How many observers do you need to create a reliable saliency map in VR attention study?

And what metrics and baselines you may use to assess VR saliency map

Andrey Bolshakov, Maria Gracheva,

Dmitry Sidorchuk

Institute for Information Transmission Problems (Kharkevich Institute)
Russian Academy of Sciences, Moscow, Russia;
e-mail: a.bolshakov@iitp.ru

40th European Conference on Visual Perception, Berlin, 2017

Institute for Information Transmission Problems (Kharkevich Institute)

Issue

Attention and saliency investigation are becoming more popular.

However, most of researches are using framed content, presented on a monitor.

But what if you have unusual viewing conditions, for example virtual reality (VR) helmet?

The main questions while conducting VR saliency study are:

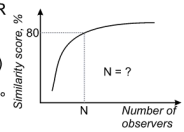
- How many observers do you need to make a reliable saliency map in 360°?
- What metrics should be used to assess 360° VR saliency prediction?
- What baselines for those metrics should be?

Puproses

1) Develop eye movement database for 360° VR video content

2) Find an 80% prediction point (similarity score)

3) Reimplement metrics and baselines for 360° VR saliency modelling estimation



Experiment

Saliency map

Source picture

Saliency map

Figures are from mathworks.com

Saliency map is a map of probabilities to attract the human gaze (visual saliency).

The most representative saliency maps may be obtained by integration of gaze tracks of real observers.

Content

20 short 360° VR videos (average duration - 99 sec)

1st day	2nd day
SET I 5 moving camera videos 5 stable camera videos	SET II 5 moving camera videos 5 stable camera videos
17 min of video ≈ 40 min in VR helmet	16 min of video ≈ 40 min in VR helmet

Total video duration: 33 min

Equipment

- SMI Mobile Eye Tracking HMD based on Samsung Gear VR
- Samsung Galaxy S7

Subjects

92 subjects (52 male, 40 female), young adults (mostly students of IITP RAS)

Most participants were incleaned to have simulator sickness, so we devided tracking procedure into two days.

Screenshots: examples of videoset

Result 1:

Database description

Content:

- 20 samples of 360° VR videos: 10 videos with moving camera and 10 videos with stationary camera.
- Panoramic landscapes, extreme sport videos, moving in some environment videos, etc (recorded by the IITP RAS).
- All videos are monoscopic (no binocular disparity).
- Average duration: 99 seconds.
- Total duration of video set: 33 min 06 sec.
- Total duration of all recordings: about 50 hours

Equipment:

- Samsung Gear 360 camera
- SMI Mobile Eye Tracking HMD based on Samsung Gear VR

Subjects:

92 subjects (52 male, 40 female), young adults.

The instruction for the experimenters were not to inform subjects about eye-tracking functionin the helmet until the end of recordings to not affect the recording data.

Procedure:

- Videos were presented in a quasirandom order in each set.
- Starting viewpoint (the initial angle of viewing in 360° of horizontal rotation) was randomized for each video and participant to minimize its influence.
- Each video was preceded by calibration procedure.

Database requesting: a.bolshakov@iitp.ru

Result 2:

Estimation of minimal number of observers

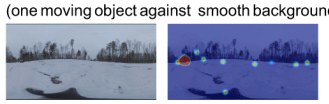
It is accepted [1, 2] that the minimum number of observers required to create a reliable saliency map begins from 8 depending on content type. However, this number has been estimated for traditional content, so it should be verified for the VR helmet case.

We have found that required number of observers for VR is varying from 25 to 45 depending on content type. As a border of reliability we used the number covering fixations of all subjects with probability of 80%.

The results obtained could be used by researchers to plan experimental procedures for VR investigations.

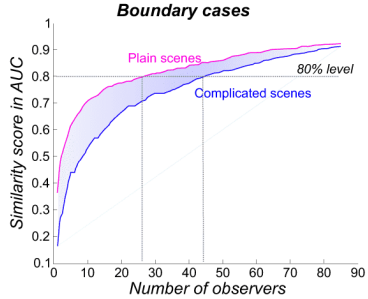
Figure shows boundary cases:

Plain scene
(one moving object against smooth background)

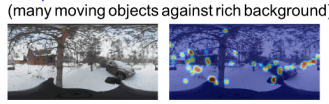


Video frame *Heat map*

Boundary cases



Complicated scene
(many moving objects against rich background)



Video frame *Heat map*

Result 3:

Metric and baseline for 360° saliency modelling estimation

Metrics & baselines intro

Saliency prediction is the most popular branch of saliency research. It is obvious that to assess the quality of prediction you need 1) metrics and 2) baselines for those metrics. For static images and usual videos there are several popular metrics, for example AUC and EMD, and the main baselines for this metrics are "chance" and "center" [4, 5].

EMD with normalized distances

AUC metric does not depend on the resolution and type of image.
Hence, EMD metric from MIT benchmark cannot be directly applied to 360° VR due to:
1) Spherical image instead of a rectangular one
2) Difference in resolution of the framed and 360° image

We have developed implementation of slightly different EMD:
1) The distance between two points is calculated as a distance on a sphere (instead of a straight line on the image plane)
2) Image was rescaled so that the area of the sphere is the same as the average area of the images used in MIT300* Saliency benchmark.

Baselines

"Chance" baseline works equally well for static images and for 360° VR videos, but "center" baseline is not as good for 360° VR. For the spherical image, there is no "center".

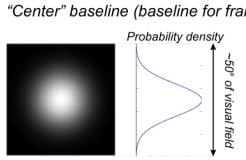
However, in practice upper and bottom parts of the sphere are less desirable locations for the viewer because they are uncomfortable to look at.
To have an internal baseline for the 360° VR videos we constructed simple method for estimating saliency.

"Equator" baseline - is a static saliency map where the saliency is constant for a fixed Y (vertical) coordinate of equirectangular projection and decreases from center to the up and bottom part of the frame.

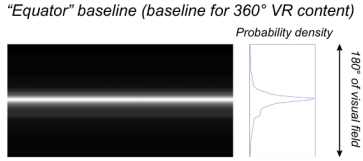
MIT benchmark:	AUC	EMD
Center	0.78	3.75
Chance	0.50	6.35

Our 360° VR baselines:	AUC	EMD
Equator	0.84	6.19
Chance	0.50	7.20

"Center" baseline (baseline for framed content)



"Equator" baseline (baseline for 360° VR content)



*We chose MIT300 dataset to get the scale of our EMD metrics because it is the most reliable and studied dataset available online with large size of held-out human fixations to estimate metrics.

Conclusions

1) The large eye movement database for 360° VR videos was gathered and may be downloaded by the request (a.bolshakov@iitp.ru).
2) The minimal number of observers for 360° VR video saliency research varies from 25 to 45 depending on content type.
3) EMD saliency metric from MIT benchmark was slightly reimplemented for 360° VR case
4) Common saliency baseline "center" were reimplemented for 360 video ("Equator" baselline).

Acknowledgments

We are very grateful to our lab team,
especially to Ivan Konovalenko and Eugeny Shvets.

REFERENCES
1. Tilke Judd,Krista Ehinger,Fre´do Durand, Antonio Torralba. Learning to predict where humans look // International Conference on Computer Vision (ICCV), 2009, pp. 2106–2113.
2. Y. Gitman, M. Erofeev, D. Vatolin, A. Bolshakov, A. Fedorov Semiautomatic Visual-Attention Modeling and Its Application to Video Compression. // IEEE International Conference on Image Processing 2014 (ICIP 2014).
3. http://cs-people.bu.edu/jmzhang/BMS/BMS_iccv13_preprint.pdf
4. http://saliency.mit.edu/baselineExp/center.html
5. http://saliency.mit.edu/results_mit300.html

FINANCIAL SUPPORT
The research was supported by the Russian Science Foundation grant (project No. 14-50-00150).

How many observers do you need to create a reliable saliency map in VR attention study?

Andrey Bolshakov, Maria Gracheva, Dmitry Sidorchuk

Attention and saliency investigations are becoming more popular with technology progress. It is widely accepted, that the minimum number of observers required to create a reliable saliency map begins from 8 depending on content type. However, this number has been estimated for traditional content, so it should be verified for the virtual reality (VR) helmet case.

Two straightforward hypotheses suggests that this number should be increased proportionally to the stimulus angular size or potential visual field (due to head rotation) enlargement. Required number for VR helmet should be about 60 observers according to the first hypothesis (typical angular size of stimuli in experiments is 45-50°) and about 30 according to the second (VR helmet field of view is about 90°).

We have recorded eye movements from 91 observers during watching 360° video content (total duration - 33 minutes) using VR helmet eye tracker based on Samsung Gear VR. We have found that required number of observers for VR is varying from 25 to 45 depending on content type to obtain saliency map covering fixations of all subjects with probability of 80% (when 8 observers give from 45% to 65%).

From the data collected it may be concluded that both hypotheses seem to be plausible: the first one may be applied for complicated scenes (many moving objects against rich background), while the second one may be applied to the plain scenes (one moving object against smooth background).

The results obtained could be used by researchers to plan experimental procedures for VR investigations.

The research was supported by the Russian Science Foundation grant (project No. 14-50-00150).

Teaser:

By collecting (using virtual reality head mounted display) and analyzing eye movement data from 91 observers watching 360° videos (33 minutes in total) it has been found that required number of observers for creating reliable saliency map is varying from 25 to 45 depending on content type.