A Robust Three-stage Approach To Large-Scale Urban Scene Recognition

Jinglu Wang[1] Yonghua Lu[2] Jingbo Liu[1] Long Quan[1]

The Hong Kong University of Science and Technology[1] Wuhan University[2]

Overview

- Introduction
- Joint semantic segmentation
- Building object segmentation
- Building object abstraction
- Conclusion

Introduction

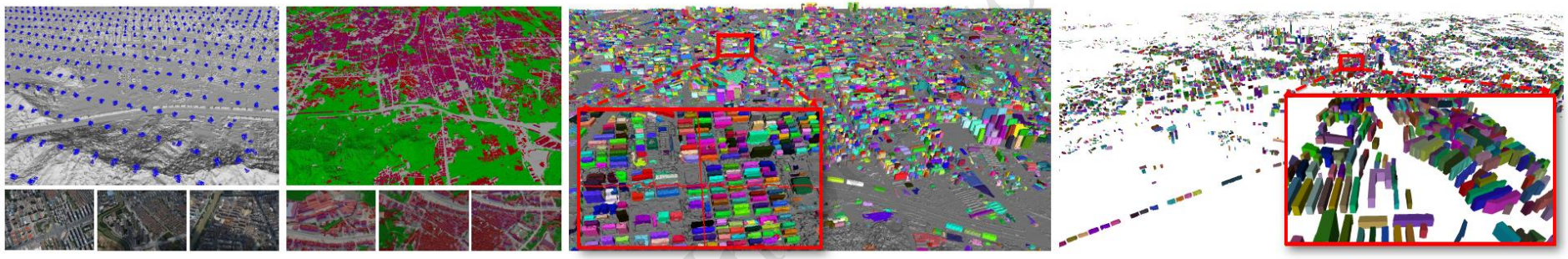
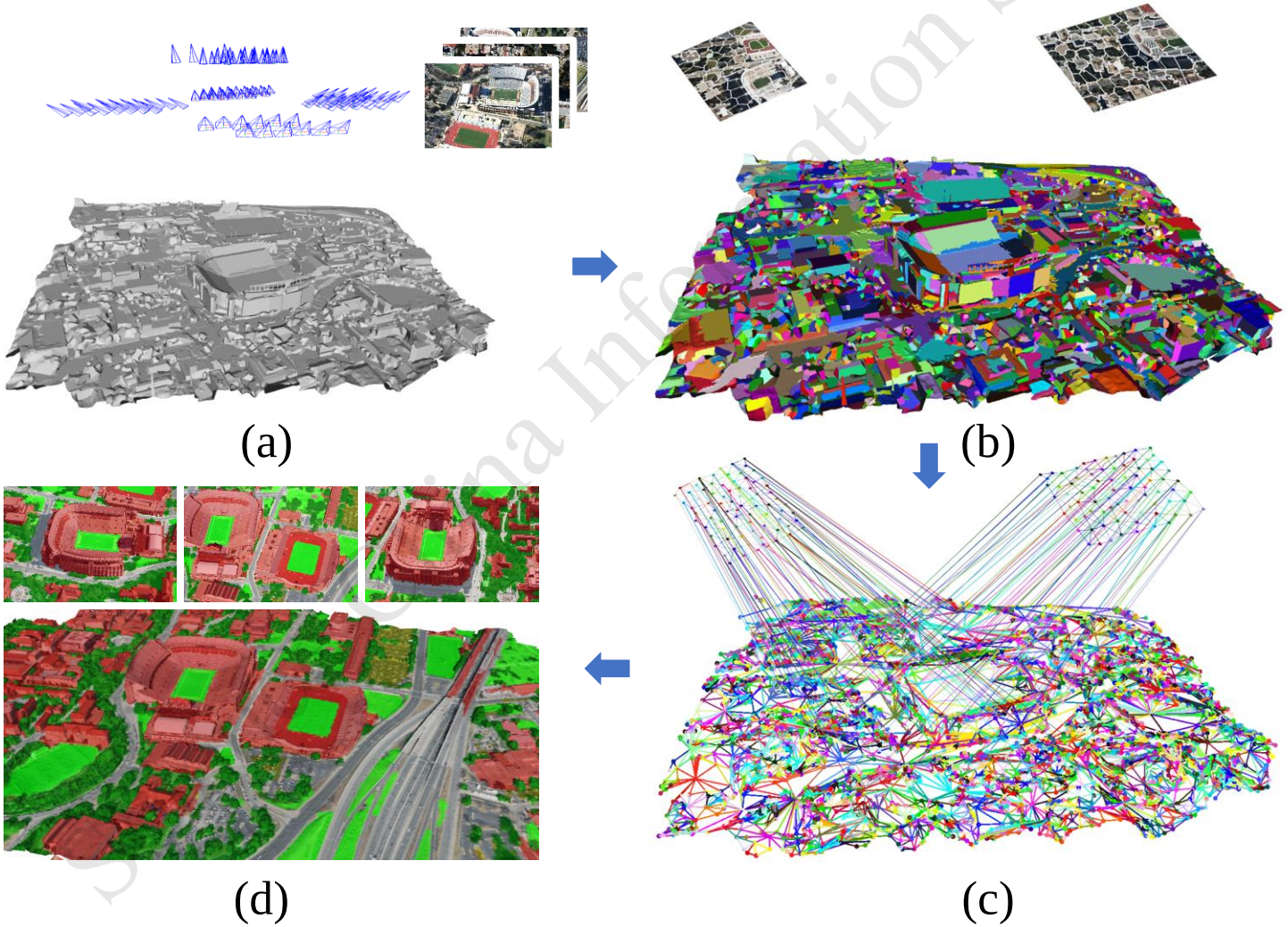


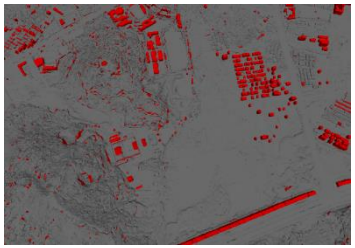
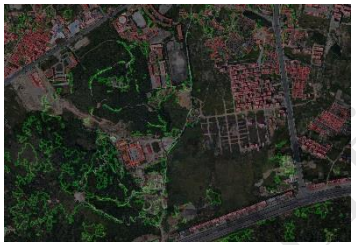
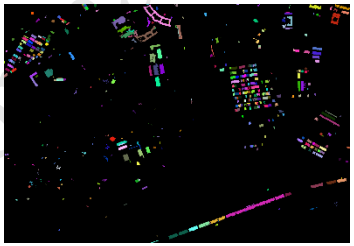
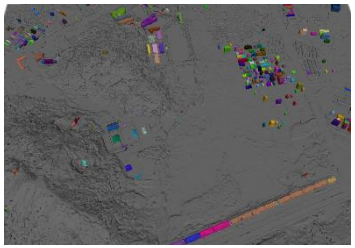
Figure 1: *The Three-stage Approach To Large-Scale Urban Scene Recognition. (a) Reconstructed mesh and input images with recovered camera poses. (b) Joint semantic segmentation. (c) Object segmentation. (d) Object Abstraction.*

Joint Semantic Segmentation



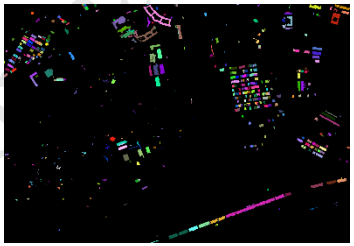
Building Object Segmentation

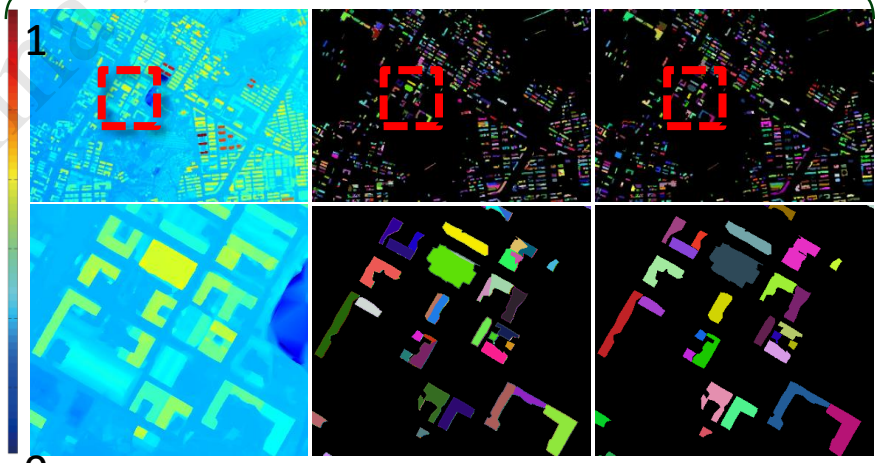
Terrain extraction → Urban object classification → 2D map consolidation → segmentation 2D to 3D

Building Object Segmentation

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

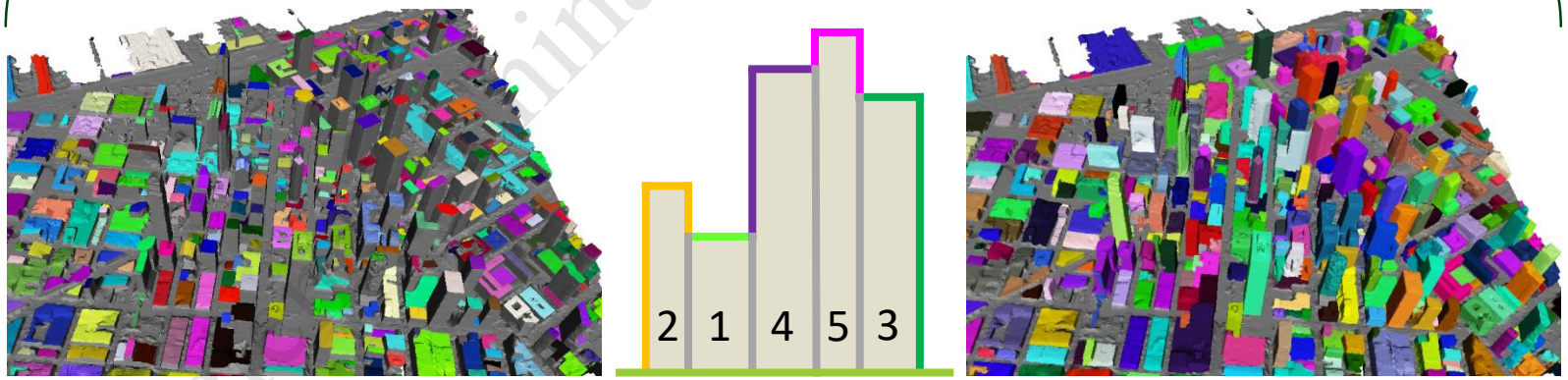




(a) Height map (b) Over-segmented map (c) Final segmented map

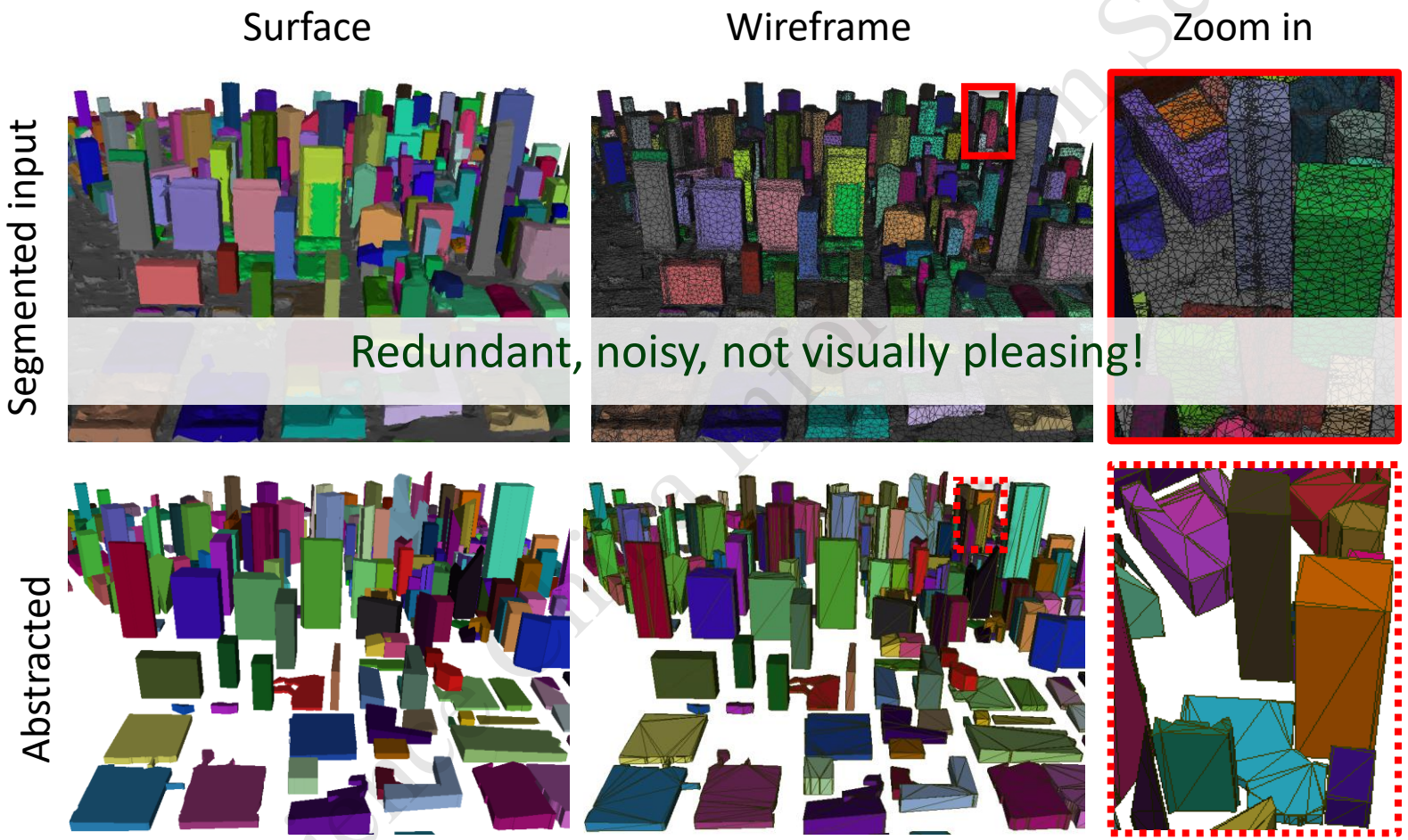
Building Object Segmentation

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

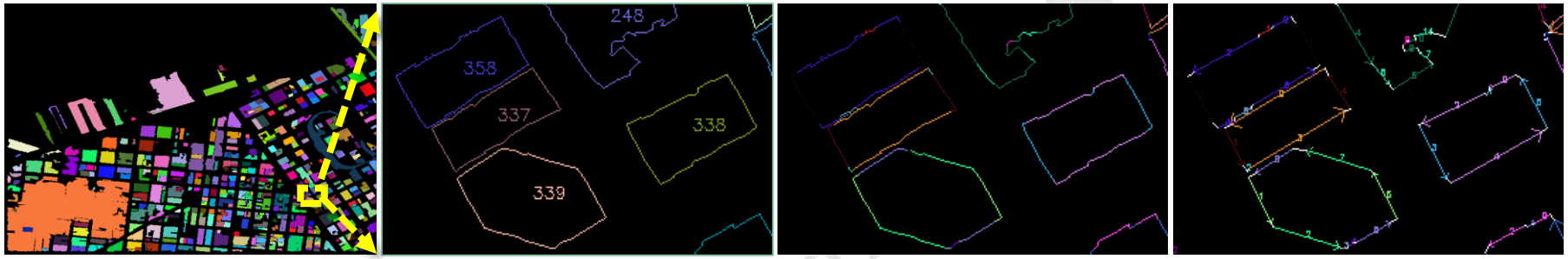


(a) Segmented roofs (b) Flooding illustration (c) Segmented objects

Building Abstraction Motivation



Contour Abstraction



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

- Higher order CRF formulation

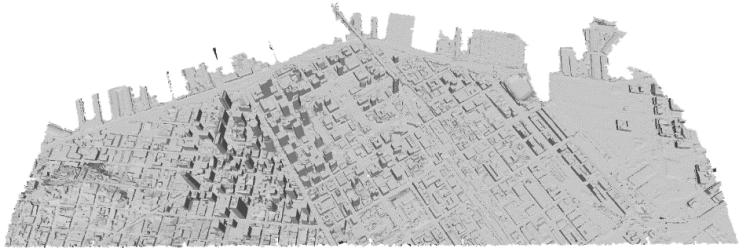
$$E = \sum_{x_i \in \mathcal{S}} \phi_i(y_i) + \sum_{x_j \in \mathcal{N}(x_i)} \psi_{ij}(y_i, y_j) + \sum_{c \in \mathcal{C}} \psi_R(\mathbf{y}_c)$$

Higher order regularity term

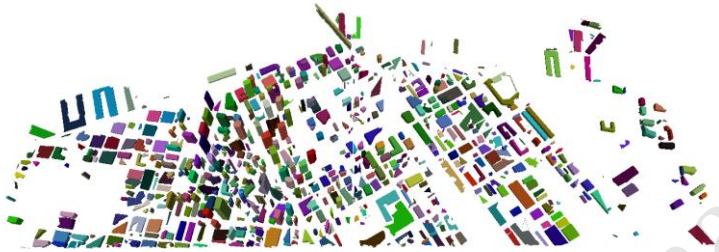
- Higher order regularity

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

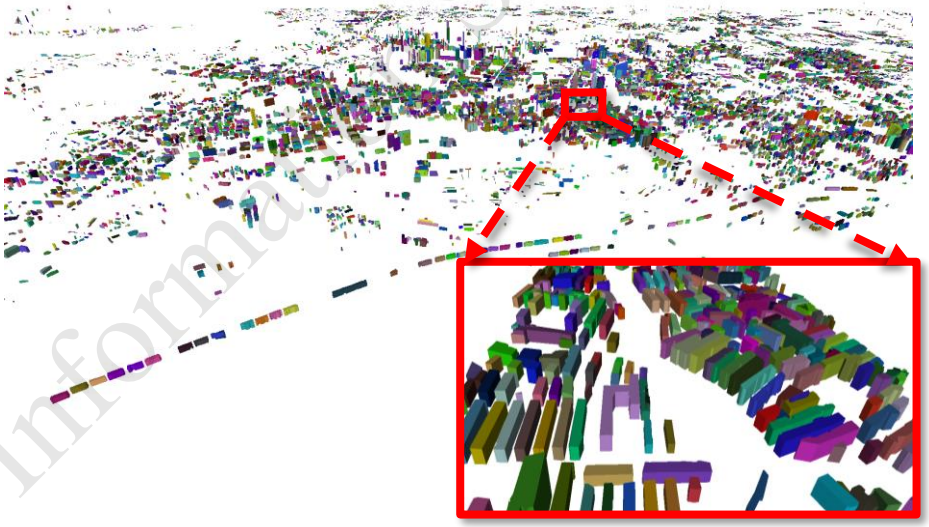
Results and Evaluation



(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

Conclusion

- A robust three-stage urban recognition approach able to deal with large scale urban reconstructed models.
- Exploit the multi-view consistency to impose labeling consistency between 2D images and 3D mesh by integrating the constraints in a CRF model.
- The building abstraction algorithm encodes the contextual information of the urban scene structure in the higher-order CRF formation.