An investigation of visual selection priority of objects with texture and crossed and uncrossed disparities

Dar’ya Khaustova, Jérôme Fournier, Emmanuel Wyckens
Orange Labs, France

Olivier Le Meur
University of Rennes 1, France
Motivation of the study

Goal – compare visual attention (VA) between 2D and 3D fully controlled contents having objects with crossed disparity in comfortable and uncomfortable conditions.

Conducted experiments:
Exp.1 – Uncrossed disparity (UD);
Exp.2 – Crossed disparity (CD);
Exp.3 – Experiment with spheres.
Crossed disparity experiment (Exp. 2)
3D stereoscopic systems: shooting and visualization

- **Shooting side**
  - Real object
  - Convergence distance
  - Inter-cameras baseline
  - Focal length
  - Cameras

- **Visualization side**
  - Screen
  - Screen size
  - Viewing distance
  - Ocular distance
  - Eyes
Exp2. Scenes parameters

<table>
<thead>
<tr>
<th>Scene</th>
<th>DoF, dpt</th>
<th>Baseline, mm</th>
<th>Convergence distance, m</th>
<th>Disparity on the screen, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
</tr>
<tr>
<td>cartoon</td>
<td>0.1</td>
<td>220</td>
<td>8</td>
<td>-15</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>660</td>
<td>8</td>
<td>-46</td>
</tr>
<tr>
<td>hall</td>
<td>0.1</td>
<td>342</td>
<td>8</td>
<td>-15</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>1020</td>
<td>8</td>
<td>-46</td>
</tr>
<tr>
<td>pigs</td>
<td>0.1</td>
<td>53</td>
<td>2.2</td>
<td>-15</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>162</td>
<td>2.2</td>
<td>-47</td>
</tr>
<tr>
<td>table</td>
<td>0.1</td>
<td>325</td>
<td>5.45</td>
<td>-16</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>1000</td>
<td>5.45</td>
<td>-45</td>
</tr>
</tbody>
</table>

Maximum disparity on the screen:

- Comfortable DoF=0.1 dpt 1.5 cm (1.6% of display width)
- Uncomfortable DoF=0.3 dpt 4.5 cm (4.8% of display width)

Amount of perceived depth is the equal for all the scenes with same DoF.
Exp2. Stimuli generation

Cartoon

Table

Hall

Pigs

4 scenes $\times$ 3 depth levels = 12 stimuli
Exp2. Experimental design

Methodology:

• 12 images divided into 3 sets;
• Each scene was seen only once by every observer;
• 51 observers (17 observers per image).
Exp2. Experimental methodology

Test stages:
1. Instruction sheet
   • free-viewing task;
2. Calibration (5 point calibration + calibration chart)
3. Training 1 min 20 s
4. Test
   • duration 1 min 40 s
   • 1 set (4 images: cartoon, hall, table and pigs)
   • 20 s per image + 5 s of gray screen

51 people were tested
17 people per image

Calibration chart

Unsuccessful calibration chart
Successful calibration chart
Exp2. Qualitative results of the heat maps

<table>
<thead>
<tr>
<th>2D</th>
<th>3D DoF=0.1 (comfortable)</th>
<th>3D DoF=0.3 (uncomfortable)</th>
</tr>
</thead>
</table>

[Images of heat maps]
We performed eye-tracking experiments using fully controlled 3D content with crossed disparity and compared visual attention between 2D, 3D comfortable and 3D uncomfortable conditions:

### Conclusions for Exp.1(UD) and Exp.2(CD)

<table>
<thead>
<tr>
<th>Is there an influence on VA in comparison with 2D? Of:</th>
<th>Exp.1 UD</th>
<th>Exp.2 CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disparity</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Texture</td>
<td>YES</td>
<td>?</td>
</tr>
<tr>
<td>Discomfort</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saccade length (4 s)</th>
<th>YES \downarrow</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saccade length (20 s)</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Fixation durations (20 s)</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>CC and AUC (20 s) add 4s</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Observers reported that it was interesting to look at objects with excessive disparity.

So what is more influential in guiding our gaze: texture or depth?
Experiment with spheres (Exp. 3)
Exp3. Stimuli generation

Arrangement of spheres in depth

Parameters to study:
1. **Texture**
   - check board
   - gray
2. **Depth**
   - 2D
   - Uncrossed disparities (UD)
   - Mixed disparities (MD)
   - Crossed disparities (CD)

Front view = 2D

Top view

Scenes
Exp3. Eye tracking data analysis

Position

Texture

<table>
<thead>
<tr>
<th>Position</th>
<th>s1</th>
<th>s2</th>
<th>s3</th>
<th>s4</th>
</tr>
</thead>
<tbody>
<tr>
<td>11_MD</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11_2D</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11_CD</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>11_UD</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Observer</th>
<th>Position</th>
<th>Depth</th>
<th>Order</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>11_MD</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11_MD</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>11_MD</td>
<td>5</td>
<td>3</td>
<td>-1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11_MD</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Exp3. Influence of depth on visual attention

CD significantly influences on visual attention in comparison with 2D – $F(1, 3264)=13.14, p<0.05, p=0.0003$.

Texture significantly influences sphere selection order: $F(1, 4456)=12.31, p<0.05, p=0.0005$.
Exp3. Influence of the position of the spheres on test results

Influence of position and texture on the order

Influence of position and depth on the order

CD
2D
UD
s1
s2
s3
s4
1,4
1,6
1,8
2,0
2,2
2,4
2,6
Order

gray
check board

S1
S3
S2
S4

Order

CD
2D
UD
Exp3. Saccade length and fixation duration (5 s)

Average saccade length, deg

<table>
<thead>
<tr>
<th></th>
<th>2D</th>
<th>MD</th>
<th>CD</th>
<th>UD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Average fixation duration, ms

<table>
<thead>
<tr>
<th></th>
<th>2D</th>
<th>MD</th>
<th>CD</th>
<th>UD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300</td>
<td>320</td>
<td>340</td>
<td>360</td>
</tr>
</tbody>
</table>

No significance

Paired comparison with 2D (t-test):

MD: $t(13) = -4.06$, $p<0.05$, $p=0.0013$

CD: $t(13) = -2.68$, $p<0.05$, $p=0.019$

UD: $t(13) = -3.98$, $p<0.05$, $p=0.0016$
Conclusions Exp.3

We performed eye-tracking experiment using low-level visual stimuli and studied which factor was more influential in guiding our gaze: texture or depth (2D vs. CD or 2D vs. UD).

<table>
<thead>
<tr>
<th>Is there significant influence on visual attention? Of:</th>
<th>UD</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disparity (in comparison with 2D)</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Texture (Gray vs. Check Board)</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Position ( S1 vs. S2 vs. S3 vs. S4)</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

- Saccade length (5 s) (in comparison with 2D): NO, NO
- Fixation durations (5 s) (in comparison with 2D): YES↑, YES↑

If scene remains behind the display plane there is no difference in visual attention between 2D and 3D.

Perspective: design a test to avoid position bias.
THANK YOU
### Exp3. Scenes parameters and configurations

<table>
<thead>
<tr>
<th>Scene</th>
<th>DoF, dpt</th>
<th>Baseline, mm</th>
<th>Convergence distance, m</th>
<th>Disparity on the screen, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>UD</td>
<td>+0.15</td>
<td>315</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>MD</td>
<td>±0.1</td>
<td>500</td>
<td>8</td>
<td>-15</td>
</tr>
<tr>
<td>CD</td>
<td>-0.1</td>
<td>300</td>
<td>9.5</td>
<td>-15</td>
</tr>
</tbody>
</table>

#### Scenes

1. ![Scene 1](image1.png)
2. ![Scene 2](image2.png)
3. ![Scene 3](image3.png)
4. ![Scene 4](image4.png)
5. ![Scene 5](image5.png)
6. ![Scene 6](image6.png)
7. ![Scene 7](image7.png)
8. ![Scene 8](image8.png)
9. ![Scene 9](image9.png)
10. ![Scene 10](image10.png)
11. ![Scene 11](image11.png)
12. ![Scene 12](image12.png)
13. ![Scene 13](image13.png)
14. ![Scene 14](image14.png)

14 scenes × 4 depth configurations = 56 stimuli
Exp3. Experimental methodology

Test stages:
1. Instruction sheet
   • free-viewing task;
2. Calibration (5 point calibration + calibration chart)
3. Training 1 min 40 s
4. Test
   • 56 stimuli
   • duration 9 min 30 s
   • 5 s per image + 5 s of gray screen

28 people took part in the experiment
Conclusions Exp3

We evaluated features influencing the saliency of the objects in stereoscopic conditions by using content with low-level visual stimuli:

• We detected that texture is the most important feature for selection of objects;

• Objects with crossed disparity are significantly important for selection process, but objects with uncrossed disparity are less important for visual attention;

• Any significant difference between 2D and 3D conditions were revealed for average saccade length. Though, average fixation duration was higher when viewing stimuli with spheres in 3D; there was no difference in average fixation duration for the content with crossed disparity.
Impact of image acquisition on QoE

Optimization rules:

**Rule 1** - Optimize the stereoscopic geometry on the Region of Interest depth plane

**Rule 2** - Optimize the perceived scene to be located in the comfortable viewing zone

**Priority rule** - Rule 2 is prior to Rule 1

Depth of focus (DoF) describes the limits of the accommodative output under natural viewing conditions which concurs with the range of fusion.
Visual attention

Visual Attention: mechanism allowing to select and analyze some parts of our visual field

Visual attention is used:
- to select important areas of our visual field (alerting);
- to search a target in cluttered scenes (searching);
- allows to orient eye-movements.

There are two kinds of visual attention:
- **Overt visual attention**: involving eye movements;
- **Covert visual attention**: without eye movements. Attention can be voluntarily focused on a peripheral part of the visual field.

Types of the eye movements:
- **Saccade**: quick eye movements from one fixation location to another.
- **Fixation**: phase during which eyes is almost stationary.
Visual attention

A. Yarbus [Yarbus, 1967] demonstrated how eye movements changed depending on the question asked to the subject.

**Bottom-up attention:** some things draw attention reflexively, in a task-independent way:
- Involuntary;
- Very quick;
- Unconscious.

**Top-down attention:** some things draw attention intentionally, in a task-dependent way:
- Voluntary;
- Very slow;
- Conscious.
Motivation

Goals:

• Compare visual attention between 2D and 3D conditions using low-level visual stimuli;

• Find out which factor is more influential in guiding our gaze: texture or amount of depth.