Salient360!
Visual Attention for 360° Content

Jesús Gutiérrez, Patrick Le Callet

Image, Interaction, Perception Group (IPI)
Laboratoire des Sciences du Numérique de Nantes (LS2N)
Université de Nantes
Introduction and motivation

• Visual attention allows to know the important regions of the scene for the observers.

• Proxy for visual fidelity widely used for 2D and 3D content:
  • Coding and transmission: protection based on saliency.
  • Quality evaluation:
    • Weighting most important regions.
    • Consider artistic intentions.


Introduction and motivation

- Even more important for 360° content: Not everything may be seen!
- Different scenarios compared to conventional image/video viewing:
- No direct use of 2D VA models for 360° content.
Introduction and motivation

• Even more important for 360° content: Not everything may be seen!
  • Applications:
    • Tile-based coding and streaming, non-uniform quality streaming...
    • Foveated rendering
    • Storytelling, artistic intent, movie editing...

• Evaluation of quality using tracking data: weighting the metrics.

$$\text{PSNR} = 10 \log \left( \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} (\text{error}(i, j))^2 \right)$$

«Saliency» weight
Introduction and motivation

• Need of **datasets** and **benchmarking**:
  • Images:
    • MIT Saliency benchmark
  • Video & Higher Resolution

T. Vigier *et al.*, “Impact of visual angle on attention deployment and robustness of visual saliency models in videos: From SD to UHD”, *ICIP 2016.*

Available at: [http://ivc.univ-nantes.fr/en/]
Introduction and motivation

• Need of datasets:
  • HDR / TM Images/videos

  • 3D TV Images/videos


• 360° content?

State of the art

• VA studies for 360° content:
  - ...
State of the art

• Datasets for 360° content:

  - X. Corbillon, et al., “360-Degree Video Head Movement Dataset”, ACM MMSys’17 → [7 videos from Youtube, 70 seconds, 59 observers]
  - W-C. Lo et al., “360 Video Viewing Dataset in Head-Mounted Virtual Reality” ACM MMSys’17 → [10 videos from Youtube, 1min, 50 observers]
  - C. Wu et al., “A Dataset for Exploring User Behaviors in VR Spherical Video Streaming”, ACM MMSys’17 → [18 videos from Youtube, 2-8 min, 48 observers]
  ✓ A. Serrano et al., “Movie editing and cognitive event segmentation in virtual reality video”, ACM TOG 2017. → [216 clips (2 shots), 6+6 s, 49 observers]

• Need of more video datasets with head and eye tracking data.
Importance of eye-tracking data

- Head movement can be a proxy of VA for some applications, but...
  - Observers explore within the viewport.
  - Not high correlations between head-only and head-eye saliency maps → How to approximate?
- Eye data is important for many applications:
  - Quality assessment: weighting of metrics
  - Coding, streaming, foveated rendering, movie editing, cinematography, etc.
Dataset of 360° images

• Dataset for still images:
  • 85 equirectangular images.
  • Processed data from **head and eye movements**:
    • Head saliency maps.
    • Head-Eye saliency maps.
    • Scanpaths (Head-Eye).
• Tools:
  • Parsing the data:
  • Compare saliency maps.
  • Compare scanpaths.
• To access it: email to salient360@univ-nantes.fr

Dataset of 360° images

- Dataset for still images:
  - 85 equirectangular images.
  - Processed data from **head and eye movements**:
    - Head saliency maps.
    - Head-Eye saliency maps.
    - Scanpaths (Head-Eye).
- Tools:
  - Parsing the data:
  - Compare saliency maps.
  - Compare scanpaths.
- To access it: email to salient360@univ-nantes.fr


**Evaluation Toolbox**
- parseSaliencyMapScanpaths.m
- CompareSaliencyMaps
  - CompareHeadEyeSaliencyMaps.m
  - CompareHeadSaliencyMaps.m
- SpiralSampleSphere.m
- CC.m
- KLDiv.m
- NSS.m
- AUC_Judd.m
- CompareScanpaths
  - scanpathMetric.m
  - checkVectorSimilarity.m
  - munkres.m
  - arrow.m
Dataset of 360° images

• Dataset for still images:
  • 85 equirectangular images.
  • Processed data:
    • Head saliency maps
    • Head-Eye saliency maps
    • Scanpaths (Head-Eye)

• ICME’17 Grand Challenge Salient360!
  • Align Saliency modeling community
  • Provide dataset
  • Benchmark of models
  • Modeling approach

• Special issue in Elsevier Signal Processing: Image Communications (To appear soon)
Dataset of 360° images

**Equipment:**
- HMD Oculus Rift DK2
  - Horizontal and vertical FoV: 100°
  - Resolution: 960x1080 per eye.
  - Refresh rate & head-tracking data rate: 75Hz.
- SMI Eye-tracker
  - Binocular eye-tracking at 60Hz.

**Execution of the test:**
- Free-viewing: “view as naturally as possible”.
- Seated in a rolling chair.
- Each stimulus: 25 seconds (6 seconds between stimulus).
- Total duration: 35 minutes + 5 minutes pause.

**Observers:**
- 63 (24 females / 39 males).
- Average age 30 (from 19 to 52).
- 40 observers per image.
- Expertise: 32/63 used HMD less than 2 times, 8 experts.
Salient360! – 2017

- Results

### Table I

RESULTS FOR MODEL TYPE 1

<table>
<thead>
<tr>
<th>Team</th>
<th>KL</th>
<th>CC</th>
<th>ROC</th>
<th>Rank KL</th>
<th>Rank CC</th>
<th>Rank NSS</th>
<th>Rank ROC</th>
<th>Rank Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhejiang University [11]</td>
<td>0.44</td>
<td>0.69</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Wuhan University [12]</td>
<td>0.51</td>
<td>0.71</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SJTU [13]</td>
<td>0.65</td>
<td>0.67</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>TU Munich (1) [14]</td>
<td>0.75</td>
<td>0.62</td>
<td></td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>TU Munich (2) [14]</td>
<td>0.74</td>
<td>0.60</td>
<td></td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>TU Munich (3) [14]</td>
<td>0.64</td>
<td>0.56</td>
<td></td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Roma Tre University [15]</td>
<td>0.81</td>
<td>0.52</td>
<td></td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>USTC [17]</td>
<td>1.05</td>
<td>0.67</td>
<td></td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Anonymous</td>
<td>0.92</td>
<td>0.58</td>
<td></td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>TU Munich (4) [14]</td>
<td>0.91</td>
<td>0.50</td>
<td></td>
<td>6</td>
<td>11</td>
<td>11</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Jiangxi University of Finance and Economics [16]</td>
<td>1.14</td>
<td>0.57</td>
<td>13</td>
<td>6</td>
<td>9.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TU Munich (5) [14]</td>
<td>1.09</td>
<td>0.44</td>
<td></td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>IRISA</td>
<td>1.07</td>
<td>0.41</td>
<td></td>
<td>9</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

### Table II

RESULTS FOR MODEL TYPE 2

<table>
<thead>
<tr>
<th>Team</th>
<th>KL</th>
<th>CC</th>
<th>NSS</th>
<th>ROC</th>
<th>Rank KL</th>
<th>Rank CC</th>
<th>Rank NSS</th>
<th>Rank ROC</th>
<th>Rank Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TU Munich (1) [14]</td>
<td>0.45</td>
<td>0.58</td>
<td>0.81</td>
<td>0.73</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TU Munich (2) [14]</td>
<td>0.42</td>
<td>0.61</td>
<td>0.81</td>
<td>0.72</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SJTU [13]</td>
<td>0.48</td>
<td>0.53</td>
<td>0.92</td>
<td>0.73</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.25</td>
</tr>
<tr>
<td>TU Munich (3) [14]</td>
<td>0.50</td>
<td>0.55</td>
<td>0.92</td>
<td>0.75</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3.25</td>
</tr>
<tr>
<td>TU Munich (4) [14]</td>
<td>0.48</td>
<td>0.56</td>
<td>0.70</td>
<td>0.71</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Zhejiang University [11]</td>
<td>0.70</td>
<td>0.53</td>
<td>0.85</td>
<td>0.71</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>5.75</td>
</tr>
<tr>
<td>Trinity College (1) [17]</td>
<td>0.49</td>
<td>0.54</td>
<td>0.76</td>
<td>0.70</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>6.25</td>
</tr>
<tr>
<td>USTC [17]</td>
<td>2.02</td>
<td>0.51</td>
<td>0.92</td>
<td>0.69</td>
<td>17</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
</tr>
<tr>
<td>Trinity College (3) [17]</td>
<td>0.49</td>
<td>0.58</td>
<td>0.53</td>
<td>0.66</td>
<td>5</td>
<td>1</td>
<td>13</td>
<td>12</td>
<td>7.75</td>
</tr>
<tr>
<td>TU Munich (5) [14]</td>
<td>0.58</td>
<td>0.41</td>
<td>0.69</td>
<td>0.69</td>
<td>5</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>7.75</td>
</tr>
<tr>
<td>Anonymous</td>
<td>0.59</td>
<td>0.41</td>
<td>0.68</td>
<td>0.69</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>TU Munich (6) [14]</td>
<td>0.64</td>
<td>0.40</td>
<td>0.63</td>
<td>0.68</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>12</td>
<td>10.5</td>
</tr>
<tr>
<td>Yonsei University</td>
<td>0.78</td>
<td>0.59</td>
<td>0.51</td>
<td>0.63</td>
<td>14</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>IRISA</td>
<td>0.59</td>
<td>0.45</td>
<td>0.51</td>
<td>0.64</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>11.75</td>
</tr>
<tr>
<td>DAIICT</td>
<td>1.16</td>
<td>0.19</td>
<td>0.42</td>
<td>0.65</td>
<td>16</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>14.25</td>
</tr>
</tbody>
</table>

### Table III

RESULTS FOR MODEL TYPE 3

<table>
<thead>
<tr>
<th>Team</th>
<th>SimMetric</th>
<th>Rank SimMetric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insight Centre for Data Analytics - UPC [18]</td>
<td>2.8697</td>
<td>1</td>
</tr>
<tr>
<td>SJTU [13]</td>
<td>4.6565</td>
<td>2</td>
</tr>
<tr>
<td>Wuhan University [12]</td>
<td>5.9517</td>
<td>3</td>
</tr>
</tbody>
</table>

VQEG Meeting
20/03/2018
Salient360!: Visual Attention for 360° Content

• **Salient360!** – Grand Challenge at ICME’18:
  • 360° content: images and video.
  • 4 types of models:
    1. Head-only saliency maps.
    2. Head and eye saliency maps.
    3. Head and eye scanpaths.
    4. Head-only “scanpaths” (trajectories).

• More info:
  • Webpage: [www.salient360.ls2n.fr](http://www.salient360.ls2n.fr)
  • Email to salient360@univ-nantes.fr
Conclusions

• Ongoing work:
  • Dataset of eye and head movements for 360 video.
  • Toolbox for:
    • Processing eye/head data and generate saliency maps and scanpaths
    • Comparing saliency maps and scanpaths

• Benchmarking platform for VA in 360 content.
Salient360!
Visual Attention for 360° Content

Jesús Gutiérrez, Patrick Le Callet

Image, Interaction, Perception Group (IPI)
Laboratoire des Sciences du Numérique de Nantes (LS2N)
Université de Nantes