3D Semantic Labeling of Photogrammetry Meshes Based on Active Learning

Mengqi Rong[1,2], Shuhan Shen[1,2], Zhanyi Hu[12]

[1] University of Chinese Academy of Sciences, Beijing, 100049, China
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

Session Title: Computer Vision Robotics and Intelligent Systems

Session Type: Poster

ICPR 2020 Virtual Conference

Content

- Introduction
- Method
    - ➢ Image Semantic Segmentation
    - ➢ 2D-3D Semantic Fusion
        - Back-Projection
        - Geometric Constraint
        - Energy Minimization
    - ➢ Active Learning Based Image Selection
        - Least-Scoring Subset/Uncertainty
        - Largest Coverage Area/Divergence
        - Annotation suggestion
- Experimental Results

Introduction

- The traditional geometry-based 3D reconstruction has reached a relatively mature stage, off-the-shelf commercial and open-source 3D reconstruction and photogrammetry softwares could help us generate large scale city models from massive aerial images captured by drones.
- Many scholars are not satisfied with just obtaining the structural information of the scene, and then focus on the expression and understanding of 3D scenes.
- There is no doubt that an urban model with richer information can be better applied to smart city, urban planning, virtual reality, autonomous driving and so on.

Introduction

- A straightforward strategy is to perform semantic segmentation directly on the 3D mesh or 3D point cloud, which usually needs to involve 3D deep neural network and so on.
  - the irregular, unstructure and orderless format of the point cloud
  - still cannot solve such large-scale 3D models well
  - the lack of large amounts of fine-labeled training data

- For the reason that the photometry meshes are obtained through image reconstruction, so apart from the 3D mesh models, there are also calibrated images. The second feasible method is to segment on those 2D images for pixel-wise semantic information, and then use the calibrated camera parameters to back-project the 2D segmentation results to the 3D mesh model and continue to fuse them together for a 3D semantic model.

Method

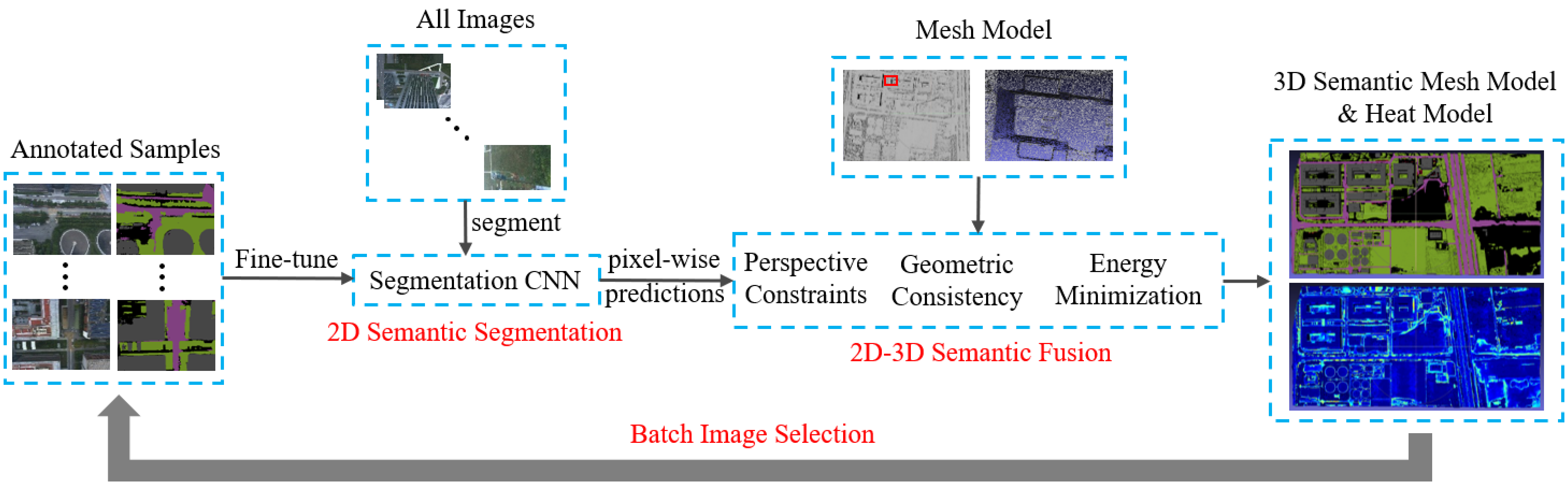
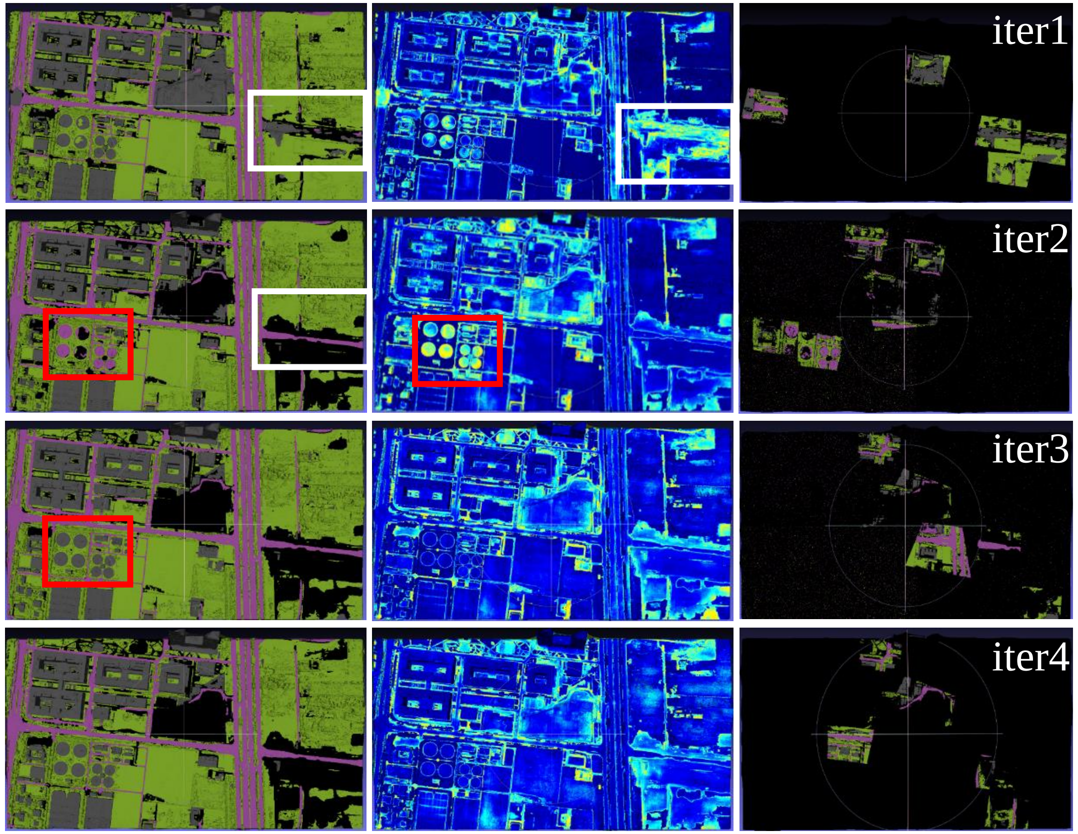


Fig.2 The pipeline consists of three main steps: fine-tuning the 2D semantic segmentation network with an ever-enlarging annotated image set, back-projecting the pixel-wise predictions onto 3D mesh model for semantic fusion based on geometric consistency, selecting a batch of images for annotation and adding them into the training set for the next iteration.

Experimental Results



(a) (b) (c)

Fig. 5. The experimental results of the proposed method on Urban1. (a) Semantic mesh models after MRF optimization. (b) Heat models corresponding to each semantic mesh model. (c) 5 images chosen according to the active learning based image selection.

Trend of some model parameters as the iterations increasing on Urban1

iterations	2D_Seg_Acc	Number / Percentage	3D_Seg_Acc
iter1	0.8894	----	0.7036
iter2	0.8886	770108 / 0.1569	0.8168
iter3	0.8633	317575 / 0.0647	0.8485
iter4	0.8351	164582 / 0.0335	0.8739