Semantic Alignment of LiDAR Data at City Scale

Fisher Yu, Jianxiong Xiao, Thomas Funkhouser
Princeton University

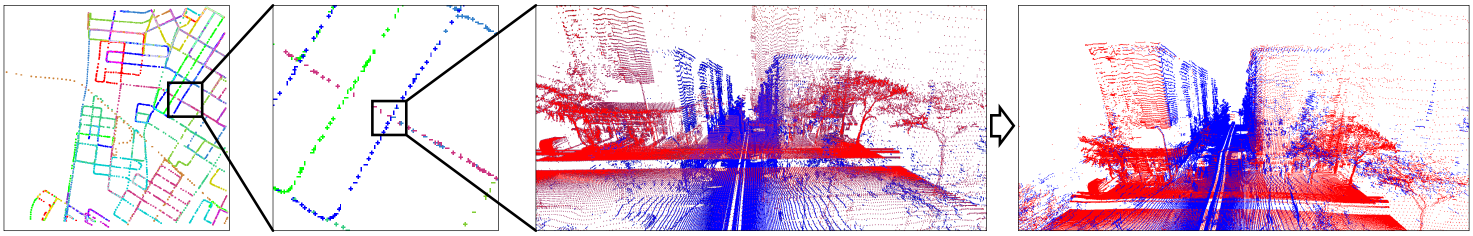

Figure 1: Our system aligns LiDAR captured by StreetView cars throughout a large area of New York City. The two images on the left show overhead maps of the car trajectories. The third image shows the original alignments provided by Google. The rightmost image shows our alignment. Different colors represent different scans of the same surfaces.

There has been a recent explosion in worldwide efforts to acquire 3D models of real-world urban environments. The motivation for these efforts is to provide 3D representations of cities that can be used for mapping, urban planning, virtual tourism, security analysis, and commerce applications. For most of these applications, the 3D model must be detailed, true-to-life, and globally consistent.

Acquiring this type of model has been a long-standing goal in computer vision, and much progress has been made towards it using structure-from-motion (SfM). For example, SfM has been used to reconstruct sparse 3D point clouds for tourist sites using collections of images found on the Internet, and it has been used to co-register Street View images captured from car-mounted camera over wide areas of a city. However, methods based on SfM are limited in accuracy, resolution, completeness, and robustness by the difficulties of finding point correspondences in images, which is particularly challenging in an urban environment with complex occlusions (e.g., trees), repeated structures (e.g., windows), specular reflectance (e.g., glass buildings), and wide baselines between images.

With this motivation, several companies routinely acquire LiDAR data from scanners mounted on cars driving systematically through cities. The problem is that the absolute point position acquired by the 3D scanners depends on the scanner poses predicted by GPS, inertial sensors, and SfM, which are notoriously inaccurate in urban environments. Different runs (continuous capture sessions) through the same region of a city can be misaligned by tens of meters due to inaccuracies of GPS, and sets of points acquired within any single run can suffer warps due to drifts in pose estimations. The misalignments are usually so serious that traditional point cloud registration methods such as Iterative Closest Points (ICP) fail to converge to correct solutions.

The goal of our paper is to investigate robust methods to perform registration of LiDAR scans collected from car mounted scanners in urban environments. Our approach is to align *semantic features* as shown in Figure 2. We detect semantic objects commonly found in cities (roads, facades, poles, cars, segments, etc.) to establish features within the LiDAR, and then we align those features across all scans simultaneously with an all-to-all ICP algorithm. This approach is advantageous in situations where data is noisy and seen from disparate views because it considers the entire shape of an object for feature detection, which can be recognized repeatably and distinctively for many types of objects in urban environments.

We demonstrate the value of semantic features in an ICP framework for large-scale alignment of LiDAR scans. Our variant of ICP leverages the fact that different semantic features are distinctive within neighborhoods of different sizes (scales). Therefore, it executes a coarse-to-fine refinement by selecting different semantic features to match at every step – i.e., first it successively aligns mutually closest roads, facades, and poles, which can be matched robustly even for gross initial misalignments. Then, it successively matches mutually closest cars and other small objects, which require better initial alignments to find correct correspondences. This multi-stage ICP

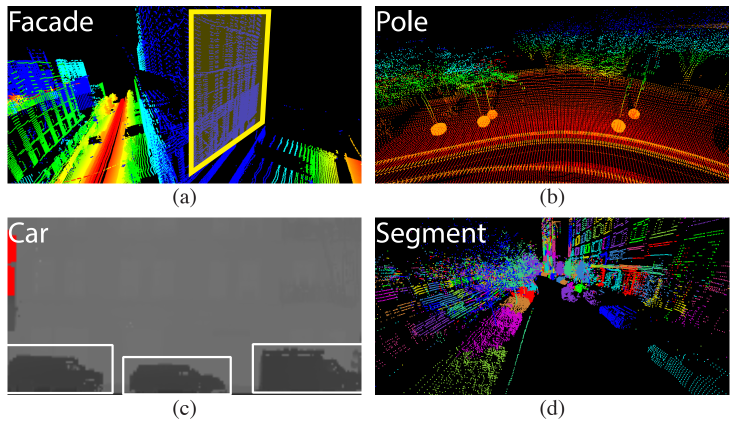

Figure 2: Examples of features for alignment. We extract semantic features such as building facades, poles, cars and objects proposed by segmentation

algorithm can globally align many scans across a large area simultaneously.

The main contribution of the paper is the idea that semantic features based on object detectors can be used effectively for alignment of LiDAR data. Secondary contributions include a method for introducing increasingly finer-scale features in successive stages of an ICP algorithm, and a framework for aligning features of many different types in a non-rigid, all-to-all global registration of many Street View LiDAR scans covering large regions of a city. We manually labeled point correspondences covering all the intersections in our data to evaluate the alignment and compare different methods as shown in Figure 3. Our experimental results indicate that the proposed methods can achieve significantly better alignments than those provided by alternative methods for four different cities.

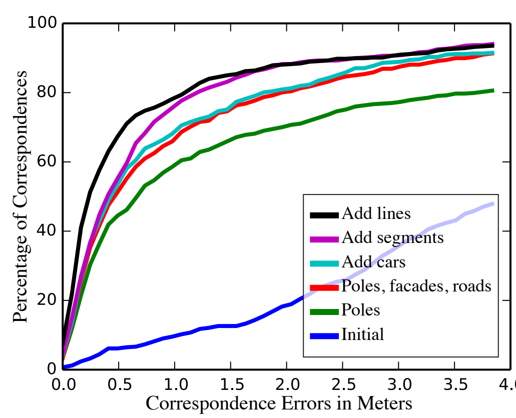

Figure 3: Plots of the percentage of ground truth features (vertical axis) aligned within different distance thresholds (horizontal axis) using different combinations of features (different curves) during tests on scans of New York. Higher curves are better.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.