Visual attention in daylit architecture

Dr. Caroline Karmann
Postdoctoral Researcher
at LIPID

Prof. Marilyne Andersen
Head of LIPID

Paris - October 2019
Laboratory of Integrated Performance in Design (LIPID)

- Comfort
- Perception
- Health
- Energy
Example:

Ryerson Student Learning Centre, Toronto, CA
Zeidler Partnership Architects + Snøhetta

Images: L. Bridgman, Doublespace Photography
Example:

Ryerson Student Learning Centre, Toronto, CA
Zeidler Partnership Architects + Snohetta

Illustration from (Rockcastle et al., 2017) (adapted) | Photograph from L. Bridgman, Doublespace Photography
LIPID previous work

Study by (Rockcastle et al., 2017)

Objectives
- Evaluate perceptual responses
- Effect of sky condition / space
- Develop a contrast-based metric (mSC5)

Procedure
- VR-based human-subject experiment
- 360° fully immersive (Oculus rift, cubemaps)
- Exposure to stimuli: participant's choice

Dataset
- Visual stimuli: 16 scenes (B/W renderings)
- Headtracking logs: ~12 participants/scene

LIPID previous work

Process

Headtracking logs → Methodology to define a reference → Ground-truth

Comparison effect of sky

Ground-truth → predicted vs. ground-truth

Scene characteristics → Selection of saliency prediction algorithms → Predicted visual attraction

Comparison predicted vs. ground-truth
Saliency prediction

Do existing saliency models accurately predict visual attention in daylit architectural spaces?
Establishing ground truth

Coordinate systems
- Equi-projection
- Account for distortions

Exposure to visual stimuli
- 5 sec, 25 sec, no limit

Fixations vs. saccades
- Angular velocity threshold
  - 15°/sec, 60°/sec, no limit

Gaussian filter
- Standard deviation (Rai et al., 2017)
- Corrected filter (Upenik et al., 2017)
- Rounding and flattening (Upenik et al., 2017)

Illustration from Oculus Rift Developer Guide
Upenik et al., 2017, A simple method to obtain visual attention data in head mounted virtual reality
Rai et al., 2017, A Dataset of Head and Eye Movements for 360 Degree Images
Establishing ground truth
Establishing ground truth
**Saliency prediction**

**Visual attentional mechanism**

**Bottom-up attention**
- low-level visual features
  - e.g., intensity, color, orientation, texture, directions

**Top-down attention**
- high-level features (recognition)
  - e.g., faces, cars, objects, furniture, etc.

**Early models:**
- Itti-Koch model (2001)
- Graph-based Visual Saliency (GBVS) (Harel et al., 2006)

Saliency prediction

Visual attentional mechanism

**Bottom-up attention**
low-level visual features  
- e.g., intensity, color, orientation, texture, directions

**Top-down attention**
high-level features (recognition)  
- e.g., faces, cars, objects, furniture, etc.

**Deep learning**
Convolutional neural networks (CNN)

Learn from images / multiple layers  
(input > output)
Allows large scale object recognition corpses
Perform better than traditional saliency models

Saliency prediction

Visual attentional mechanism

Bottom-up attention
low-level visual features
  e.g., intensity, color, orientation, texture, directions

Top-down attention
high-level features (recognition)
  e.g., faces, cars, objects, furniture, etc.

Publicly available and pre-trained CNN-based models from VR data

SalNet360 (Monroy et al. 2018)

SaltiNet (Assens et al. 2017)

Results
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Douglas House | Clear
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

SalNet360

SaltiNet

Douglas House | Overcast
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Ryerson | Clear

SalNet360

SaltiNet
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Ryerson | Overcast

SalNet360

SaltiNet
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Serpentine pavilion | Clear

SalNet360

SaltiNet
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Serpentine pavilion | Overcast
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Zollverein | Clear

SalNet360

SaltiNet
Saliency prediction vs. ground truth

Ground Truth (stimuli 25s. | fixations <60°/s.)

Zollverein | Overcast

SalNet360

SaltiNet
Insights and limitations

Qualitative analysis

SalNet360 finds salient spots
SaltiNet identifies larger zones Work in different ways

Both model embed equator bias
Artefacts not corrected (e.g., cubes, border)
Low-level visual features identified

Quantitative analysis

Distribution-based statistical metrics

Linear correlation coefficient

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean CC</th>
<th>Range (Abs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SalNet360</td>
<td>0.25</td>
<td>(0.03 - 0.44)</td>
</tr>
<tr>
<td>SaltiNet</td>
<td>0.46</td>
<td>(0.23 - 0.58)</td>
</tr>
<tr>
<td>Laplacian</td>
<td>0.66</td>
<td>(0.26 - 0.80)</td>
</tr>
</tbody>
</table>

Kullback-Leibler divergence

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean KL</th>
<th>Range (Abs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SalNet360</td>
<td>3.60</td>
<td>(2.04 - 5.99)</td>
</tr>
<tr>
<td>SaltiNet</td>
<td>2.58</td>
<td>(1.30 - 4.35)</td>
</tr>
<tr>
<td>Laplacian</td>
<td>1.77</td>
<td>(0.63 - 2.83)</td>
</tr>
</tbody>
</table>

Sitzman et al., 2018, Saliency in VR: How do people explore virtual environments? IEEE TVCG
Insights and limitations

How are saliency algorithms trained?

How does this compare to our data?
Effect of sky conditions

Does our viewing behavior change with sky conditions?
Comparing ground truth output for different sky conditions

Mean $\text{CC}_{\text{Clear-Overcast}} = 0.69 \ (0.46 - 0.87)$
Mean $\text{KL}_{\text{Clear-Overcast}} = 1.96 \ (1.17 - 2.97)$

Clear vs. overcast
(same scene)

Viewing patterns presents similarities despite changing sky conditions
Comparing ground truth output for different sky conditions

Linear model output and plots

Higher correlations for the most ‘horizontal’ scenes
Conclusions

Existing saliency models could not accurately predict visual attention in rendered black and white architectural scenes.

Good insights to be gained.

Discrepancies in the experimental protocol (limitation)
→ Adapt our protocol if we like to further test/use saliency models.

Tendency to look outside.

Viewing patterns in a space remains somehow consistent despite varying sky conditions.

Validation under real conditions (missing).
Thank you!
caroline.karmann@epfl.ch
Acknowledgments

Data collection

Kynthia Chamilothori
Assistant Professor at the Eindhoven University of Technology (TU/e)
Laboratory of Integrated Performance in design, EPFL (previously)

Siobhan Rockcastle
Assistant Professor at University of Oregon
Laboratory of Integrated Performance in design, EPFL (previously)

Data processing and expertise

Seungryong Kim
Image and Visual Representation Lab, EPFL

Evgeniy Upenik
Multimedia Signal Processing Group, EPFL