A Multilayer-Based Framework for Online Background Subtraction with Freely Moving Cameras
Yizhe Zhu Ahmed Elgammal
Rutgers University, Department of Computer Science
yizhe.zhu@rutgers.edu elgammal@cs.rutgers.edu

Motivation
- An increasing amount of videos are captured from moving platforms.
- Videos need to be analyzed as they being streamed in real time.
- Most general Background Subtraction methods are restricted to binary segmentation.

Contributions
- We formulate Background Subtraction as an online multi-label segmentation problem by modeling multiple foreground objects in different layers.
- We design an processing block to handle the information in each layer separately and simultaneously.

Our Proposed Framework
- The accumulated appearance models and probability maps of each layer are fed to “processing blocks”
- With the collection of probability maps produced by each “processing block”, we use Multi-label Graphcut to carry out the segmentation.
- Add the new color feature (RGB) in the current frame to update the corresponding appearance model.

Processing Block
- Trajectory labeling:
  - Recursive normalized cuts to get the initial labels of trajectories.
  - Label propagation to infer labels in the following frames.

- Motion estimation:
  - In each layer, Motion estimation is performed based on Gaussian Belief Propagation with motion of trajectories as the evidence. 
  
  \[P\left(M^{i\mid t} \mid P_t\right) \propto \prod_{(i,j) \in E} \Psi(m^{k_{i,j}}, m^{k_{j,i}}) \prod_{i \in S_{k_{i,j}}} \phi(m^{k_{i,j}})\]

- Shift the appearance model and the prior probability from the previous one with the estimated motion model.

Results Compared with State-of-the-art
- Two-label Background Subtraction
- Multi-label Background Subtraction

Table 1: Two-label background subtraction performance comparison on the videos with different numbers of moving objects.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>84.05</td>
<td>80.70</td>
<td>59.20</td>
<td>54.22</td>
<td>47.82</td>
<td>31.20</td>
</tr>
<tr>
<td>2</td>
<td>91.14</td>
<td>81.51</td>
<td>83.19</td>
<td>71.48</td>
<td>80.30</td>
<td>62.75</td>
</tr>
<tr>
<td>≥ 3</td>
<td>74.99</td>
<td>65.05</td>
<td>66.26</td>
<td>57.02</td>
<td>55.72</td>
<td>45.05</td>
</tr>
</tbody>
</table>

Figure 1: Qualitative comparisons.

![Figure 1](image)

Table 2: Performance comparison of Multi-foreground segmentation on Hopkins Dataset.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>First 10 Frames</td>
<td>ours SLT Baseline</td>
<td>89.79</td>
<td>83.00</td>
</tr>
<tr>
<td>All Frames</td>
<td>ours SLT Baseline</td>
<td>89.51</td>
<td>78.24</td>
</tr>
</tbody>
</table>

Figure 2: The Multilabel segmentation performance comparison with SLT[6]