Learning Spatiotemporal Filters to Track Visual Saliency

Khaled Aboumerhi – ECE Ph.D. Candidate
Ralph Etienne-Cummings, Ph.D. – Professor of ECE
Agenda

▪ Visual Saliency Overview

▪ Prophesee.ai Dataset

▪ Handcrafting Filters

▪ Pre-processing for Visual Saliency Model

▪ Spatiotemporal Filters

▪ Applications of Proposed Unsupervised Visual Saliency Model
Visual Saliency

- Tendency to gaze in a particular direction or toward an interesting feature
- Applications in high-accuracy drone cameras, real-time traffic, criminal investigation, tele-tourism and more
- Why is it humans gaze toward specific features?

(Kadam et. al, 2020)
**Dataset – Prophesee.ai**

- ATIS Camera
- APS events (left) and real visualization (right)
- 10 µs resolution

Streetcar Dataset

Motorway Dataset
Preprocessing Technique

[Flowchart showing decision-making process for finding and following new targets based on event occurrences and distance calculations.]
Spatiotemporal Filters

Vertical (handcrafted) Filter

\[ S_{u,i} = e^{(-t_i - T_{u,i})/\tau} \]  \hspace{1cm} (1)

Learned Filter

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>a</th>
<th>b</th>
<th>b</th>
<th>a</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

-1 -0.9 -0.4 -0.1 0 0
-0.9 -0.5 -0.2 0.1 0 0
-0.6 -0.3 0.1 0.29 0.5 0
-0.1 0.2 0.3 0.6 0.8 0
0 0 0.6 0.8 0.9 0
0 0 0 0.8 1 0
Learning Spatiotemporal Filters

\[
\eta_k = \det|S_{u,i} - S_{u,i-1}| \quad (2)
\]

\[
C_k = \arg\min|\eta, k| \quad (3)
\]

\[
F_A(\delta) = 0 \quad (4)
\]

\[
\epsilon_A = \det|\delta F_A| \quad (5)
\]

\[
\sigma = \text{rand}(\Sigma \epsilon) \quad (6)
\]
But what is lifelong learning?

- Human brains use lifelong learning to keep important information based on many parameters, including
  - Recency
  - Relevance
  - Intense/reactive response

- How can we properly manage distinct, unique information over a timeless process without losing different types of important filters?

- Decision trees are great for simple decision making, but how can we incorporate many parameters for lifelong learning?

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon_n(t) \rightarrow 0$</td>
<td>consistency</td>
</tr>
<tr>
<td>$\epsilon_n(t &gt; t') = \epsilon_n(t)$</td>
<td>no catastrophic forgetting</td>
</tr>
<tr>
<td>$\epsilon_{n+m}(t &gt; t') \rightarrow 0$</td>
<td>continual learning</td>
</tr>
<tr>
<td>$#(f_{n+m}) - #(f_n) \in \Omega(\log(m))$</td>
<td>space efficient</td>
</tr>
</tbody>
</table>

(Vogelstein et. al, 2019)

Lifelong learning has few, yet strong parameters, of which no catastrophic forgetting and continual learning are incredibly important for timeless and online learning.

Creating stronger and less correlated trees using relaxation constraints clears up with computational resources.
Large Dataset Management

- Prophesee data set is large (~$10^9 \times 1$ vector)
- Parsed to 10 μs blocks of online learning – online application
- Note: after a full 20 μs period of learning, time surface management becomes complicated
  - How do we incorporate previous filters in future learning/application?
  - How can we implement desired filter inputs?
Learned Spatiotemporal Filters

Streetcar

Motorway

A
N = 4, R = 4

B
N = 4, R = 16

A
N = 4, R = 4

B
N = 4, R = 16

C
N = 16, R = 16

C
N = 16, R = 16
Learned Spatiotemporal Filters

- Large region (R) filters
  - Not used, but shows clearly the movement of event-based data through a field of vision
Implementations of Filters - Motorway

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>IE</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>7.1</td>
<td>2.3</td>
</tr>
<tr>
<td>red</td>
<td>3.403</td>
<td>4.13</td>
</tr>
<tr>
<td>yellow</td>
<td>24.10</td>
<td></td>
</tr>
</tbody>
</table>
Implementations of Filters - Streetcar

Abstract centroid placement comparison between lowest and highest generality filters

Experienced Observer

Inexperienced Observer
Future Work

- acquire the first spike-based saliency data-set
- Visual saliency algorithms that track and attempt to expose the brain’s tendencies cannot be truly verified without a ground-truth.
- We can do this using eye-tracking devices in a closed environment, such as the HTC Vive or Google HoloLens.
- Breaking up the TD events in groups of two, four, and eight seconds, while applying the centroids during the intermittent latent phases, could prove to be a more accurate and quick process.

