Enhancing the AR Experience with Machine Learning Services

Web3d’19, July 26-28, 2019, Los Angeles CA, USA
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung
Introduction

- Existence of AR application increases
- AR frameworks only provide a local coordinate system, but many applications require an application-oriented coordinate space
- AR frameworks suffer from drift
- Web service to easy integrate state-of-the-art machine learning techniques
  - Initialization of complex AR scenes
  - Improve tracking quality
Introduction
Outline

1. Estimating Camera Pose and Rooms
2. The Service Architecture
3. Preparing a new Scene
4. Generating the Training Data
5. Learning ML Models from Generated Training Data
6. Evaluation of Artificial Training Data
Estimating the Camera Pose

Sending a Snapshot to the Web-Service

Calculating Camera Pose using solvePnpRansac method from OpenCV

Performing a Segmentation
Estimating the Room

Sending a Snapshot to the Web-Service

Performing a Classification

Sending the estimated room

Room 001

Enhancing the AR Experience with Machine Learning Services
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung
The Service Architecture

Enhancing the AR Experience with Machine Learning Services
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung
Preparing a new Scene (1/2)
Room Estimation

- Web-based 2d editor to define different areas
- Support of simple elements
  - Wall, Window, Door
  - Floor, Roof
  - Navigation elements
- Auto. calculation of the volumes of each room
Preparing a new Scene (2/2)
Pose Estimation

Creating projects from 3D Models

- Web-based 3d editor to define the labels that are used for the pose estimation
- Support of a simple explosion mechanism to pick all parts
Generating the Training Data (1/2)
Room and Pose Estimation

- Collecting data during application runtime using our AR tool
- Every snapshot is automatically located to a predefined area
- Using pretrained ML Models that are retrained to the new labels
- Refinement of data that will be continuously used to retrain during every new session

Enhancing the AR Experience with Machine Learning Services
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung
Generating the Training Data (2/2)
Pose Estimation using Artificial Data

- Blender to produce artificial training data
- Previously prepared config controls the rendered output
  - Rendering tuples of images and labels
  - Varying colors
  - HDR environment maps
  - PBR definitions for materials
  - Ambient occlusion
  - Variance in camera position, rotation and roll
  - Variable lighting
Learning ML Models

- Inception V3 or Inception Resnet V2 for classification of areas
- DeepLabV3+ is used for semantic image segmentation

Training Hardware:
- Intel Xeon CPU E5-1660 v4 3.2GHz
- 32 GB DDR3 RAM
- NVIDIA GeForce GTX 1080 (8GB GDDR5)
Evaluation (1/4)
Using Artificial Training Data

- Precision of pose estimation is dependent from segmentation quality
- Different datasets were used to measure the impact of rendering features to the quality of the segmentation of a DeepLab v3+ model

Enhancing the AR Experience with Machine Learning Services
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung
Evaluation (2/4)
Used Artificial Datasets

Uniform Material
Uniform Background

Simple Material
Monochrome Background

PBR Material
Real Background
AR-Tool to create Tuples
- For every tuple the registration has to be done automatically
- Tuples are captured and stored for later manual rating

Manual Rating of quality of each Tuple
- Software shows overlay images
- Image right shows an acceptable tuple (left) and an unacceptable tuple (right)
## Evaluation (4/4)

**Results**

Enhancing the AR Experience with Machine Learning Services  
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
<th>MeanDevFC</th>
<th>Pose Estimation Successful</th>
<th>Pose Distance</th>
<th>Pose Orientation Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBR-R-MIX</td>
<td>97.96</td>
<td>94.48</td>
<td>99.57</td>
<td>57.87</td>
<td>77.4 %</td>
<td>8.93 cm</td>
<td>10.34°</td>
</tr>
<tr>
<td>PBR-R</td>
<td>97.07</td>
<td>92.96</td>
<td>99.16</td>
<td>58.14</td>
<td>80.6 %</td>
<td>10.52 cm</td>
<td>12.71°</td>
</tr>
<tr>
<td>PBR-R-30k</td>
<td>97.25</td>
<td>92.15</td>
<td>99.3</td>
<td>59.41</td>
<td>71.0 %</td>
<td>9.67 cm</td>
<td>11.73°</td>
</tr>
<tr>
<td>PBR-R-RV</td>
<td>97.36</td>
<td>93.86</td>
<td>98.56</td>
<td>60.93</td>
<td>80.6 %</td>
<td>8.66 cm</td>
<td>10.28°</td>
</tr>
<tr>
<td>S-M</td>
<td>97.05</td>
<td>90.18</td>
<td>99.70</td>
<td>61.05</td>
<td>64.5 %</td>
<td>16.53 cm</td>
<td>15.91°</td>
</tr>
<tr>
<td>S-R</td>
<td>96.43</td>
<td>87.67</td>
<td>99.92</td>
<td>61.59</td>
<td>56.5 %</td>
<td>19.31 cm</td>
<td>17.26°</td>
</tr>
<tr>
<td>PBR-M</td>
<td>97.49</td>
<td>96.03</td>
<td>98.55</td>
<td>65.53</td>
<td>69.4 %</td>
<td>11.93 cm</td>
<td>14.31°</td>
</tr>
<tr>
<td>U-M</td>
<td>94.29</td>
<td>88.82</td>
<td>99.77</td>
<td>86.92</td>
<td>11.3 %</td>
<td>34.44 cm</td>
<td>34.80°</td>
</tr>
<tr>
<td>U-R</td>
<td>94.08</td>
<td>73.05</td>
<td>98.84</td>
<td>92.99</td>
<td>21.0 %</td>
<td>33.63 cm</td>
<td>34.80°</td>
</tr>
<tr>
<td>S-U</td>
<td>33.75</td>
<td>96.16</td>
<td>14.09</td>
<td>101.94</td>
<td>3.2 %</td>
<td>39.72 cm</td>
<td>34.80°</td>
</tr>
<tr>
<td>PBR-U</td>
<td>76.74</td>
<td>79.95</td>
<td>32.82</td>
<td>106.02</td>
<td>0.0 %</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>U-U</td>
<td>77.59</td>
<td>0.91</td>
<td>0.84</td>
<td>127.78</td>
<td>0.0 %</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Questions and Answers

Thank you

Q & A

Enhancing the AR Experience with Machine Learning Services
Michael Englert, Marcel Klomann, Kai Weber, Paul Grimm, Yvonne Jung