Visualization and Labeling of Point Clouds in Virtual Reality

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="https://dx.doi.org/10.1109/34.121791">https://dx.doi.org/10.1109/34.121791</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Association for Computing Machinery</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Nov 24 08:52:37 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/117394">http://hdl.handle.net/1721.1/117394</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Visualization and Labeling of Point Clouds in Virtual Reality

Jonathan Dyssel Stets
Technical University of Denmark

Yongbin Sun
Massachusetts Institute of Technology

Wiley Corning
MIT Media Lab

Scott Greenwald
MIT Media Lab

Figure 1: Screen capture from the Virtual Reality application. A point cloud of jugglers is displayed, the left and right images shows the point cloud before and after the juggling pins has been annotated by the user.

ACM Reference Format:

1 INTRODUCTION

We present a Virtual Reality (VR) application for labeling and handling point cloud data sets. A series of room-scale point clouds are recorded as a video sequence using a Microsoft Kinect. The data can be played and paused, and frames can be skipped just like in a video player. The user can walk around and inspect the data while it is playing or paused. Using the tracked hand-held controller, the user can select and label individual parts of the point cloud. The points are highlighted with a color when they are labeled. With a tracking algorithm, the labeled points can be tracked from frame to frame to ease the labeling process. Our sample data is an RGB point cloud recording of two people juggling with pins. Here, the user can select and label, for example, the juggler pins as shown in Figure 1. Each juggler pin is labeled with various colors to indicate different labels.

2 MOTIVATION

We consider the use case of viewing animated point cloud data in VR. Although individual frames appear grainy and low-resolution, when they are animated, they produce a lifelike effect that is suitable for representing dynamic real-world scenes. However, without any labeling or metadata, it is difficult to edit or modify these scenes. Labeling a person with a name, or changing the color of an object would be examples of basic editing tasks. Segmenting the point clouds in each frame, and establishing how the segments in each frame correspond with those in surrounding frames, makes it possible to perform such tasks.

Furthermore, the same set of user interface affordances can be used for other use cases as well. The demand for large labeled point cloud data sets is increasing as the need for classification algorithms for such data is gaining popularity, especially for data-driven approaches like deep neural networks. Navigation of self driving cars, social robots and robot interaction in general often rely on 3D data that needs to be classified or segmented. Additionally, the use of 3D sensors is increasing, while these sensors are getting more mobile, better and more available. This enables the potential for more and better 3D point cloud data. The website Semantic3D [Semantic3d 2016] provides examples of large labeled 3D point cloud data sets. Oftentimes, such 3D data sets are labeled manually using a 2D computer monitor and a computer mouse or similar input device. We argue that a two-dimensional human-computer interaction with three-dimensional data can be inefficient in many cases. For example, the labeling problem is more difficult with a 2D interface when the orthogonal x\text{y}, x\text{z}, or y\text{z} planes do not separate the points to be labeled from the others surrounding them.
We implemented two algorithms to fit labeled points. The first one is an Iterative Closest Point (ICP) algorithm [Besl et al. 1992], which iteratively detects the closest point in frame$_{i+1}$ labeled point in frame$_i$ and estimates the current best rotation and translation to update the positions of the labeled points. This iterative process terminates until a predetermined stop criteria satisfies, for example, the spatial difference between labeled points and matched points converges. The second algorithm follows the same pipeline as the first one, but is augmented with color support to improve tracking accuracy. Instead of detecting the spatially closest point, the second algorithm selects, for each query point, the point with the most similar color from k-nearest neighboring points as the match. We build a k-d tree [Brown 2014], which can be pre-calculated for each point cloud, to expedite the nearest neighbor search.

Our method is not limited to Kinect data, thus data of varying scale and resolution can be visualized and labeled. We also emphasize the potential of this application for modifying point clouds, such as deleting noisy points, or cutting out relevant parts of the data using the same type of labeling and selection method.

4 CONCLUSION

We have realized an application for visualizing and labeling point clouds in Virtual Reality. The application demonstrates that it is possible to label and inspect large sets of room-scale point clouds in a fast and intuitive way. With the application, it is possible to select and track parts of the data to speed up labeling in sequences of point clouds. We believe that is a step in the direction towards incorporating point clouds into an efficient production pipeline for realistic animated virtual reality content, and more high precision labeled point cloud data sets for a variety of other applications as well.

5 ACKNOWLEDGEMENTS

Thanks to the MIT Juggling Club for allowing us to record the data.

REFERENCES

