Online 3D Object Mesh Pose Recognition

Mike Hagenow (mhagenow 9079441078) and Kevin Welsh (kpwelsh 9081472293)

February 14, 2020

Abstract

Many of the existing methods for 3D object pose recognition from point clouds have deficits that make them not ideal for a aviation manufacturing environment. In this work, we propose an extension to efficient RANSAC methods that directly recognizes irregular object meshes by evaluating a novel set of mesh face-based features. We will demonstrate that our method can recognize real objects in a point cloud captured by a Microsoft Kinect. Finally, we also will provide systematic comparisons to two state of the art methods: a neural net approach and a irregular geometry RANSAC method.

1 Motivation

There are many instances in which 3D object recognition may be a critical capability. One compelling recent example has been the autonomous driving market where there is a need to make split-second decisions based on deducing candidate objects from a combination of visual sources. In other autonomous industries, such as robotics, it may be further required to recognize and accurately estimate object poses to determine task semantics and plan low-level motion including grasp strategies, path planning, and motion characteristics.

We are interested in this problem as a stepping stone for enhanced human robot interaction, since humans would prefer to communicate semantically (e.g., "bring me that screwdriver") rather than specifying task parameters such as the location of the object and how it should be grasped. The ability to provide semantic instructions requires a system that can recognize and precisely locate scene objects. We are interested in a solution that works well in a manufacturing settings, where accurate 3D models of manipulated objects are known ahead of time, which we refer to as a semi-structured environment. Notably, these environments use hardware (e.g., bolts, rivets) and surfaces (e.g. aluminum, composites) that have little color variation, which makes geometric inference an attractive approach.

2 Related Work

Estimating object and pose recognition from a point cloud is a area of large recent interest. We provide a partial review of relevant, recent work to appropriately contextualize the key ideas of our method. Given that point clouds often contain large amounts of noise, it is common to use robust fitting methods in determining scene geometry. Many methods (e.g., [4][1]) use a modified version of RANSAC, where primitive geometry is estimated based on evaluating random minimal sets of points against continuous surfaces (e.g., planes, spheres, cones, etc). In a few methods, the general RANSAC method is extended to arbitrary mesh geometry. For example, Papazov et al. [2] developed a method which sample an arbitrary mesh as a point cloud and uses two-point geometric correspondences to determine where a mesh is located in the scene.

Other recent methods use neural network approaches to fit an arbitrary mesh. Methods commonly use a convolutional neural network (CNN) approach. For example, PointNet++ [3] can identify objects in an scene by first segmenting the scene and then classifying each of the segments through a hierarchical network struture. In contrast to the manufacturing environment described above, neural network approaches are often employed to allow for variance in the specification of objects.

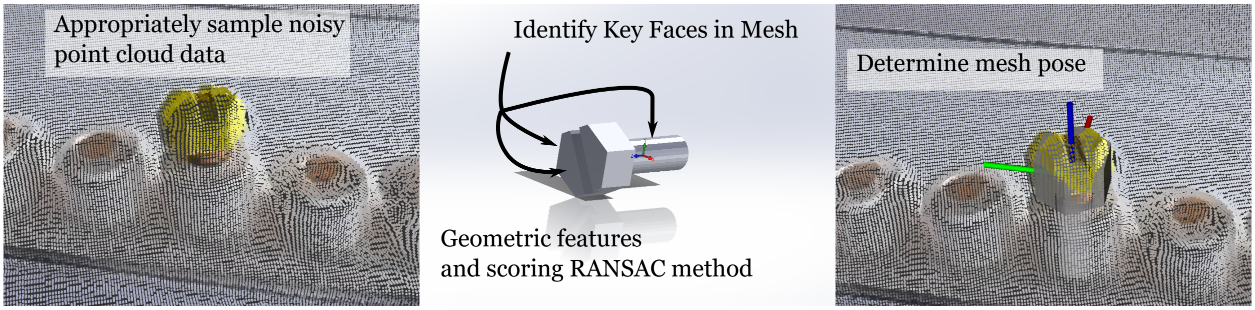

Figure 1: General problem statement to fit an arbitrary mesh to noisy point cloud data. Our method will describe geometric features and scoring to find likely objects and poses in the point cloud.

3 Proposed methods

Many state of the art algorithms for 3D object pose detection suffer from at least one significant drawback for application to a manufacturing setting. Neural network based methods are often highly effective, but they offer very few guarantees and critically depend on accurate training on relevant data to function. This is something that cannot yet be reliably performed by general consumers of the technology. Algorithms for feature based registration fail when the environment has many similar features. Manufacturing settings often have low color gradients and repeated geometry, resulting in erroneous correspondences. Offline solutions clearly don't facilitate effective human computer interactions, or responsiveness to the environment. Lastly, many RANSAC algorithms have a limited set of geometries , which results in an accurately segmented scene that is of little semantic use without further shape composition.

In our work, we will explore fitting an arbitrary mesh using geometric correspondences, most similar to the work of Papazov et al. [2]. Our method is proposed as an extension of Efficient Ransac and will differ from existing methods in two main ways. First, we plan to greatly reduce the candidate set of geometric features on the mesh to allow for faster calculations, by optimizing mesh face combinations based on higher order combinations of faces. Second, we plan to provide geometric features that define face correspondences as opposed to correspondence of points on a mesh that is the used is often used as the mesh sampling strategy. Based on this key difference, we refer to this method as *Face-Based-Features RANSAC* (FBF-RANSAC). For our application, we define the following requirements:

- Does not require retraining on local data.
- Robust to repeated geometries and color patterns.
- Calculate scene objects for a specified region of interest in less than 2 seconds (maximum 10 seconds).
- Can identify multiple small scale (<1 meter) meshes with arbitrary geometry.

Our method aims to generalize to semantically relevant objects, minimize the number of false positives, and operate in a sparse scene. We will do that by leveraging fundamental geometric information of the meshes and extracting unique identifiable features for quick correspondence checking. This method is targeted for use in conjunction with a visual user interface, and as such, we will rely on the human to localize the object search in a potentially much larger scene.

4 Proposed Evaluation

To evaluate our method, we propose to first demonstrate functionality on the intended use case: a semi-structured environment where 3D models of key objects are known and color information provides little differentiation. We follow-up by comparing to two state-of-the-art object pose recognition methods.

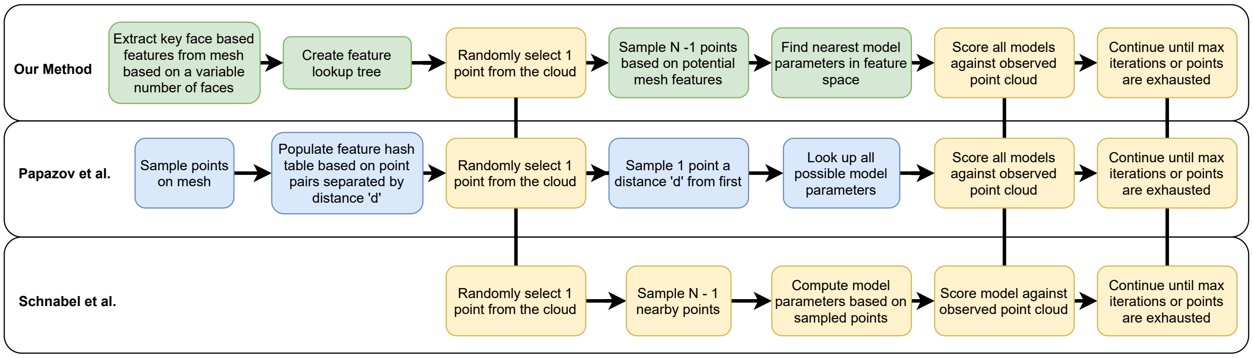

Figure 2: Our method attempts to iterate again on the efficient RANSAC algorithm by improving the way we identify possible mesh configurations by using an optimized set of features.

4.1 Evaluating efficacy in manufacturing use case

We have developed a small prototype manufacturing environment where a robot can interact with several toy tools, including bolts and a screwdriver. All interactions are recorded by a Kinect camera. We will perform regression tests that cross-validates our method for a variety of object configurations, sets, and repeated geometries.

4.2 Comparison to state of the art methods

We desire to show that our method can perform similarly to state of the art geometric feature-based RANSAC methods. We also desire to show that our algorithm can also perform in other situations where there are high-performing state of the art methods. As our algorithm performs scene matching on intuitive mesh features, we hope that it will be easier to implement in a new setting than PointNet++ while maintaining acceptable performance. We plan to use the following metrics for comparison:

- Accuracy of pose estimation compared to ground truth.
  - Across varying rates of occlusion.
  - Across varying noise intensities.
- Likelihood of false negatives.
- Likelihood of false positives.

4.2.1 Rigid 3D Geometry Matching

We intend to leverage the existing implementation by Papazov et al. [2]. The existing library works best on an environment where all of the objects in the scene are estimation targets, and when there is a single viewpoint. We will directly compare it throughout several different scenarios: the example scenes that they provide, the target lab scene from a single view point, the target lab scene from multiple view points, and simulated scenes that allow us to modulate the occlusion rates, noise, and sparsity of identifiable objects.

4.2.2 PointNet++

We will leverage the existing implementation and training data for analysis. This method supports both classification of entire point clouds to identify which object it is, and semantic segmentation of large point clouds to support the classification. Because of the architecture of the network, we will compare each of the modalities separately on fabricated scenes. This allows us to avoid the complex retraining step of PointNet++ to fit our local objects, while still comparing its core feature set.

5 Deliverables and Timeline

In order to implement our method, we will first develop and test a form of an efficient RANSAC method. We will then develop our method, FBF-RANSAC. For our method, we aim to have a well-engineered prototype that can be shared and implemented by other groups. Since implementations are not readily available in this area, we want to make an easily usable implementation that can benefit other research groups exploring this area. All code will be hosted on Github. Finally, for comparison, we will also set up and properly tune PointNet++ and the method described in Papanov et al.

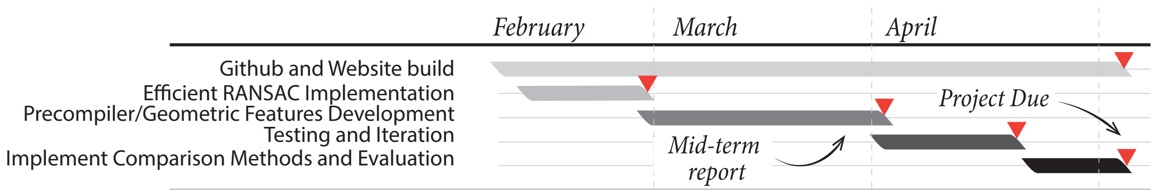

Figure 3: Tentative project timeline.

References

[1] Liangliang Nan and Peter Wonka. Polyfit: Polygonal surface reconstruction from point clouds. 07 2017. doi: 10.1109/ICCV.2017.258.

[2] Chavdar Papazov, Sami Haddadin, Sven Parusel, Kai Krieger, and Darius Burschka. Rigid 3d geometry matching for grasping of known objects in cluttered scenes. *The International Journal of Robotics Research*, 31(4):538–553, 2012. doi: 10.1177/0278364911436019. URL `https://doi.org/10.1177/0278364911436019`.

[3] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[4] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient ransac for point-cloud shape detection. *Comput. Graph. Forum*, 26:214–226, 06 2007. doi: 10.1111/j.1467-8659.2007.01016.x.