Saliency and Human Fixations: State-of-the-art and Study of Comparison Metrics

Motivation

1/ Analysis of validation metrics for saliency models
   • Dozens of saliency models recently published...
   • Finding a way to fairly evaluate all of these models became a challenge

2/ Redundancy analysis on validation metrics
   • Major contributions in saliency benchmarks: Borji, Judd, Toets...
   • None deeply discussed the relevance of their similarity metrics mix.

Literature review of similarity metrics

1/ Analysis of metrics consistency:

<table>
<thead>
<tr>
<th>Common Metrics</th>
<th>Value-based Metrics</th>
<th>Location-based Metrics</th>
<th>Distribution-based Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid Metrics</td>
<td>NSS / Percentile / Pf</td>
<td>AUC-Judd</td>
<td>KL-DiV / CC / Spear / Similarity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specific Metrics</th>
<th>NSS / Percentile / Pf</th>
<th>AUC-Judd</th>
<th>KL-DiV / CC / Spear / Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Torralba A.</td>
<td>2006</td>
<td>CC:</td>
<td>Ouerhani N.</td>
</tr>
<tr>
<td>AUC-Zhao</td>
<td>2011</td>
<td>Similarity:</td>
<td>Judd T.</td>
</tr>
<tr>
<td>AUC-Judd</td>
<td>Judd T.</td>
<td></td>
<td>2012</td>
</tr>
<tr>
<td>AUC-Li</td>
<td>Ji L.</td>
<td></td>
<td>2012</td>
</tr>
<tr>
<td>AUC-Borji</td>
<td>Borji A.</td>
<td></td>
<td>2012</td>
</tr>
</tbody>
</table>

Experimental Setup

1/ Database: Jian Li’s dataset
   • 235 images
   • 20 viewers
   • small, medium and large objects of interest

2/ Saliency Maps: 12 models
   • cognitive approach: ITTI
   • Baysian models: SUN and Torralba
   • information theoretic category: AIM, DVA and RARE
   • spectral approach: SR, PFT, PQFT and Achanta
   • other models: HFT and AWS

Conclusion

Conclusion 1:
We reviewed 12 state-of-the-art similarity metrics for visual saliency models validation. The comparison is based on the ranking between models using Kendall’s W coefficient (PCA provides poor results).

Conclusion 2:
Our study shows that only metric alone is not enough to evaluate the saliency model ranking on eye fixation data.

Conclusion 3:
Our approach can compare saliency models rankings only, thus it is not possible here to test if the differences between the saliency models are statistically significant.

Conclusion 4:
The minimal set of similarity metrics which should be used is a) one of the metrics composing the Cluster, b) AUC-Borji and c) KL-Div. The use of those three metrics is enough to cover most of the space and provide a fair ranking result.

http://tcts.fpms.ac.be/attention

Nicolas Riche, Matthieu Duvinage, Matei Mancas
nicolas.riche@umons.ac.be | matthieu.duvinage@umons.ac.be | matei.mancas@umons.ac.be