Automated Three Dimensional Object Scanning

Arnav Jain, Ashay Lokhande, and Brahma S. Pavse
arnav_jain@utexas.edu
ashay.j.lokhande@gmail.com
brahmasp@utexas.edu
University of Texas at Austin
Computer Science

1. Abstract

This papers hovers over the idea of optimizing a methodology to extract the most data from an object via various camera viewpoints in the least amount of cost to make the object "meaningful" to the robot. At a high level, we plan to collect all camera views for a given object and after developing our optimization algorithm, acquire a subset of these views that enhances the meaning and give us the most information about the object. Ideally, we would like to add functionality to the robot that will give it the ability to correctly and efficiently reconstruct an object that it recently viewed. We see a wide array of potential uses after this development, such as associating the identified object with various actions that can be performed on it. Towards the end of this paper, we explore in detail some future implementations following the development of this system.

2. Introduction

Three dimensional scanning is a necessary feature for robots to eventually navigate through and work within the human world. With three dimensional reconstruction a robot would be able to better interact with and manipulate the environment around it. Our goal is to create an automated procedure to find the most efficient set of camera angles of an object to create an accurate three dimensional representation of that object. Apart from the practical applications of this optimization procedure, which we will introduce soon, there is a tremendous amount of cost saved via the efficiency achieved; here we consider cost to be the robot energy and time that is expended.

A great example of the tremendous benefits that this optimization can result is a military situation. Consider such a situation with regards to national defense and security. Most war / battle situations that occur in an area desolate of resources (living and nonliving) and usable equipment. In this type of emergency situation, where this is a great lack of assistance, it is imperative for the robot to have the ability to correctly and efficiently identify objects around it and determine how to use any given item. In the long term, we would like to address both those issues. However, we are initially attempting to resolve the first issue: ability to correctly and efficiently identify objects. With an optimized manner of "perfect" object identification, in times of emergency, the robot will be able to identify the objects around it quickly and efficiently. Another advantage of utilizing images from relevant camera angles is that it is scalable to the teaching the robot to identify objects that appear similar to any given object. Given the situation of the miliary example, we now have a robot that given the algorithm developed has the ability to identify or make an educated guess of what the non-usable equipment is; the task is done in a timely manner, which is an essential component of emergencies, as well as efficiently, by reducing unnecessary movements of the robot.

3. Related Work

In this type of work of optimizing object recognition and reconstruction, there are several efforts made by researchers in this area.

A paper that is not only related to our goal, but could also help us develop a solution is titled "Point feature extraction on 3D range scans taking into account object boundaries" by Steder, Rusu, Konolige, and Burgard [1].

The focus of this paper is to present a keypoint extraction method on three dimensional point cloud data for both object recognition and pose identification. This can be used to identify interest points which can then be used to compare three dimensional images in a scan. The authors' goals were to find the similarities in two different images in order to understand which parts of the images overlap. This allows one to determine what has already been scanned and help distinguishes new data from data that the robot has already taken into account. In the experiment, the paper discusses using single range scans, as obtained with three dimensional laser range finders, where the data is incomplete and dependent on a viewpoint. In the paper they present a normal aligned radial feature (NARF), "a novel interest point extraction method together with a feature descriptor for points in three dimensional range data". The NARF relies on detecting the borders of an object and having objects placed in locations where the surfaces are stable.

The interest point detection relies on three ideas to accurately detect interest points in a three dimensional image. It must take borders and surface structure into account, select points that are reliably detectable from different angles, and the points must be in positions that provide stable areas for normal estimation. The goals for the development of the NARF Descriptor was to identify the difference in occupied and free space, to make the descriptor robust in handling different interest point positions, and enable them to extract a unique local coordinate frame at a single point. Once this procedure and its calculations were explained the paper introduces two different experiments that test object matching and stability in recognizing interest points from different distances and angles.

Another existing work that we read about and has some relevance to the development of our project is "Real-time object classification in three dimensional point clouds using point feature histograms" by Himmelsbach, Leuttel and Wuensche [2]. In this paper, the motivation that inspired the authors to pursue the given task was to assist a robot to navigate urban traffic as well as off road scenarios. In order to achieve this goal, the authors combined two dimensional and three dimensional image processing techniques. Two dimensional data was used for segmentation of point clouds into objects and three dimensional data was used as raw point clouds to classify objects. The paper further goes over fast object feature extraction which is captured by histograms over point features. The uniqueness of the paper is that it implements a way to incorporate two dimensional and three dimensional methodologies to solve the problem as opposed to using a single type of dimension. This paper assists for future purposes in that it accounts for very large point cloud sizes as much as 10^5 measurements.

4. Technical Approach

4.1 Gathering Data

Initially we will be running simulations in Rviz in order for us to gather test data on multiple objects of different sizes. In this project we will be gathering data in two different forms. In order to reach our main objective of creating a general camera planning procedure we must observe multiple objects each of different shapes and sizes. Based on their optimal solutions and our procedures for finding those solutions, we should be able to determine a general solution. Thus, the main source of data is our trials with different objects in the simulator. However, to create this data we must gather data on the different views for each object. For each object we must develop an optimal view set. In order to do this we must store and work through all the possible views for each object.

The very first step to achieving our ultimate goal is to create an environment in Rviz that simulates a room with the robot, a table, and the object we are currently testing. Each object will be imported into Rviz from Stanford's ShapeNet. ShapeNet is a database full of three dimensional models of real objects (see figure 1.1). Each object we choose will act as sample data for our eventual general solution. We will start off with smaller objects, but then continue on to examining larger objects in order to expand our trial data. Removing this size bias will allow our general planning procedure to better autonomously create three dimensional reconstructions of the objects.

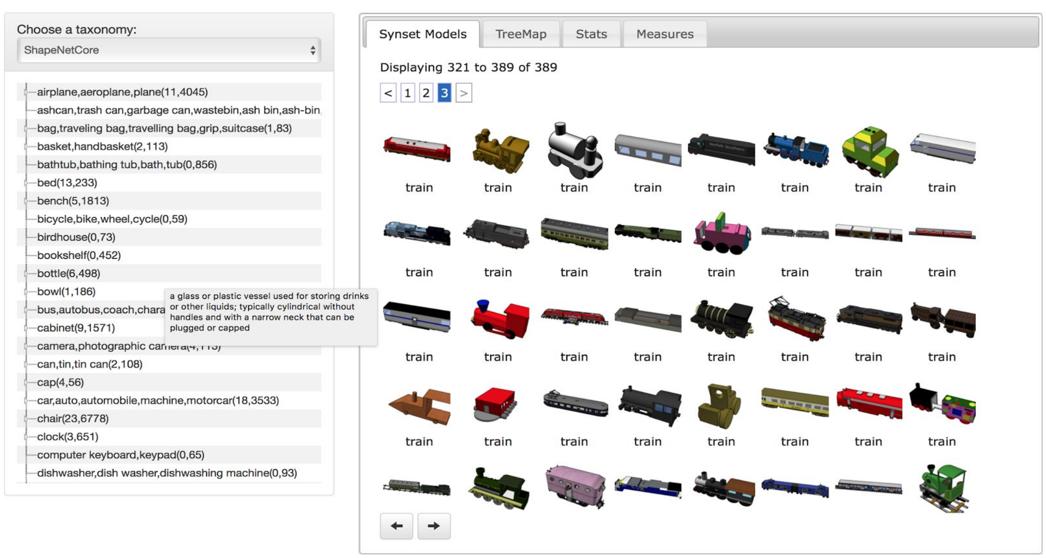

Fig. 1.1. A screenshot of the ShapeNet database containing complete three dimensional objects that may be imported into programs such as RVIZ

For each object we will be collecting data by generating planes around the object. For each plane we will look at the points on the object that face the plane. Thus, each plane will act as a representation of a camera's orientation (see figure 1.2). Initially we will look at generating planes completely around the object. Once we have mostly optimized our method of finding the most efficient set of planes, we will start testing the solution on situations where we limit the areas at which the planes can be generated. By preventing planes from being generated in certain areas, we can remove camera angles that some robots would not be able to reach.

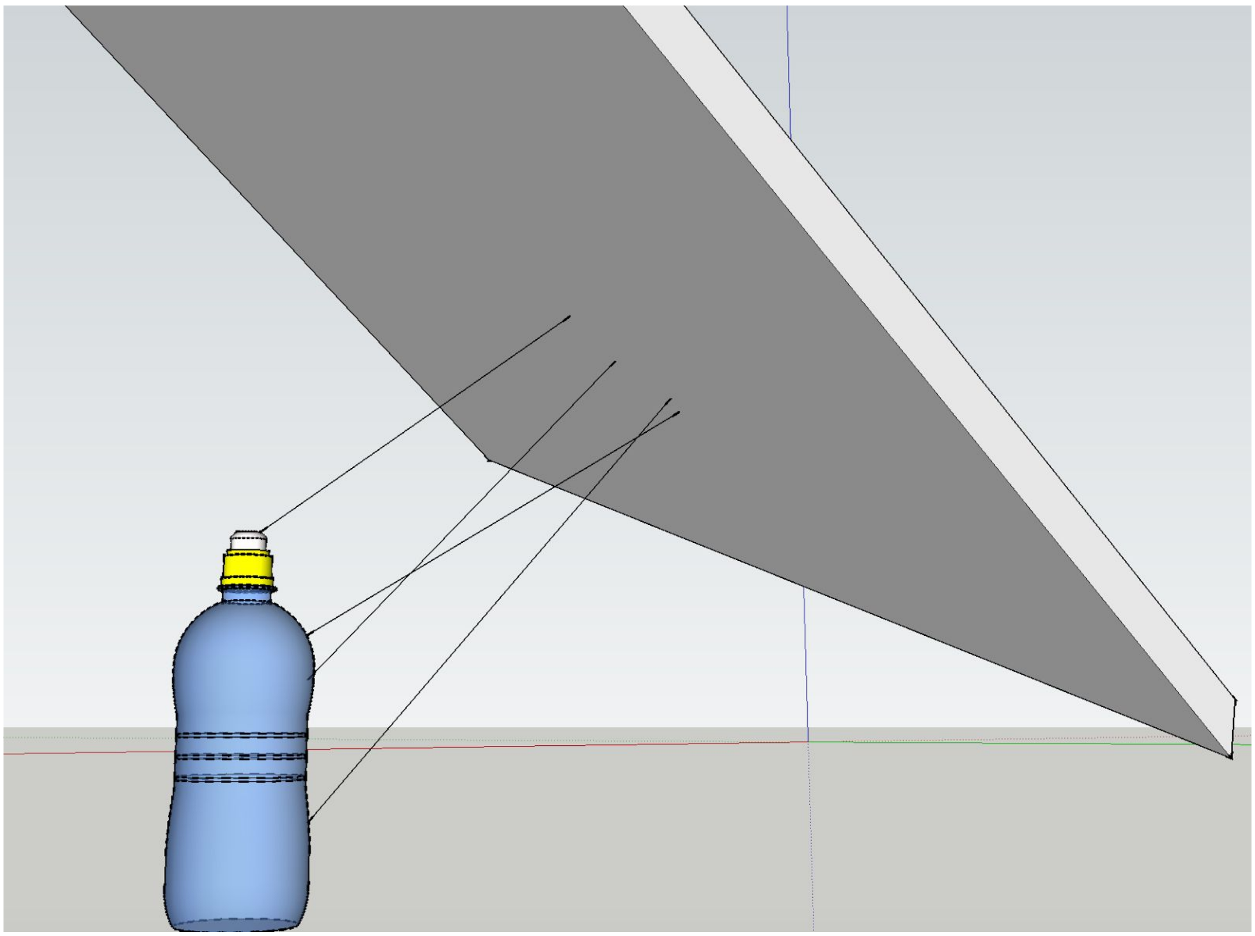

Fig. 1.2. A diagram made to illustrate what the planes look like in reference to the object as well as depicting the points on the bottle that would be identified as parts of the object visible from the plane's orientation.

4.2 Data Analysis and Optimization

Once we have generated the views for each object, we must develop a method to determine the optimal set of camera angles that creates the best reconstruction of the object with the least cost. This statement itself must be broken down into three components, the first of which is: defining what constitutes the "most best views" of an object. As far as our applications are concerned, we will define the “best views” as those that produce the most visual data from the object. Next, we must determine our method of attributing costs to each view. Each view will be given a cost based on its coordinate and rotational distance from the other views in the set. Finally we must tackle the procedure in which to find the optimal set of views for scanning the object.

The overall method of determining the optimal set of views will be based on the cost of each view as well as how much visual data on the object the set of view produces. While much of this portion of the project has yet to be decided, we will generally develop a method of creating a smaller set of views to optimize from. We would be eliminating views that provide relatively less information and cannot be reached by the robot. From here we would ideally divide the planes into groups based on their position relative to the object. For example, objects that are to the right of the object from the original view may be grouped together. We would then create a method of creating sets of views based on how much visual data on the object the produce. Within these sets we would attribute the cost based on the sum of rotational and coordinate distance of each view from each other. Thus we are trying to find the set of camera angles that creates the best reconstruction of the object with the least movement.

4.3 Evaluation

Since our goal is to achieve an idea of “How should I look at this object to get the most meaning”, we must develop a way to compare “meaning” collected by the robot to the actual “meaning” of the object.

A manner in which we will go about this is in the following fashion: consider a real life object that is similar to an object in our database of objects - from ShapeNet. After we perform our algorithm on the object to find the best the camera angles that extract the most meaning / data of the object, we will store the set of planes / camera angles that have been deduced by the algorithm. Given the set of these images, we will attempt to reconstruct the object that was just looked at. The reconstructed object is then compared to the similar object from ShapeNet. We plan to conduct this process of evaluation by importing the shape in question from ShapeNet into the simulated environment of Rviz along with the reconstructed object from our algorithm. Once both are in the environment together, we desire to formulate a procedure or sequence of steps to follow that will be able to make point to point (or plane to plane) comparisons between both the items. Given some form of metric such as means of percentage, we are hopeful the numeric value will be high to conclude the experiment as a success.

Walking through the point to point (or plane to plane) comparison will be as follows. Of course, when looking at both items under speculation we must look at them at the same angle of view for the same reason if consider bit equality we look at one bit at a time at a given index. So for a given angle view, we plan to compare factors such as color shade, shine angle of light etc. Of course this methodology presents some levels of restriction and limitation of scalability to test the algorithm under more cumbersome circumstances. For that given reason, we considered this in a more simplified manner and will accordingly expand as we develop the system.

These type of comparison will present some numeric value which we will use to account for difference and incongruence between two of the items under consideration. With the numeric values, as mentioned earlier, we can construct a percentage that will assist us to determine the accuracy of the algorithm and solution to this problem.

5. Future Work / Extensions

Our project provides the base for numerous applications that we see can be beneficial and worth researching.

A primary research topic is providing the robot with the capability to decide all possible actions or procedures that it can implement on an object. To go more into detail, if we were to successfully develop a system that could accomplish our goal, we would have a robot that could, given an object, be able to decide “what is the best way of looking at this object to get the most meaning?” Once the robot is able to find the “meaningfulness” of an object, we think that it is imperative that a robot know what are all the actions that it can perform on the given object. For example, when we, as humans, look at an object that is similar to a bottle, we start to think of the key characteristics of the bottle and once it has been correctly identified we associate the bottle to certain set of actions such as drinking from it, or storing liquid in it etc. This idea of mapping an object to a set of actions is crucial to develop a sense of understanding of objects around the robot.

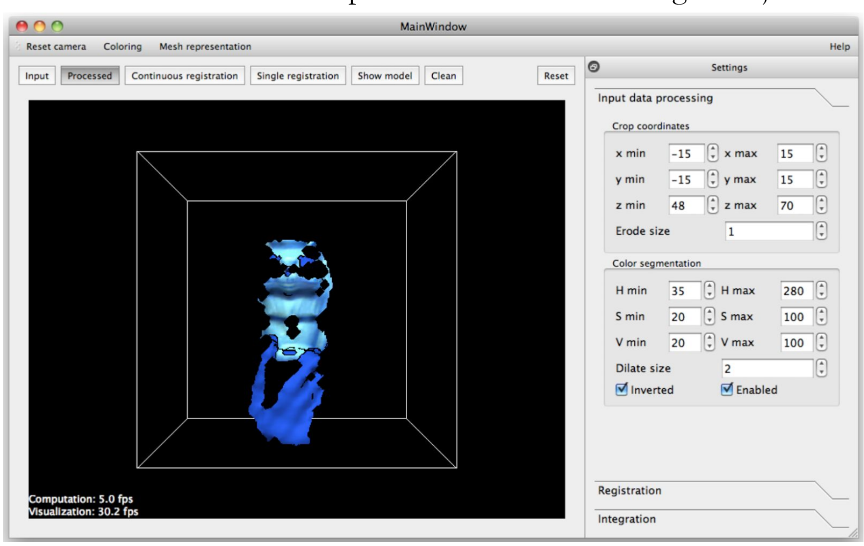

Fig. 2. An example of a three dimensional scanning suite. This particular code base incorporates the need to turn the object in order to capture the data covered by the arm holding it.

6. Conclusion

The main focus of our project is create a method of finding a set of camera angles that are the most efficient towards creating a three dimensional reconstruction of the object. We hope to find a general solution that would allow us to apply the resulting method on objects with dimensions we did not previously have, as well as being able to limit the camera angle detection to include angles at which the robot using it can actually look at the object. While a majority of our project will be spent working on creating a general solution within the simulator, we hope to apply our work into a demo. There are a plethora of extensions that could benefit from our ultimate solution. Our primary goal was once to use this procedure to develop a method of intelligently selecting a scanning procedure that best fits the object that the robot is looking at. Another extension that we are hoping to be able to start working on is to extend the representation of the model to program potential actions linked to common objects. Ultimately our project focuses on optimizing the fundamental procedure of objects scanning in order to allow developers to more efficiently work with creating and using three dimensional models.

7. References

1. B. Stedar, R. Rasu and K. Konolige, “Point feature extraction on 3D range scans taking into account object boundaries”, *Robotics and Automation (ICRA), 2011 IEEE International Conference* , 2011.
2. M. Himmelsbach, T. Luettel and H. Wuencsche, “Real-time object classification in 3D point clouds using point feature histograms”, *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference* , 2009.