

# A dataset of head and eye movements for 360° images

Yashas Rai, Jesús Gutiérrez, Patrick Le Callet

[jesus.gutierrez@univ-nantes.fr](mailto:jesus.gutierrez@univ-nantes.fr)

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 images/videos 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.

J-S. Lee, F. De Simone, T. Ebrahimi, "Efficient video coding based on audio-visual focus of attention", *Visual Communication and Image Representation*, vol 22, no. 8, pp. 704–711, Nov. 2011.

Q. Huynh-Thu, M. Barkowsky, P. Le Callet, "The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives", *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 421–431, Jun 2011

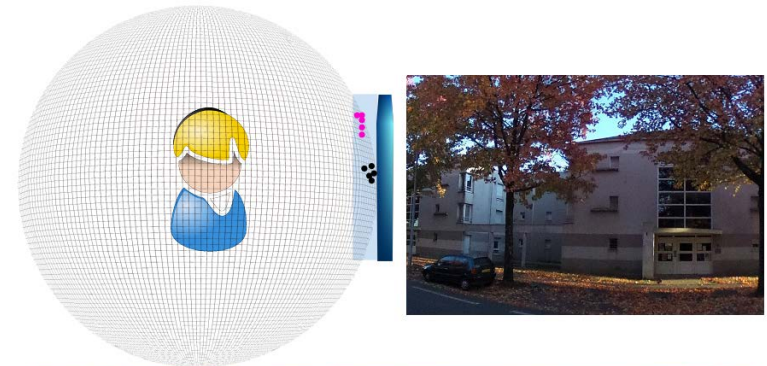
M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Pepion, "Tone mapping based HDR compression: Does it affect visual experience?," *Signal Process. Image Commun.*, vol. 29, no. 2, pp. 257–273, 2014.

# Introduction and motivation

- Even more important for **360 content**: Not everything may be seen
- Use **eye/head movement data** for:

- **Coding and streaming**: non-uniform quality streaming, use of tiles, ...

K. Kammachi Sreedhar, *et al.*, "Viewport-adaptive encoding and streaming of 360-degree video," *IEEE ISM*, Dec. 2016.



- **Quality evaluation**: weighting the metrics.

M. Yu, H. Lakshman, and B. Girod, "A Framework to Evaluate Omnidirectional Video Coding Schemes", *IEEE ISMAR*, Sep. 2015.



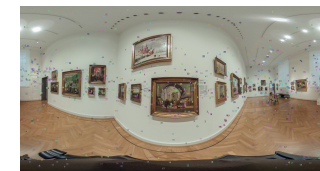
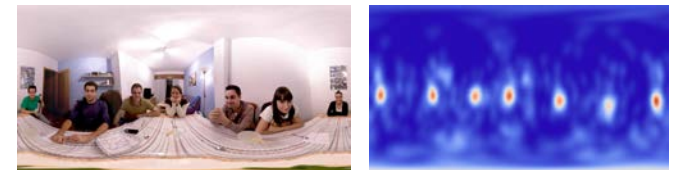
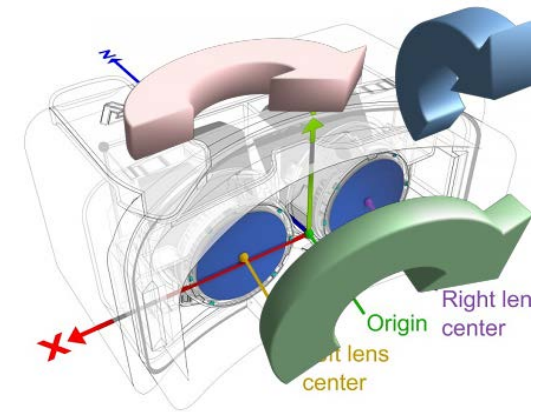
# Introduction and motivation

- **Proposed dataset:**

- Omnidirectional images.
- Exhaustive data of **head** and **eye** movements.
- Link: <https://drive.google.com/open?id=0ByRLJyAo2RFWaDFOZTIHRI9PdIU>
- Used in the ICME'17 Grand Challenge: **Salient360!**
  - Develop computational models to detect the visually salient regions for the given 360 images.

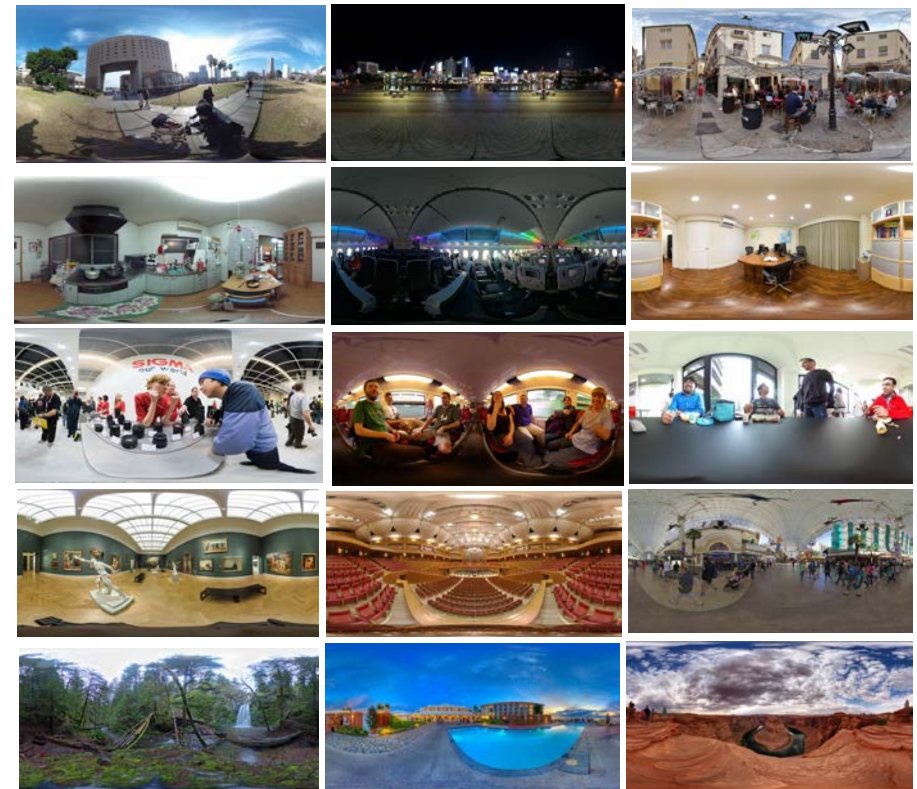
# Dataset description

- **60 images** saved in equi-rectangular format.
- Raw **eye+head** tracking data:
  - Rotational angles.
  - Translational movements.
  - [2x2] Eyes-movement within the viewport projection.
  - Time stamps of capture.
- Processed data:
  - **Head** saliency maps.
  - **Head-Eye** saliency maps.
  - **Scan-paths**.
- **Tools** for processing the data: parsing, generating and comparing saliency maps, ...



# Subjective experiment

- **Test stimuli:**
  - 98 omnidirectional images:
    - 60 released in the dataset: captured with a *Ricoh Theta S* camera.
    - 38 images from *Flicker* by professionals (CC).
    - Resolutions from 5376x2688 to 18332x9166 (equi-rectangular).
  - Wide coverage of characteristics:
    - Indoor: small rooms and wide halls.
    - Outdoor: natural landscapes and cityscapes.
    - People.



# Subjective experiment

- **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.



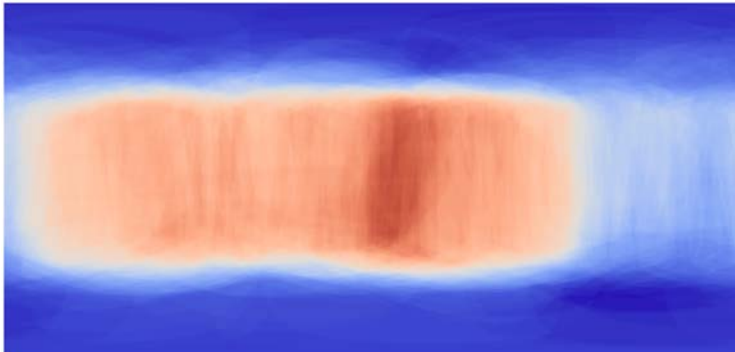


# Scan-path data



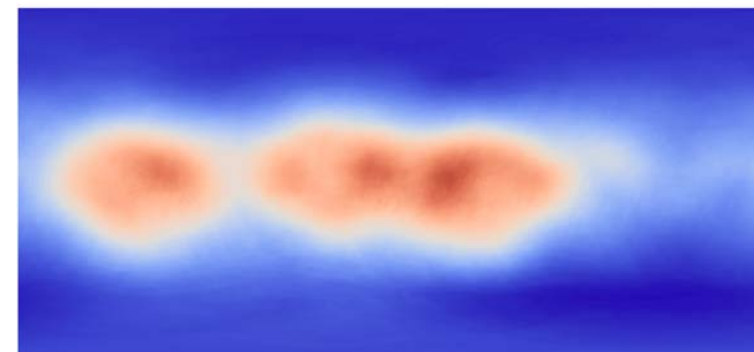
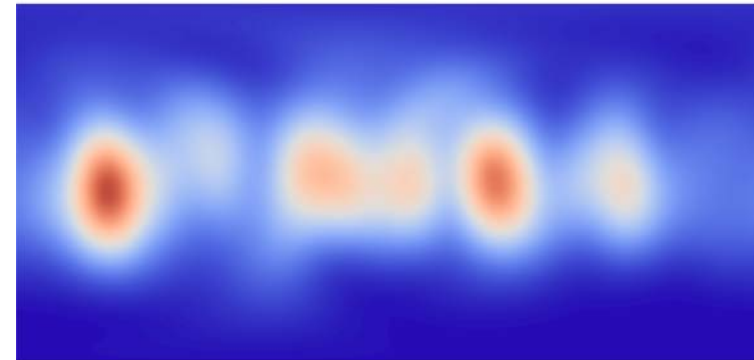
# Saliency maps

Original equi-rectangular image



Saliency map from **head movement** considering **entire viewport**

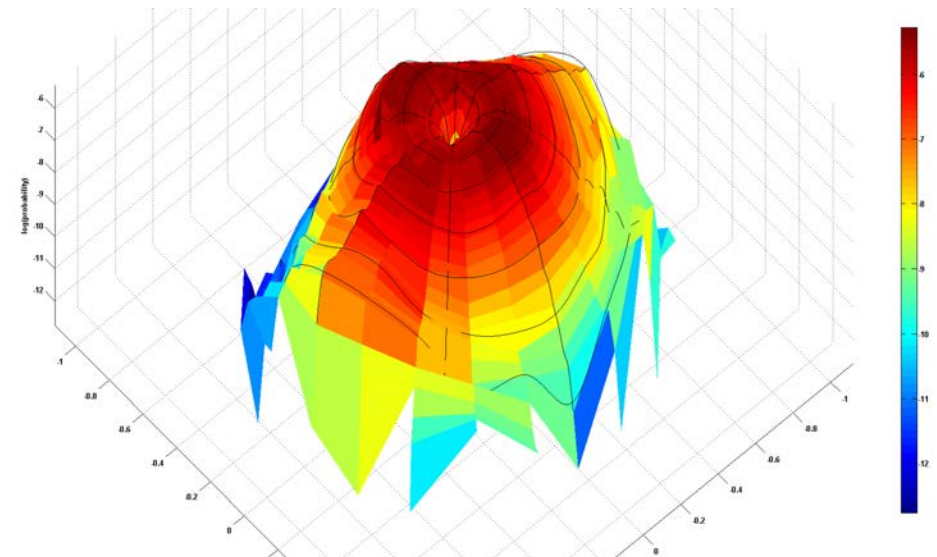
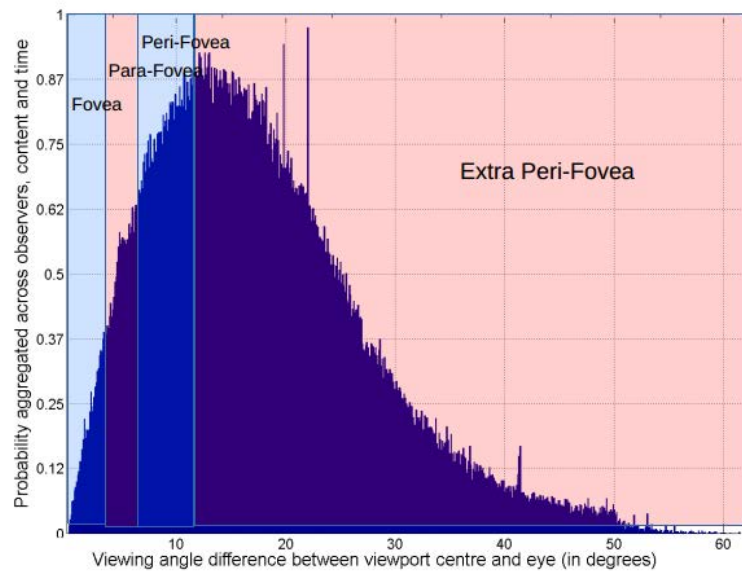
Saliency map from **head+eye** movements



Saliency map from **head movement + Gaussian** at the center of viewport

# Some statistical results

- Eye-movement within the viewport
  - Do people really look at the centre?
    - The peak is offset by 14-16 degrees from the centre.
    - Distribution has the shape of a *doughnut*: directionally isotropic.

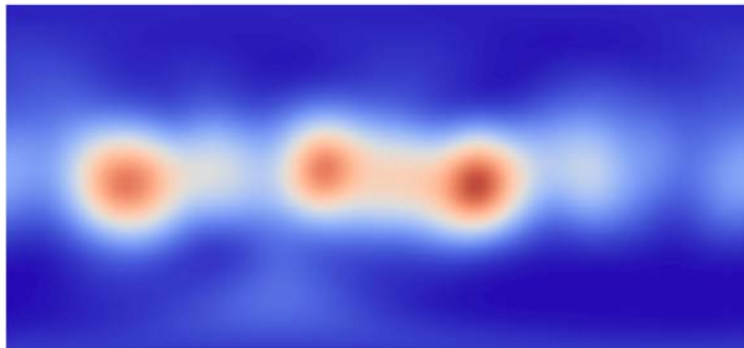
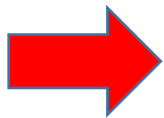
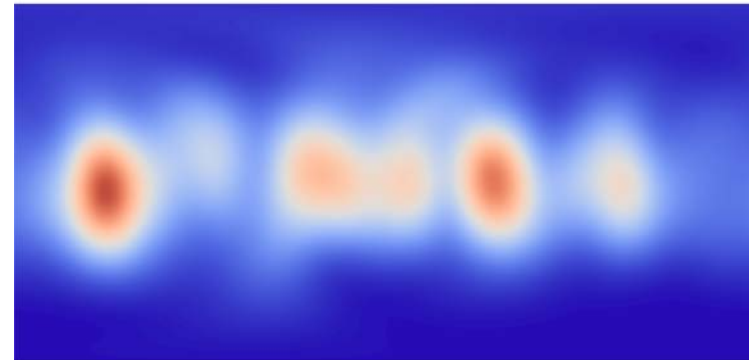


# Saliency maps

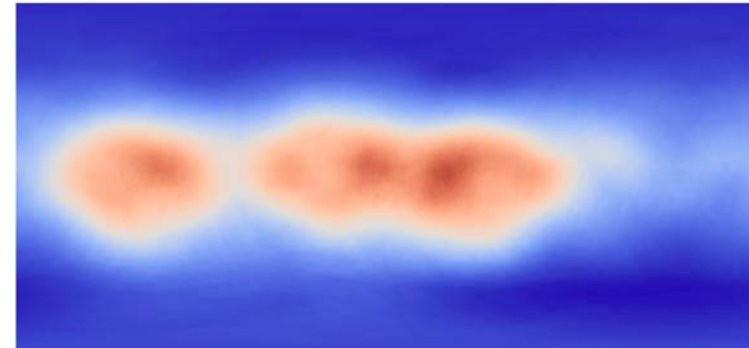
Original equi-rectangular image



Saliency map from **head+eye** movements



Saliency map from **head movement + Doughnut** at the center of viewport

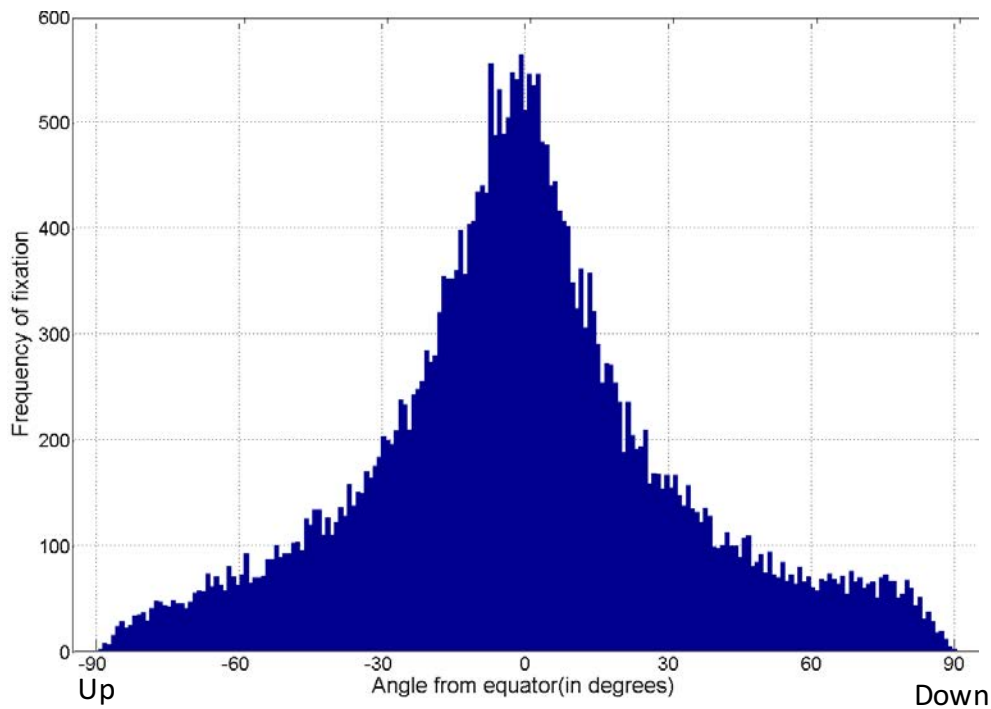


Saliency map from **head movement + Gaussian** at the center of viewport

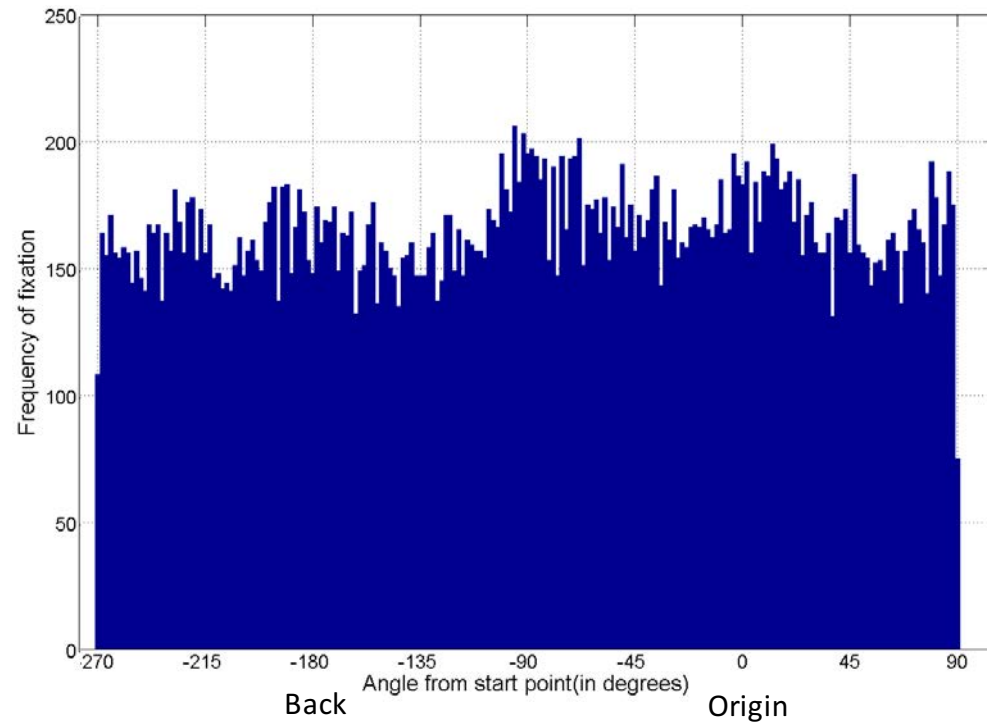
# Some statistical results

- **Scanning strategies:** frequency of fixations for a given latitude/longitude.

Latitude



Longitude

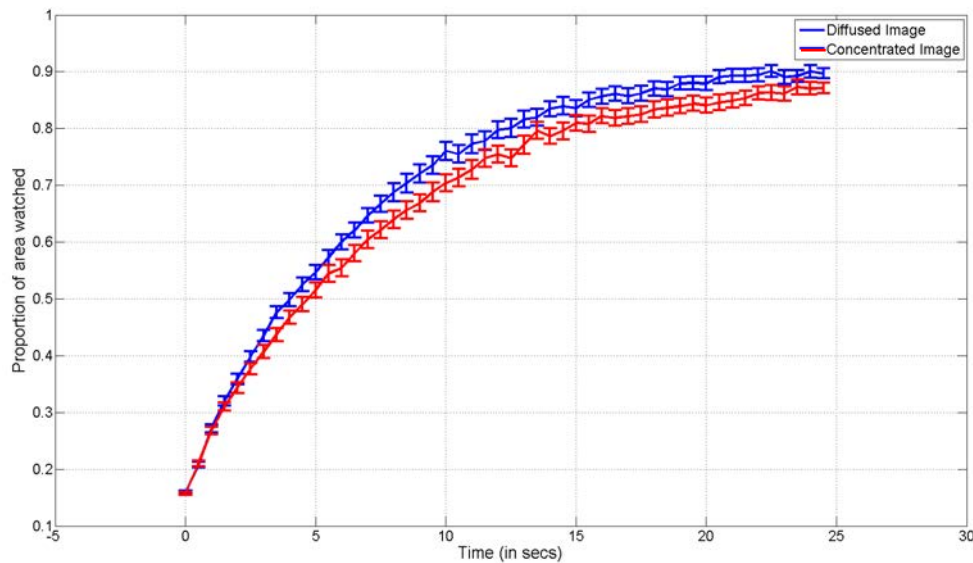


# Some statistical results

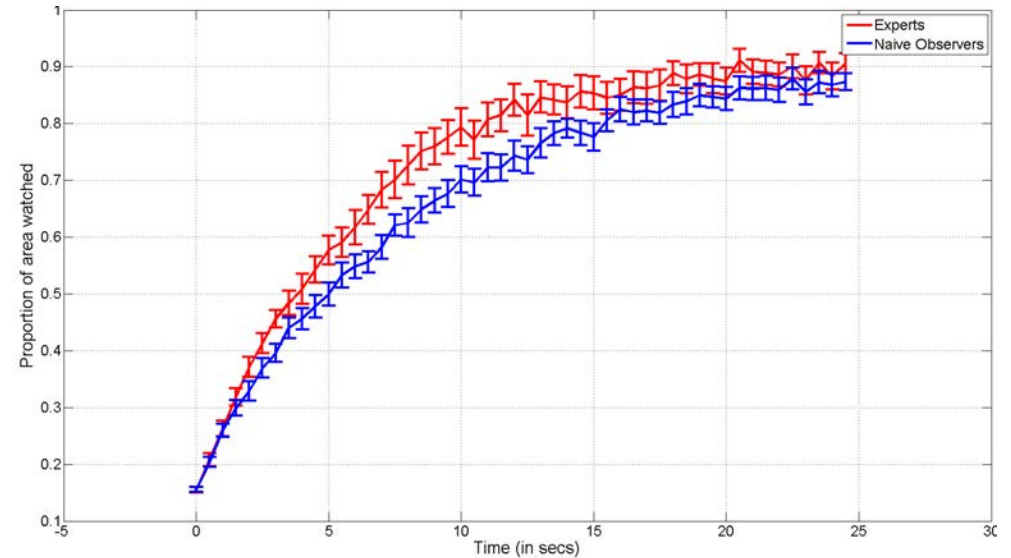
- **Speed of exploration:**

- Explored area of the sphere vs. time.
- At least ~20 seconds required for exploration (a little less for experts).

By image type



By observer type



# Conclusions

- Dataset of omnidirectional images with **head and eye movement data**.
- Subjective experiment:
  - Importance of eye tracking → **Observers explore within the viewport.**
  - Observers attend regions that are **~14 degrees away from the center of the viewport**:
    - Hypothesis of watching the center not always valid.
    - Weighting for saliency map: ***Doughnut-like* function instead of Gaussian.**
- Currently working on a similar dataset for **360° video**.

# A dataset of head and eye movements for 360° images

Yashas Rai, Jesús Gutiérrez, Patrick Le Callet

[jesus.gutierrez@univ-nantes.fr](mailto:jesus.gutierrez@univ-nantes.fr)

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