Appearance Models for People Tracking

Professor Pascal FUA
Computer Vision Lab
Zhongqi.shi@epfl.ch

Introduction
In this master project, we present a complete system of appearance models for people tracking. Firstly, we evaluate several types of appearance models based on the Siamese network architecture, to find the most suitable one. Afterwards, we present two ways of tracking using the appearance information - one that preserves appearances locally, and one that preserves appearances globally.

Theory/Method/Hypothesis
Siamese neural network[1] is used in this project, which consists of two or more identical sub-networks based on convolutional neural networks. The topology of the Siamese neural network that we use is:

![Figure 1: Model Architecture](image)

In Fig. 1, a pair of images is fed into the network which includes two twin CNN branches. The CNN blocks are used to extract the features of the input images.

![Figure 2: Convolutional Architecture](image)

Fig. 2 shows the architecture of CNNs block, which includes four convolutional layers and one fully connected layer. Finally, we save the models of Siamese network based on Station and MOT16 dataset, respectively. The output is the probability of how similar the input two images are. Then, we have two optimization schemes. The first one is tracking without clusters:

\[
edge_{cost_{ij}} = \log(\frac{P_i}{1-P_i}) \cdot (D_j - 1) \cdot \text{time\_penalty} \tag{1}
\]

Another one is tracking with clusters:

\[
edge_{cost_{ij}} = \log(\frac{P_i}{1-P_i}) \cdot (D_j - 1) \cdot \text{time\_penalty} \cdot \text{group}(i) + \text{group}(j) \tag{2}
\]

Eq. (1) only considers the appearance between image pairs. Eq. (2) run the optimization algorithm as many times as the number of clusters based on people’s global appearance.

Results
We have the models of Station dataset and MOT16 dataset, respectively. For example:

![Figure 3: Testing loss and accuracy of Station dataset](image)

Table 1: IDF of Station and MOT16 dataset

<table>
<thead>
<tr>
<th>Dataset/IDF</th>
<th>Basic</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station</td>
<td>0.5909</td>
<td>0.5605</td>
</tr>
<tr>
<td>MOT16-13</td>
<td>0.3076</td>
<td>0.3173</td>
</tr>
<tr>
<td>MOT16-11</td>
<td>0.4719</td>
<td>0.4243</td>
</tr>
<tr>
<td>MOT16-15</td>
<td>0.3977</td>
<td>0.4869</td>
</tr>
<tr>
<td>MOT16-09</td>
<td>0.3965</td>
<td>0.4217</td>
</tr>
<tr>
<td>MOT16-05</td>
<td>0.4855</td>
<td>0.4820</td>
</tr>
<tr>
<td>MOT16-04</td>
<td>0.3676</td>
<td>0.3386</td>
</tr>
<tr>
<td>MOT16-02</td>
<td>0.3003</td>
<td>0.3038</td>
</tr>
</tbody>
</table>

Table 1 presents the IDF results of Station and MOT16 dataset of two optimization schemes - tracking without clusters (‘Basic’) and tracking with clusters.

Conclusion/Perspectives
Appearance models for people tracking achieves great performance in our project. In future work, we should investigate more reliable Siamese network and appearance models to get better tracking result.

Acknowledgments
I thank Professor Pascal Fua for granting me the opportunity to work on the master project.
I thank Andrii Maksai for his mentoring.

References