Motivation & Requirements

Just like the titular Harry Potter Marauder’s map we want to provide users with the means to follow any individual as they wander around a mapped space. Our Android app and machine learning model work together with Wyze cameras to provide a Multi-Target Multi-Camera Tracking (MTMCT) system for keeping your private space safe and secure. Some example use cases include tracking at risk individuals in a long-term care home or detecting unauthorized intrusions. Our goal is to handle at least 4 different cameras at once with minimum 4 different people.

Background & Models

The first step towards achieving MTMCT is by checking each camera frame for persons. When a person is identified a bounding box can be drawn around the detected person/object. In our case we are using Detectron2 which is a Faster RCNN created by Facebook and pretrained on the COCO dataset. Detectron2 outputs the set of bounding boxes for each detection.

The next step to achieve MTMCT is through single camera tracking. We are using DeepSort which uses a combination of feature embedding and a Kalman filter to generate a predicted trajectory for each objects next video frame position. If a close match between the next frame bounding box and the DeepSort predicted position then they are associated together as the same individual.

Dataset & Training

We created our own custom dataset using 3 separate locations with 21 different cameras and 15 unique IDs. This allowed us to focus our dataset on areas we needed to improve our model. To improve the realism we have also developed different scenarios such as entering the front door, taking off your shoes and then going to sit on the couch. The 3 separate locations are summarized below.

<table>
<thead>
<tr>
<th>Location</th>
<th>IDs</th>
<th>Cameras</th>
<th>Footage/camera Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>William’s</td>
<td>0,1,2,3,4,5</td>
<td>0,1,2,3,4,5,6</td>
<td>366 seconds Multi-cam, clothes switch</td>
</tr>
<tr>
<td>Kevin’s</td>
<td>0,3,4,5,6,7,8</td>
<td>7,8,9,10,11,12,13</td>
<td>714 seconds Single-cam, leave and return</td>
</tr>
<tr>
<td>Michael’s</td>
<td>0,3,9,10,11,12,13,14</td>
<td>14,15,16,17,18,19,20</td>
<td>600 seconds Multi-cam same clothing, real skills</td>
</tr>
</tbody>
</table>

Our ReID model was first pretrained on ImageNet but we performed retraining on our own dataset to improve the accuracy. The model was trained for 50 epochs using this dataset. An example training query and the matches are shown to the right.

Results & Conclusions

The results from retraining our ReID model with our data for 50 epochs can be found below. The retraining was successful and resulted in a large improvement for our testing set.

<table>
<thead>
<tr>
<th>Training</th>
<th>Mean Average Precision</th>
<th>Rank-1</th>
<th>Rank-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-trained</td>
<td>24.9%</td>
<td>9.2%</td>
<td>61.7%</td>
</tr>
<tr>
<td>Re-trained</td>
<td>68.3%</td>
<td>80.8%</td>
<td>96.7%</td>
</tr>
</tbody>
</table>

The results from retraining our tracking model with our data for 30 epochs can be found below. The retraining was successful and resulted in a large improvement for our testing set.

MTMCT Results for full pipeline using ReID in combination with pre-trained DeepSort tracking.

Future Work, References, and Acknowledgments

- Collect and label more data with more IDs/Locations.
- Host data pipeline online.
- Live data streaming to the app.
- Examine joint detection

Embedding with FairMOT.

Faculty: Prof. Payman Kadkhodazadeh, Dr. Hung-Min Hsu
TA’s: Haosu Zhang, Shoul Miao, Daniel King