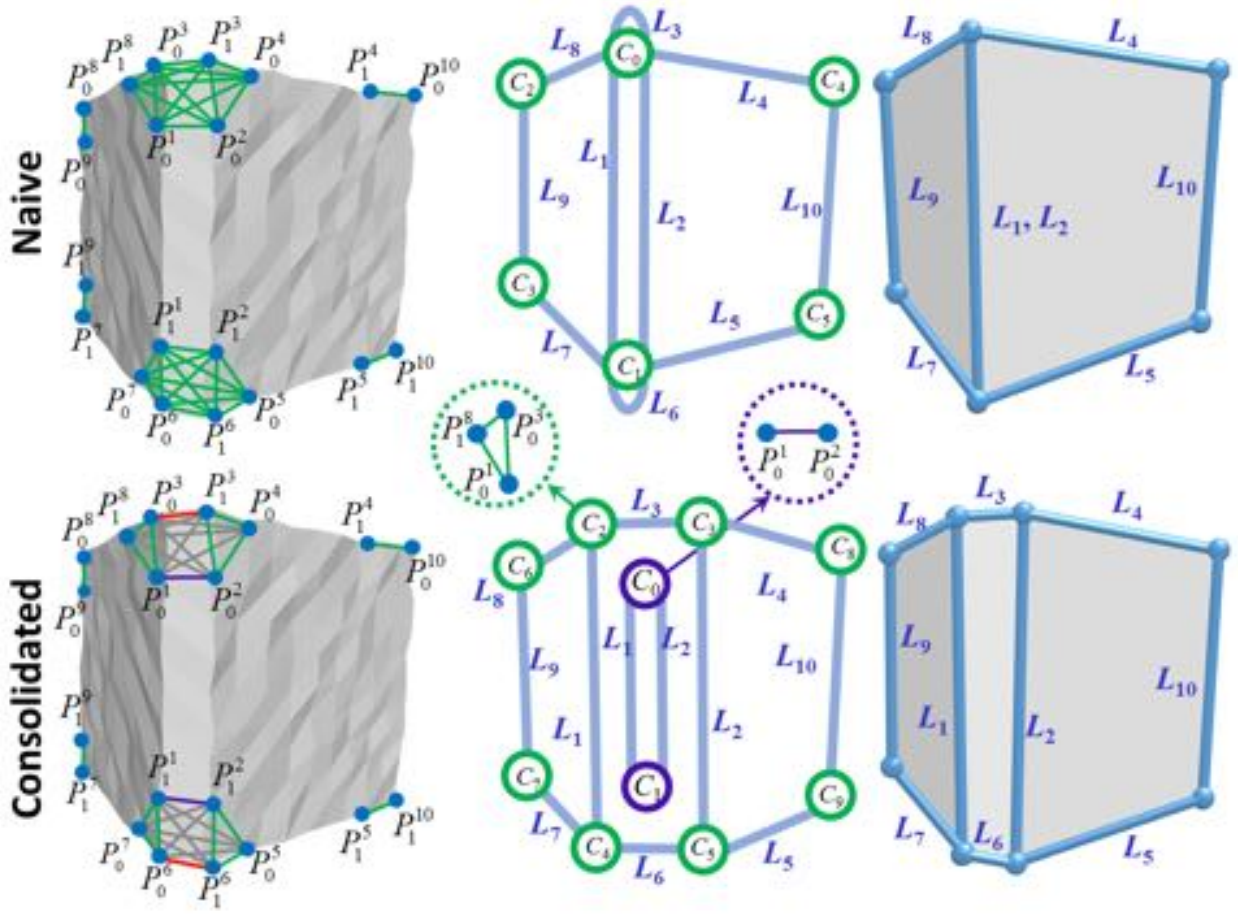
(a) Scaffold graph (b) Relation clique graph

Scaffold Topology Consolidation

LPW relation types

- *Contiguous* relation
- *Incident* relation
- *Parallel* relation
- *Other* relation

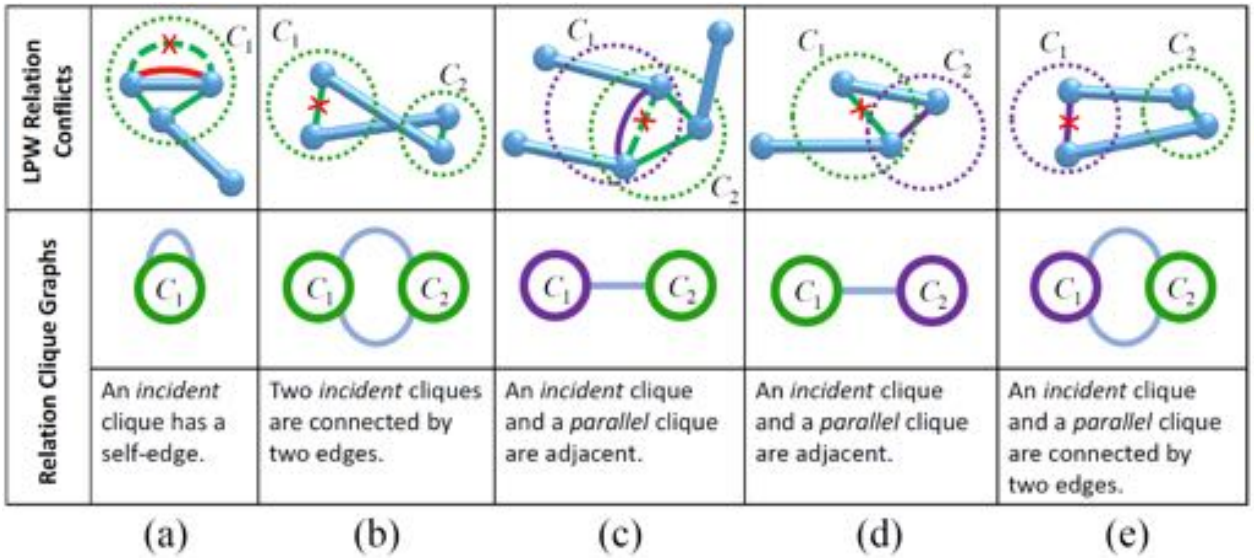
- *Incident* clique
- *Parallel* clique
- Interaction between relation cliques



(a) Scaffold graph (b) Relation clique graph (c) Recovered scaffold

Scaffold Topology Consolidation

- Relation conflict $C_i \bowtie C_j$



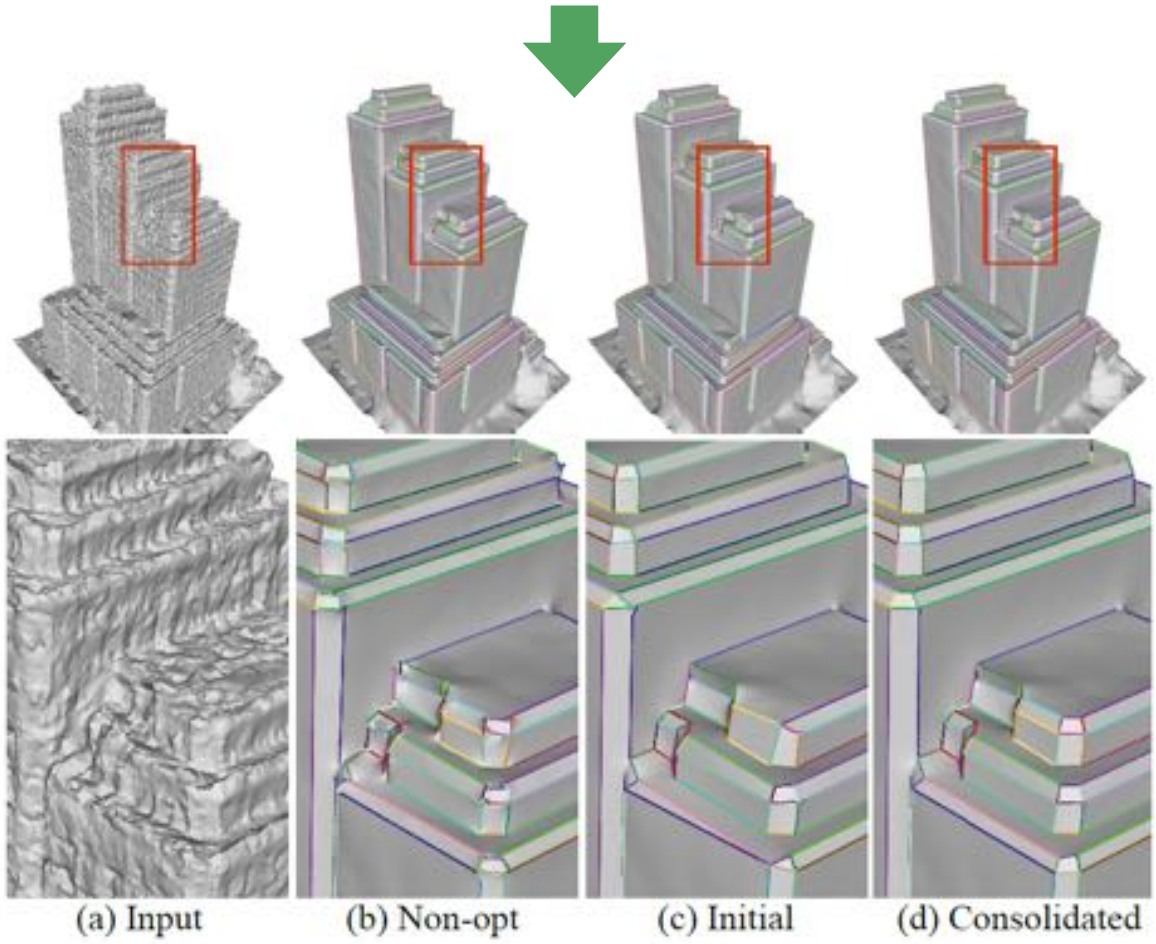
- Formulation

$$\mathbf{R}_s^* := \arg\min_{\mathbf{R}_s} \sum_{R_k \in \mathbf{R}_s} \psi(R_k) \qquad R_k \in \{y_C, y_I, y_P, y_O\}$$

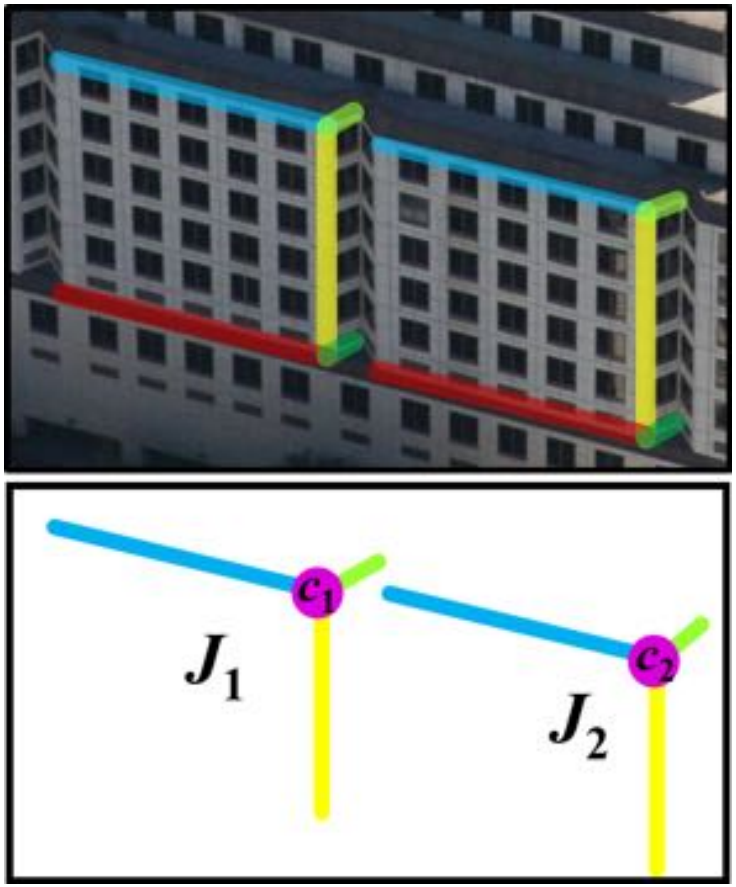
$$\text{s.t.}\ \Phi(C_i, C_j) = 0,\ \forall C_i \bowtie C_j$$

Global Shape Refinement

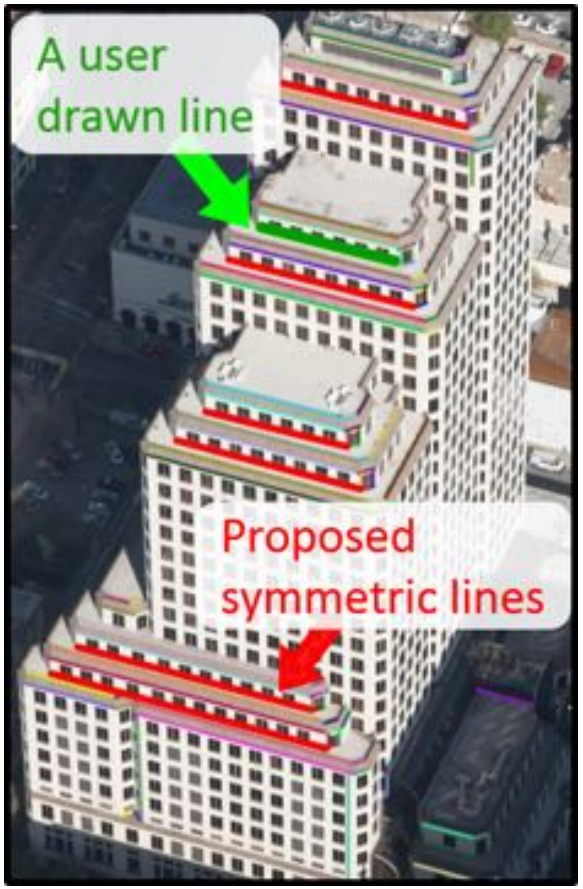
$$\min_{\{c_i\},\{\mathbf{v}_j\}} \Big(\sum_{C_i \in \mathcal{C}_I} \sum_{L \in C_i} \omega(L) dist(c_i, L) + \sum_{C_j \in \mathcal{C}_P} \sum_{L \in C_j} \omega(L) ||\mathbf{v}_j - \mathbf{v}_L||_2 \Big)$$



Interactive Line Proposal



(a) Symmetric joint structure



(b) An line proposal iteration

Result

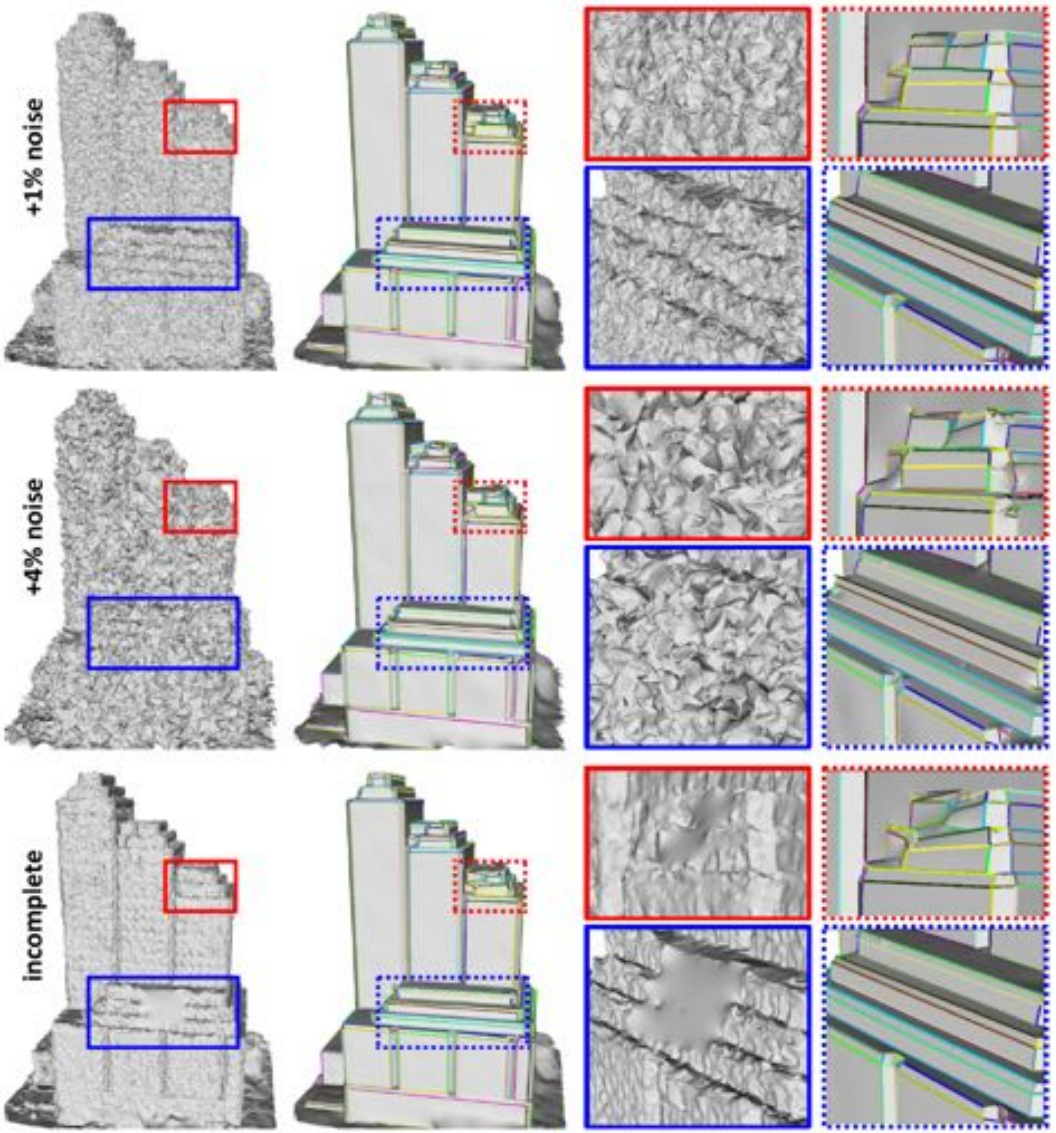
Output

Application - Simplification

Simplification Comparison

- QS: model directly simplified using QSLIM
- LS: model simplified using QSLIM after Laplacian smoothing
- BS: model simplified using QSLIM after Bilateral denoising
- Ours: model simplified using QSLIM after performing our method
- SR: simplification ratio
- #SF: number of faces in simplified model

Evaluation on Defect Laden Data



More Results

Dataset: monument

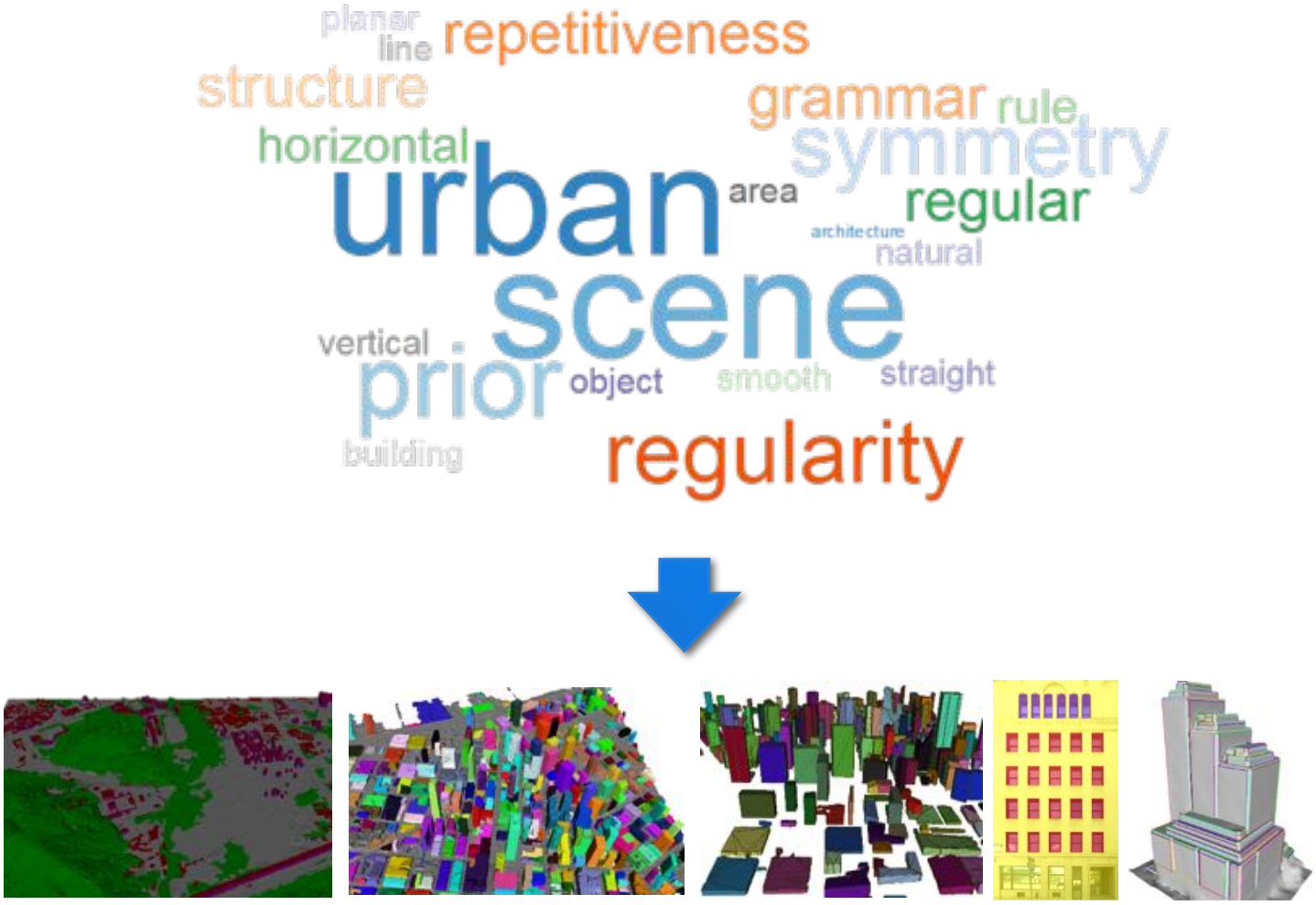
Contributions

- A novel scaffold structure
- A noise-insensitive scaffold structure consolidation method
- The interactive line proposal using urban priors

Outline

- Introduction
- Related work
- Approaches
- Conclusion and future work

Conclusion


planar
line
repetitiveness
structure
grammar
rule
horizontal
symmetry
urban
area
regular
architecture
natural
scene
vertical
prior
object
smooth
straight
building
regularity

Future Work

- Semantic segmentation
  - Deep learning techniques →deal with diverse datasets
  - Online data driven learning methods →adapt to crowd sourcing platforms
- Object recognition
  - Recognize individual tree models
  - Parsing facades in large-scale
- Remodeling
  - More complex primitives (quadratic features)

Publication

- ***Wang, J.*** *, Fang, T. , Su, Q. , Zhu, S. , Liu, J. , Cai, S. , Tai, C. , Quan, L.* Image-based Building Regularization Using Structural Linear Features. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 2015*
- ***Wang, J.****, Liu, C. , Shen, T., Quan, L.* Structure-driven Facade Parsing With Irregular Patterns. *Asian Conference on Pattern Recognition (ACPR) 2015. (**Oral**)*
- ***Wang, J.****, Li, S., Zhang, H., Quan, L.* Semantic Segmentation of Large-Scale Urban 3D Data with Low Annotation Cost. Computer Vision and Pattern Recognition (CVPR) Workshop 2015.
- *Liu, J.* ***Wang, J.****, Fang, T., Tai, C., Quan, L.* Higher-order CRF Structural Segmentation of 3D Reconstructed Surfaces. *International Conference on Computer Vision (ICCV) 2015.*
- *Zhang, H.,* ***Wang, J.****, Fang, T., Quan, L.* Joint Segmentation of Images and Scanned Point Cloud in Large-Scale Street Scenes with Low Annotation Cost. *IEEE Transactions On IMAGE PROCESSING 2014 (TIP) 2014.*
- *Zhang, H., Wang, J., Tan, P.,* ***Wang, J.****, Quan, L.* Learning CRFs for Image Parsing with Adaptive Subgradient Descent. *IEEE International Conference on Computer Vision (ICCV) 2013.*
- *Shen, T.,* ***Wang, J.****, Fang, T., Quan, L.* Color Correction for Image-Based Modeling in the Large. *Asian Conference on Computer Vision (ACCV) 2016. (To be appeared).*

References

- Verdie, Y., Lafarge, F., & Alliez, P. (2015). *LOD Generation for urban scenes*(No. EPFL-ARTICLE-206788). Association for Computing Machinery.
- Valentin, J., Sengupta, S., Warrell, J., Shahrokni, A., & Torr, P. (2013). Mesh based semantic modelling for indoor and outdoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2067-2074).
- Xiao, J., Fang, T., Zhao, P., Lhuillier, M., & Quan, L. (2009, December). Image-based street-side city modeling. In *ACM Transactions on Graphics (TOG)* (Vol. 28, No. 5, p. 114). ACM.
- Matei, B. C., Sawhney, H. S., Samarasekera, S., Kim, J., & Kumar, R. (2008, June). Building segmentation for densely built urban regions using aerial lidar data. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.
- Lafarge, F., & Mallet, C. (2012). Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International journal of computer vision*, *99*(1), 69-85.
- Lin, H., Gao, J., Zhou, Y., Lu, G., Ye, M., Zhang, C., ... & Yang, R. (2013). Semantic decomposition and reconstruction of residential scenes from lidar data. *ACM Transactions on Graphics (TOG)*, *32*(4), 66.

Thanks!
Questions?