Indoor-Outdoor 3D Reconstruction Alignment

European Conference on Computer Vision ECCV 2016 :
https://cvg.ethz.ch/research/indoor-outdoor/

Andrea Cohen , Johannes L. Sch¨onberger , Pablo Speciale, Torsten Sattler, Jan-Michael Frahm , Marc Pollefeys

ETH Z¨urich, UNC Chapel Hill, Microsoft

Introduction

A combined model would allow autonomous robots to easily transition between the indoor and outdoor world. However, state-of-the-art approaches often fail to reconstruct both parts into a single 3D model.

- Lack of visual overlap
- The change of lighting conditions between the two scenes.

Our approach leverages semantic information, specifically window detections, in multiple scenes to obtain candidate matches from which an alignment hypothesis can be computed.

Introduction

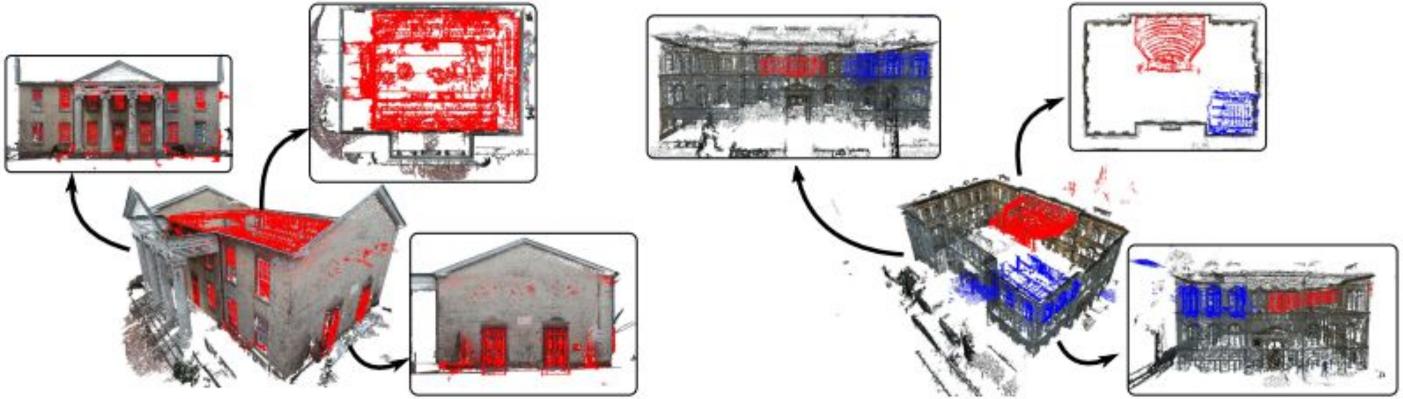


Fig. 1. The proposed method aligns disconnected Structure-from-Motion reconstructions of the inside and outside to produce a single 3D model of a building. Our approach also handles incomplete reconstructions and multiple indoor models (*c.f*. right model)

Introduction

They propose an alignment algorithm that exploits scene semantics to establish correspondences between indoor and outdoor models. More precisely, they exploit the fact that the windows of a building can be seen both from the inside and the outside.

⇒ Apply semantic classifiers to detect windows in the indoor and outdoor scenes

⇒ A single match between an indoor and outdoor window determines an alignment hypothesis (scale, rotation, translation) between the two models

⇒ Plausible alignments are then further refined using additional window matches

Related work

Trend of using higher level (semantic) information for both sparse and dense 3D reconstruction:

- Ceylan et al. Coupled Structure-from-Motion and 3D Symmetry Detection for Urban Facades. ACM Trans. Graphics (2013)
- Cohen et al. Discovering and Exploiting 3D Symmetries in Structure from Motion. In: CVPR (2012)
- H¨ane et al Joint 3D Scene Reconstruction and Class Segmentation. In: CVPR (2013)
- Savinov et al.
- Koch et al. indoor-outdoor alignment using line segment

Proposed method

Problem: Given separate indoor and outdoor models, propose a method to align the inside and outside of a building through semantic information.

⇒ Specifically, as windows are visible both from inside and outside, we use window detections to generate correspondences between the two models, which are then used to compute the alignment between the models.

Note: Door detection performs poorly

Proposed method

Window Detection:

- Apply a per-pixel classifier to detect windows in all input images using facade parsing
- Next, Use the known camera poses and the sparse 3D scene points to estimate the 3D planes containing the windows
- Leveraging the SfM points, Estimate 3D window positions for each image individually.
- Then detect overlapping 3D windows and compute consensus window positions

Proposed method

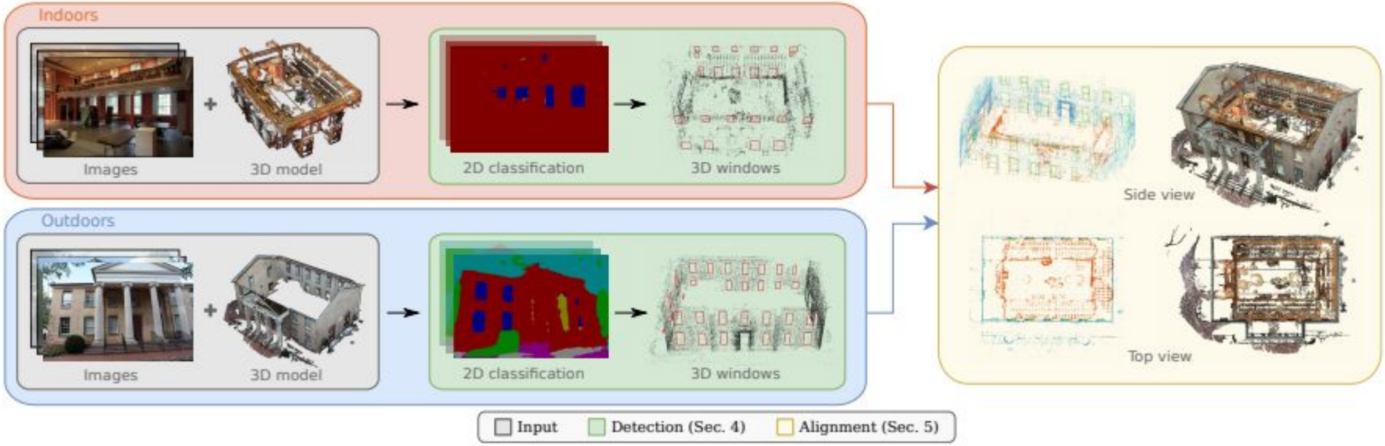


Fig. 2. Given SfM reconstructions of indoors and outdoors together with their input images, we leverage per-pixel semantic classification to detect windows in 3D. These windows are then used to compute a registration between both scenes that maximizes the number of aligned windows while avoiding that the models intersect each other

Proposed method

Window Detection:

- Image classification. They use the supervised learning method of Ladick´y et al. to obtain a pixel-wise semantic classification of the images used for reconstruction.
- a classifier trained on indoor images performs poorly on photos taken on the outside and vice-versa

⇒ Train two separate classifiers:

- For training the indoor classifier, use the annotated datasets provided by [19].
- To train the outdoor classifier, eTrims dataset.

Proposed method

Window Detection:

- Natural frame estimation. To simplify the subsequent steps of procedure, they align each 3D model into a canonical coordinate system.
- Using coordinate system that aligned to facade of the building

⇒ Calculate vanishing point for each image and vote for 3D coordinate

- The vertical axis is aligned with z
- Walls are mostly aligned with the x or y

Proposed method

Window Detection:

- Image rectified and facade parsing: search for window planes to those parallel to the x-z and y-z planes.
- Cohen et al.
- Extract the set W2D(i) of 2D windows for image i by obtaining the four corner vertices of the rectangles corresponding to window detections in the rectified image

Proposed method

Window Detection:

- Individual window projection: Let W2D(i) = {w i 1 , . . . , wi n} be the set of 2D windows detected in the previous step for image i, and let P i = {p i 0 , . . . , pi m} be the set of 3D SfM points that are visible in image i

⇒ Extract the subset P 0i ⊂ P i of points whose projections in the image fall inside any of the detected 2D windows w i j , j = 1, . . . , n.

⇒ Use P 0i to estimate the window plane π as the best fitting plane parallel to either the x-z or y-z plane.

⇒ Project all windows to plane π to obtain 3D set of windows

Proposed method

Window Detection:

- Window grouping and consensus: Windows on each image i can be overlap with other windows of image j.
- Given the sets of 3D windows W3D(i) detected for each individual image i, then group the overlapping 3D windows from all images into clusters C.
    - First, we cluster all 3D windows that overlap and are on the same plane by intersecting their areas
    - Two clusters Cs and Ct, s 6= t, are merged if there exist two overlapping windows Wi j ∈ Cs and Wk l ∈ Ct, i 6= k.
    - Compute a consensus window W(C) for each cluster C: maximum sum rectangular sub-matrix problem
    - The output is the set of consensus windows W3D = {W(C)} for each sub-model.

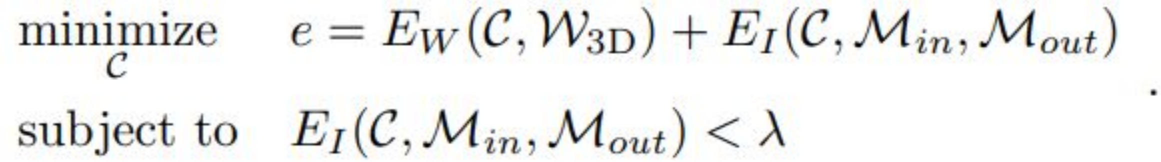
Proposed method

Model alignment:

- Objective: The goal of the alignment procedure is to transform the initially disjoint indoor and outdoor models into a common reference frame.
- Computing the alignment boils down to finding a similarity transformation between the models,
- However, the appearance of a window can change dramatically when viewed from the inside and the outside, e.g., due to illumination changes or by actually looking through the window

Proposed method

Model alignment:

- Input: The input to our alignment procedure are sets Min and Mout of axis-aligned indoor and outdoor models, consensus windows W3D set from previous step
- The output is a set of ranked configurations of aligned models Ks = {(Ci , ei) | ei < ei+1}, where the energy ei measures the cost of a configuration Ci and a lower energy denotes a better configuration
- 

$$\begin{aligned} \underset{\mathcal{C}}{\text{minimize}} \quad & e = E_W(\mathcal{C}, \mathcal{W}_{\text{3D}}) + E_I(\mathcal{C}, \mathcal{M}_{in}, \mathcal{M}_{out}) \\ \text{subject to} \quad & E_I(\mathcal{C}, \mathcal{M}_{in}, \mathcal{M}_{out}) < \lambda \end{aligned} .$$

Proposed method

Model alignment:

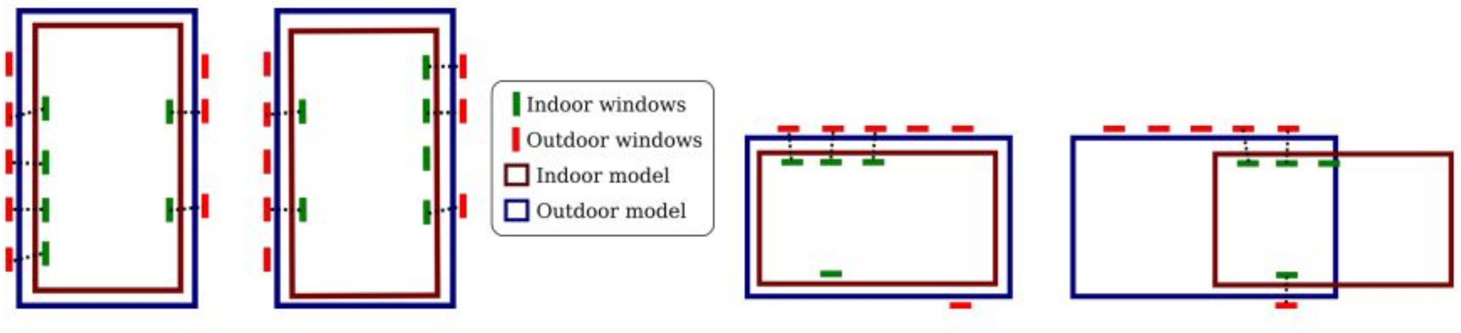


Fig. 4. (Left) Window term example. The alignment on the left has a lower cost than the one on the right. (Right) Intersection term example. Both alignments have the same E_W. The solution on the left is chosen since E_I is lower

Proposed method

Model alignment:

- Correspondence search: For correspondence search, we exhaustively explore all possible configurations Ci .
- We start by generating all unique pairwise window combinations between every unique pair of indoor and outdoor models.
- For each correspondence (Wa(mj), Wb(mk)), we estimate its associated similarity transformation Tjk from the four corresponding 3D window corners in Wa(mj) and Wb(mk).

Note: exploit the fact that the windows are already axis-aligned, i.e., the rotations around the x- and y-axes are already fixed.

Proposed method

Model alignment:

- Configuration evaluation: Given the set of unordered configurations, the next step is to determine whether they are plausible and to rank the plausible ones based on their quality.
- Propose the energy EW +EI to jointly model the quality of the window alignments EW (window term) and the amount of model intersection EI (intersection term)
- a good alignment explains as many window alignments as possible. They are number of windows that have correspondence.
-

$$E_W(\mathcal{C}_i, \mathcal{W}_{3\mathrm{D}}) = |\mathcal{W}_{3\mathrm{D}}| - 2 \cdot |\mathcal{C}_i| \quad .$$

Proposed method

Model alignment:

- Create a 3D voxel grid for each model spanning the entire reconstruction including cameras and points, using a resolution of 2003 voxels. A voxel is marked as free space if it is intersected by a viewing ray from one of the cameras to a sparse 3D point
- The intersection ratio γjk between two aligned models mj and mk is then defined as the fraction of the sparse 3D points in model mj that lie within a free-space voxel of mk
- $$E_I(\mathcal{C}_i) = \min\{1 - \epsilon, \max\{\gamma_{jk} \ \forall \ m_j \in \mathcal{C}_i, m_k \in \mathcal{C}\}\}.$$

Experimental Evaluation

- Datasets: collected a diverse set of six datasets, spanning the possible input scenarios of our approach: (1) single indoor and single outdoor model (Theatre, House-1, Chapel), (2) multiple indoor models and single outdoor model (University, House-2), and (3) single indoor and multiple outdoor models (Theatre-Split).
- All buildings have multiple floors and we use a state-of-the-art SfM pipeline
- Qualitative Evaluation:
- Quantitative Evaluation: a mean error of ≈ 0.05m (Theatre), ≈ 0.42m (University), ≈ 0.19m (Hall) and ≈ 0.54m (House-2, which results in an inaccurate alignment)
- Windows Evaluation: able to detect, on average, 73.9% of all indoor and 66% of all outdoor windows. Even for detection rates as low as 45%, our approach still works.

Experimental Evaluation