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

VQEG Meeting
20/03/2018

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.

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

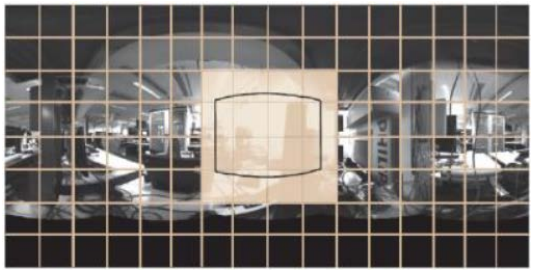
M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Pepion, "Tone mapping based HDR compression: Does it affect visual experience?," *Signal Process. Image Commun.*, vol. 29, no. 2, pp. 257–273, 2014.

Introduction and motivation

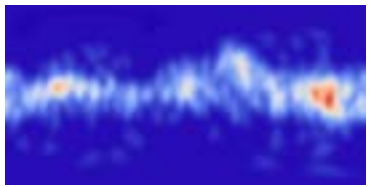
- Even more important for 360˚ content: Not everything may be seen!
  - Different scenarios compared to conventional image/video viewing:
  - No direct use of 2D VA models for 360˚ content.

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.



$$\mathrm{PSNR} = 10.\log\left(\frac{I_{\max}^2}{\frac{1}{N.M}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(error(i,j)\right)^2}\right)$$

«Saliency» weight

Introduction and motivation

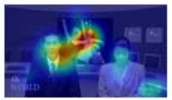
- Need of **datasets** and **benchmarking**:
  - Images:
    - MIT Saliency benchmark

NOTE: MIT Saliency Benchmark will soon switch to sorting model performances by NSS
This decision has been made at ECCV 2016 saliency tutorial. See:
Z Bylinskii, T Judd, A Oliva, A Torralba, F Durand What do different evaluation metrics tell us about saliency models? arXiv preprint arXiv:1604.03605, 2016
M KÃ¼mmerer, T Wallis, M Bethge Information-theoretic model comparison unifies saliency metrics PNAS, 112(52), 16054-16059, 2015

Model Name	Published	Code	AUC-Judd [?]	SIM [?]	EMD [?]	AUC-Borji [?]	sAUC [?]	CC [?]	NSS [?]	KL [?]	Date tested [key]	Sample [img]
Baseline: infinite humans [?]			0.92	1	0	0.88	0.81	1	3.29	0		
Deep Gaze 2	Matthias Kümmerer, Thomas S.A. Wallis, Matthias Bethge. DeepGaze II: Reading fixations from deep features trained on object recognition [arXiv 2016]		0.88 (0.84)	0.46 (0.43)	3.98 (4.52)	0.86 (0.83)	0.72 (0.77)	0.52 (0.45)	1.29 (1.16)	0.96 (1.04)	first tested: 26/11/2015 last tested: 13/09/2016 maps from authors (model without center bias in parentheses)	
SALICON	Xun Huang, Chengyao Shen, Xavier Boix, Qi Zhao		0.87	0.60	2.62	0.85	0.74	0.74	2.12	0.54	first tested: 19/11/2014 last tested: 15/11/2015 maps from authors	
DeepFix	Srinivas S S Kruthiventi, Kumar Ayush, R. Venkatesh Babu DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations [arXiv 2015]		0.87	0.67	2.04	0.80	0.71	0.78	2.26	0.63	first tested: 02/10/2015 last tested: 02/10/2015 maps from authors	

- 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.*



HD_UHD_Eyetracking_Videos 37 videos viewed in HD or in UHD with eyetracking data

A new HD and UHD video eye tracking dataset composed of 37 high quality videos observed by more than 35 naive observers. This dataset can be used to compare viewing behavior and visual saliency in HD and UHD, as well as for any study on dynamic visual attention in videos. Read more.

Available at: http://ivc.univ-nantes.fr/en/

Introduction and motivation

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

M. Narwaria, *et al.* "Tone mapping based HDR compression: Does it affect visual experience?," *Signal Process: Image Commun.*, vol. 29, no. 2, pp. 257–273, 2014.

ETHyma Eye tracking data for HDR and Tone Mapped images

Eyetracker database for 11 HDR images (displayed on an real HDR display) and 88 tone mapped images with 37 observers. Image resolution is between 760x900 and 1620x1080 pixels. Read more.

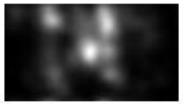
  - 3D TV Images/videos

J. Wang *et al.*, "A computational model of stereoscopic 3D visual saliency", *TIP* 2013.

Eyetracking_For_Stereoscopic_Videos Fixation density maps on 3D videos

This database contains 47 HD stereoscopic videos, theirs associated gaze point map and theirs raw eye tracking data of 40 observers. Read more.

Fang *et al.*, "Visual Attention Modeling for Stereoscopic Video: A Benchmark and Computational Model", *TIP* 2017.



3D_Gaze Free task eye tracking experiment on 3D images

Eye-tracking data of 35 observers in free task for 18 original 3D still images in HD format from the NAMA3DS1 and Middlebury databases. Read more.

  - **360˚ content?**

Available at: http://ivc.univ-nantes.fr/en/

State of the art

- VA studies for 360˚ content:
  - G. Marmitt, A. Duchowski, "Modeling visual attention in VR: Measuring the accuracy of predicted scanpath", Eurographics 2002.
  - M. Yu *et al.,* "A Framework to Evaluate Omnidirectional Video Coding Schemes", ISMAR2015.
  - Y-C. Su, et al., "Pano2Vid: Automatic cinematography for watching 360 videos", ACCV2016.
  - B. Hu *et al.*, "Head movements during visual exploration of natural images in virtual reality", CISS 2017.
  - ✓ **Y. Rai *et al*. 2017. Which saliency weighting for omni directional image quality assessment?", QoMEX 2017.**
  - ✓ V. Sitzmann *et al*., "Saliency in VR: How do people explore virtual environments?", IEEE Transactions on Visualization and Computer Graphics, 2018.
  - E. Upenik, T. Ebrahimi. "A simple method to obtain visual attention data in head mounted virtual reality", ICMEW 2017.
  - C. Wu, *et al.,* "A Dataset for Exploring User Behaviors in VR Spherical Video Streaming", ACM MMSys'17.
  - ✓ A. Serrano *et al.,* "Movie editing and cognitive event segmentation in virtual reality video", ACM TOG 2017.
  - X. Corbillon *et al.* "Viewport-adaptive navigable 360-degree video delivery", ICC2017.
  - …

Legend:
Green: Videos
Blue: Images
- : Head tracking
✓ : Head-Eye tracking

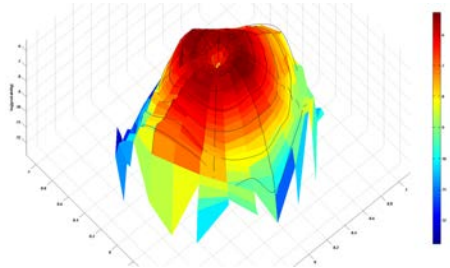
State of the art

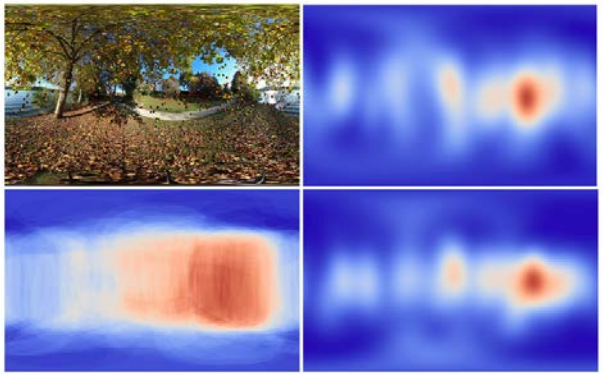
- Datasets for 360˚ content:
  - ✓ **Y. Rai, *et al.*, "A dataset of head and eye movements for 360 degree images ", ACM MMSys2017.**
  - ✓ V. Sitzmann *et al.*, "Saliency in VR: How do people explore virtual environments?", IEEE Transactions on Visualization and Computer Graphics, 2018.
  - - X. Corbillon, *et al.*, "360-Degree Video Head Movement Dataset", ACM MMSys'17 → [7 videos from Youtube, 70 seconds, 59 observers]
  - - B. Li, et al.. "A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures", Frontiers in Psychology, Dec. 2017 → [73 videos from Youtube, 29-668 seconds, 95 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.

Legend:
Green: Videos
Blue: Images
- : Head tracking
✓ : Head-Eye tracking

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
      - **Y. Rai *et al*. 2017. Which saliency weighting for omni directional image quality assessment?", QoMEX 2017.**
    - Coding, streaming, foveated rendering, movie editing, cinematography, etc.
      - A. Serrano *et al.*, "Movie editing and cognitive event segmentation in virtual reality video", ACM TOG 2017.





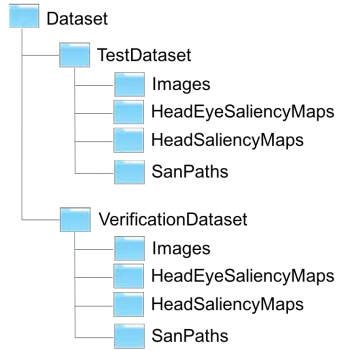
Dataset of 360° images

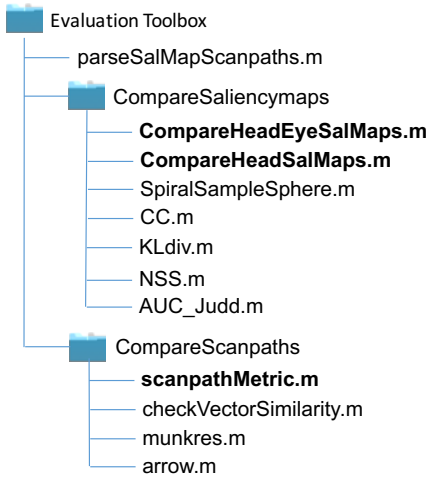
- Dataset for still images:
  - 85 equirectangular images.
  - Processed data from **head and eye movements**:
    - Head saliency maps.
    - Head-Eye saliency maps.
    - Scanpaths (Head-Eye).
  - Tools:
    - Parsing the data:
    - Compare saliency maps.
    - Compare scanpaths.
  - To access it: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

- Y. Rai, J. Gutiérrez, and P. Le Callet, "A dataset of head and eye movements for 360 degree images", *ACM MMSys2017*.
- Y. Rai, P. Le Callet, and P. Guillotel, "Which saliency weighting for omni-directional image quality assessment?", *QoMEX*2017, pp. 1–6.
- Presented in VQEG Meeting at Netflix, Los Gatos, US, May 2017.

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

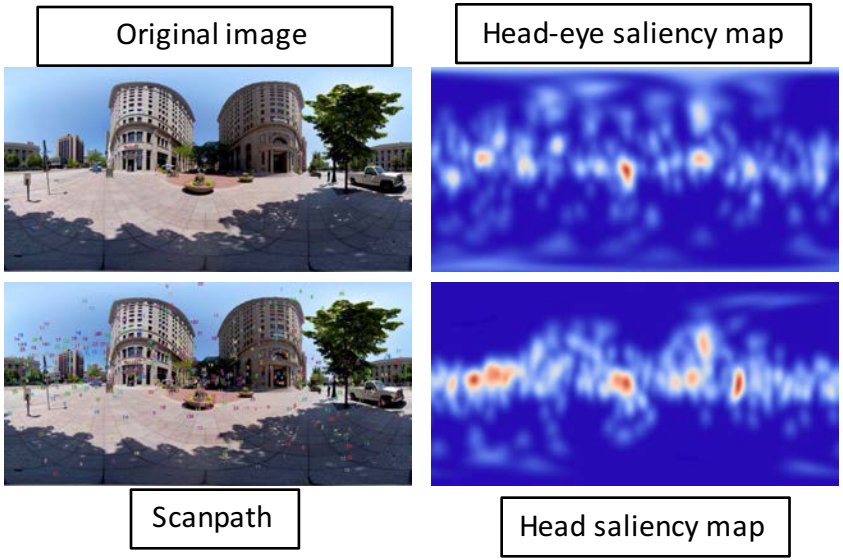




- Y. Rai, J. Gutiérrez, and P. Le Callet, "A dataset of head and eye movements for 360 degree images", *ACM MMSys2017*.
- Y. Rai, P. Le Callet, and P. Guillotel, "Which saliency weighting for omni-directional image quality assessment?", *QoMEX*2017, pp. 1–6.
- Presented in VQEG Meeting at Netflix, Los Gatos, US, May 2017.

Dataset of 360° images

- Dataset for still images:
  - 85 equirectangular images.
  - Processed data:
    - Head saliency maps → Model Type 1: 9 teams
    - Head-Eye saliency maps → Model Type 2: 11 teams
    - Scanpaths (Head-Eye) → Model Type 3: 3 teams
  - 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.
- **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.
- **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.

Salient360! – 2017

- Results

TABLE I
RESULTS FOR MODEL TYPE 1

Team	KL	CC	Rank KL	Rank CC	Rank Avg.
Zhejiang University [11]	0.44	0.69	1	1	1
Wuhan University [12]	0.51	0.71	2	1	1.5
SJTU [13]	0.65	0.67	3	3	3
TU Munich (1) [14]	0.75	0.62	6	3	4.5
TU Munich (2) [14]	0.74	0.60	3	6	4.5
TU Munich (3) [14]	0.64	0.56	3	6	4.5
Roma Tre University [15]	0.81	0.52	6	6	6
USTC	1.05	0.67	9	3	6
Anonymous	0.92	0.58	9	6	7.5
TU Munich (4) [14]	0.91	0.50	6	11	8.5
Jiangxi University of Finance and Economics [16]	1.14	0.57	13	6	9.5
TU Munich (5) [14]	1.09	0.44	9	11	10
IRISA	1.07	0.41	9	13	11

TABLE II
RESULTS FOR MODEL TYPE 2

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

TABLE III
RESULTS FOR MODEL TYPE 3

Team	SimMetric	Rank SimMetric
Insight Centre for Data Analytics - UPC [18]	2.8697	1
SJTU [13]	4.6565	2
Wuhan University [12]	5.9517	3

Salient360!: Visual Attention for 360˚ Content

- **Salient360!** – Grand Challenge at ICME'18:
  - 360˚ content: images and video. NEW
  - 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). NEW

- More info:
  - Webpage: 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 an 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

VQEG Meeting
20/03/2018