1. Introduction

In recent years, the traditional geometry-based 3D reconstruction has reached a relatively mature stage. 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. To deal with this problem, we propose a procedural approach for 3D semantic expression of urban scenes based on active learning. We first start with a small labeled image set to fine-tune a semantic segmentation network and then project its probability map onto a 3D mesh model for fusion, finally outputs a 3D semantic mesh model in which each facet has a semantic label and a heat model showing each facet’s confidence. Our key observation is that our algorithm is iterative, in each iteration, we use the output semantic model as a supervision to select several valuable images for annotation to co-participate in the fine-tuning for overall improvement. In this way, we reduce the workload of labeling but not the quality of 3D semantic model.

2. Overview

The optimized 3D semantic mesh model incorporates both 2D semantic segmentation and 3D geometry information, it could be used as a more reliable supervisor to measure the segmentation quality and help us to determine the next batch of training data for high quality performance.

A. least-Scoring Subset/Uncertainty

A straightforward strategy for finding the most valuable annotation areas is to use uncertainty samplings, which is obtained by re-projecting the label and its confidence of each facet onto different 2D images.

\[ u_p = \begin{cases} 1 - s_i, & p \cap F = f, \\ 0, & p \cap F = \emptyset \end{cases} \]

\[ U_{l_c} = \sum_{i \in I} \sum_{p \in I} u_p \]

B. largest Coverage Area/Divergence

The selected areas are expected to carry as many useful characteristics or features of the unannotated images as possible. We use the coverage area as another measure, that is, the ratio of the intersection of the visible facet by the selected images subset, as:

\[ \tilde{C}_{l_c} = \frac{\tilde{F}_i}{\tilde{F}_U} \]

3. Experimental Results

The regions that were assigned to incorrect semantic labels and accompanied by lower confidences would be selected by our annotation suggestion approach and revised in the next iteration. Four iterations later, the semantic mesh model reached a relatively stable level and the heat model also became smoother.

<table>
<thead>
<tr>
<th>iterations</th>
<th>2D. Seg. Acc</th>
<th>Number / Percentage</th>
<th>3D. Seg. Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>iter0</td>
<td>0.8894</td>
<td>0.7036</td>
<td></td>
</tr>
<tr>
<td>iter1</td>
<td>0.8808</td>
<td>770/108 / 0.1569</td>
<td>0.8168</td>
</tr>
<tr>
<td>iter2</td>
<td>0.8633</td>
<td>317/575 / 0.0647</td>
<td>0.8485</td>
</tr>
<tr>
<td>iter4</td>
<td>0.8351</td>
<td>165/882 / 0.0335</td>
<td>0.8739</td>
</tr>
</tbody>
</table>

4. Conclusion

We have presented a complete framework of semantic modeling from meshes of large urban scenes. Inspired by active learning, we make our algorithm iterative and propose an annotation suggestion algorithm for selecting the most effective data which would greatly improve the quality of semantic segmentation and then the 3D semantic mesh model. It demands limited human labor but not reduces the labeling quality of 3D models.

5. References

