Eye Tracking in 360: Methods, Challenges, and Opportunities

Instructors:

Eakta Jain University of Florida

Olivier Le Meur, Univ. of Rennes, CNRS, IRISA

Introduction: Olivier Le Meur

- PhD, University Of Nantes (Fr)
- HDR, French post-doctoral degree, University of Rennes 1
- Associate Professor, University of Rennes 1
- Team leader: PERCEPT / IRISA
- More than 10 years at Technicolor R&D
- Research Interests:
  - Computational modelling of visual attention
  - Image processing (quality, inpainting, HDR…)

http://www-percept.irisa.fr/

Sponsors:

Eye Tracking in 360

Topic	Learning Objectives
Part 3: Methods to generate saliency maps from eye tracking data [Le Meur]	1. Explain why 2D saliency map methods need to be generalized for omnidirectional viewing 2. Discuss the pros and cons of the selected methods 3. Compare the performance of different methods using standard metrics 4. Computational saliency models for 360° images

Generating saliency maps from eye-tracking data

- How to generate saliency map for 2D images? [Le Meur & Baccino, 2013]
  - Fixation map and Gaussian convolution

- How to generate saliency map for 360° images? [John et al., 2018]
  - Equirectangular image and Gaussian convolution
  - Equirectangular image and modified Gaussian convolution
  - Viewport and Gaussian convolution
  - Cubemaps and Gaussian convolution
  - Kent distribution, normal distribution defined on a sphere

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, *45*(1), 251-266.
John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data – the 2D case

- Discrete fixation map f^i for the ith observer:

$$f^i(\boldsymbol{x}) = \sum_{k=1}^{M} \delta(\boldsymbol{x} - \boldsymbol{x_k})$$

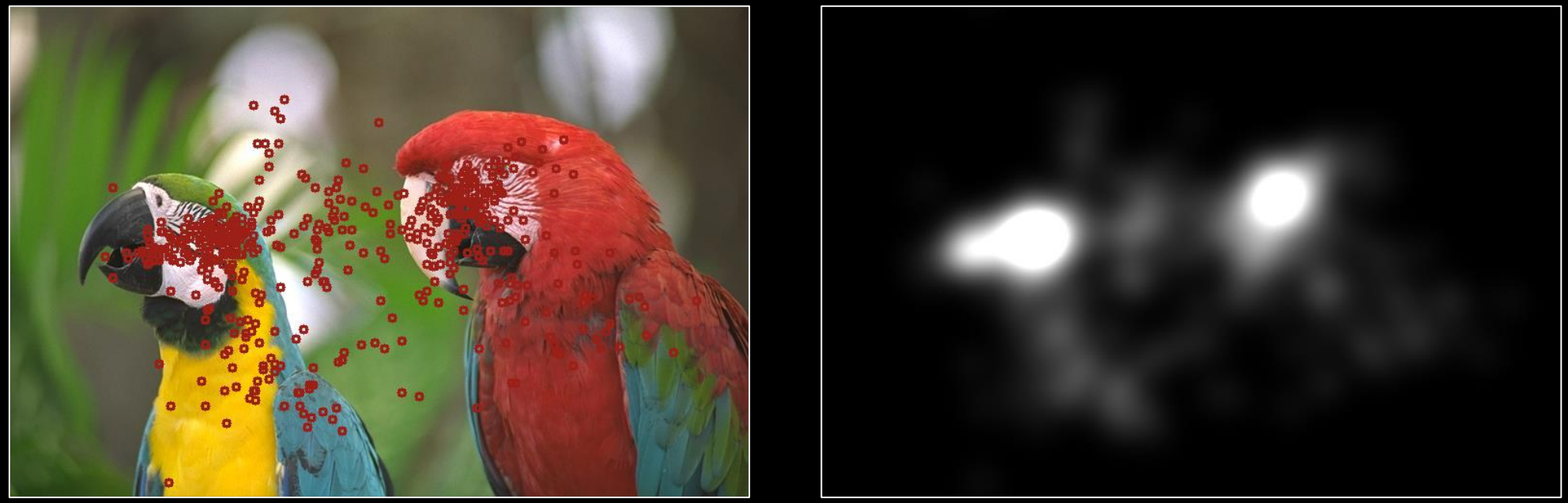
Where M is the number of fixations and $\boldsymbol{x}_k = (x_k, y_k)$ is the kth fixation.

- Continuous saliency map S:

$$S(\boldsymbol{x}) = \left(\frac{1}{N}\sum_{i=1}^{N} f^i(\boldsymbol{x})\right) * G_\sigma(\boldsymbol{x})$$

Where N is the number of observers, G_σ the Gaussian kernel and σ the number of pixels per degree of visual angle (nppd).

Generating saliency maps from eye-tracking data – the 2D case



Discrete fixation map to continuous saliency map

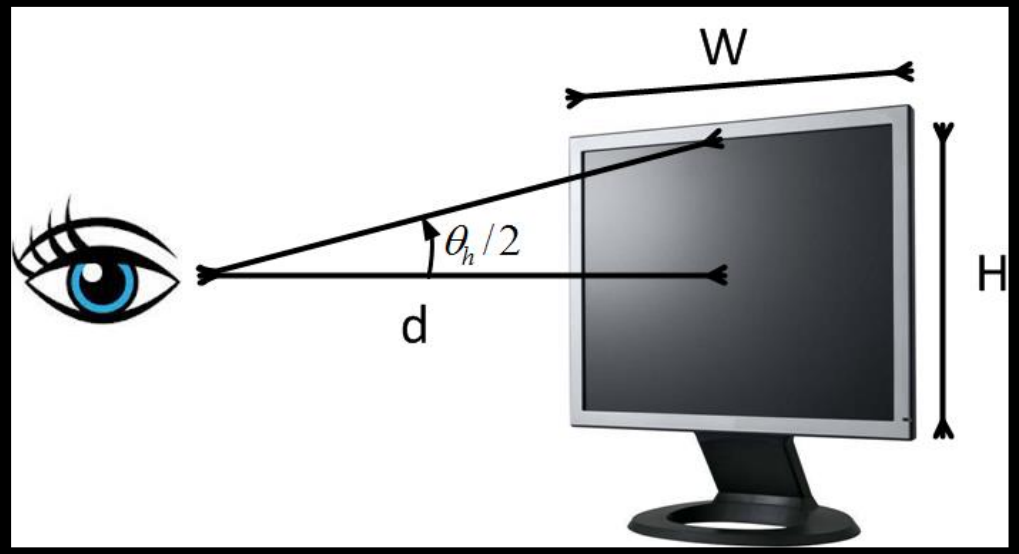
Generating saliency maps from eye-tracking data – the 2D case

- What is the nppd value?

$$\theta_H = 2 \times \arctan\left(\frac{H}{2d}\right)$$

$$\theta_W = 2 \times \arctan\left(\frac{W}{2d}\right)$$

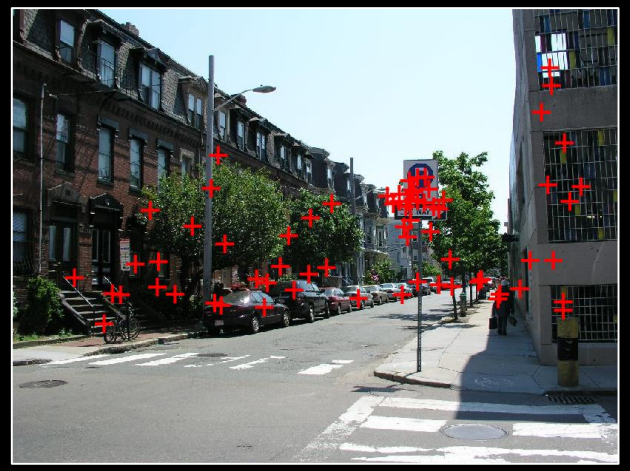
The nppd is equal to the screen resolution divided by the subtended angle



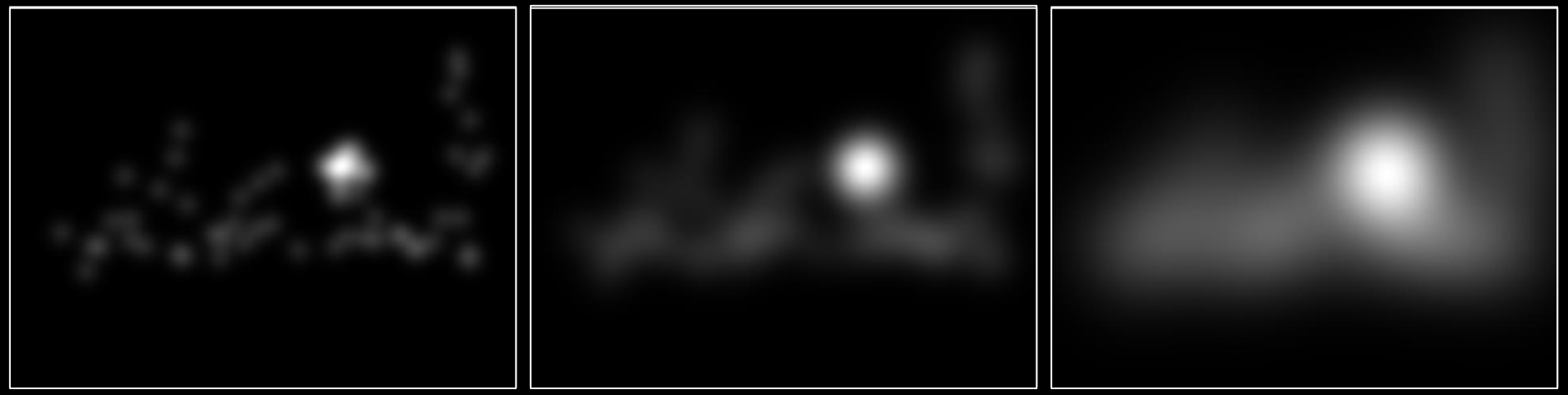
Assuming that

- the screen resolution is equal to the image resolution
- display full-screen
- the human fovea is mostly symmetric and round structure…

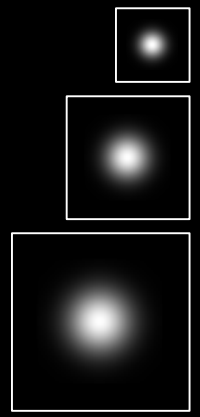
Generating saliency maps from eye-tracking data – the 2D case



From left to right, $\sigma = \{19,38,76\}$



Gaussian kernel



Generating saliency maps from eye-tracking data

- How to generate saliency map for 2D images? [Le Meur & Baccino, 2013]
  - Fixation map and Gaussian convolution

- How to generate saliency map for 360° images? [John et al., 2018]
  - Equirectangular image and Gaussian convolution
  - Equirectangular image and modified Gaussian convolution
  - Viewport and Gaussian convolution
  - Cubemaps and Gaussian convolution
  - Kent distribution, normal distribution defined on a sphere

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, *45*(1), 251-266.
John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

The question is now

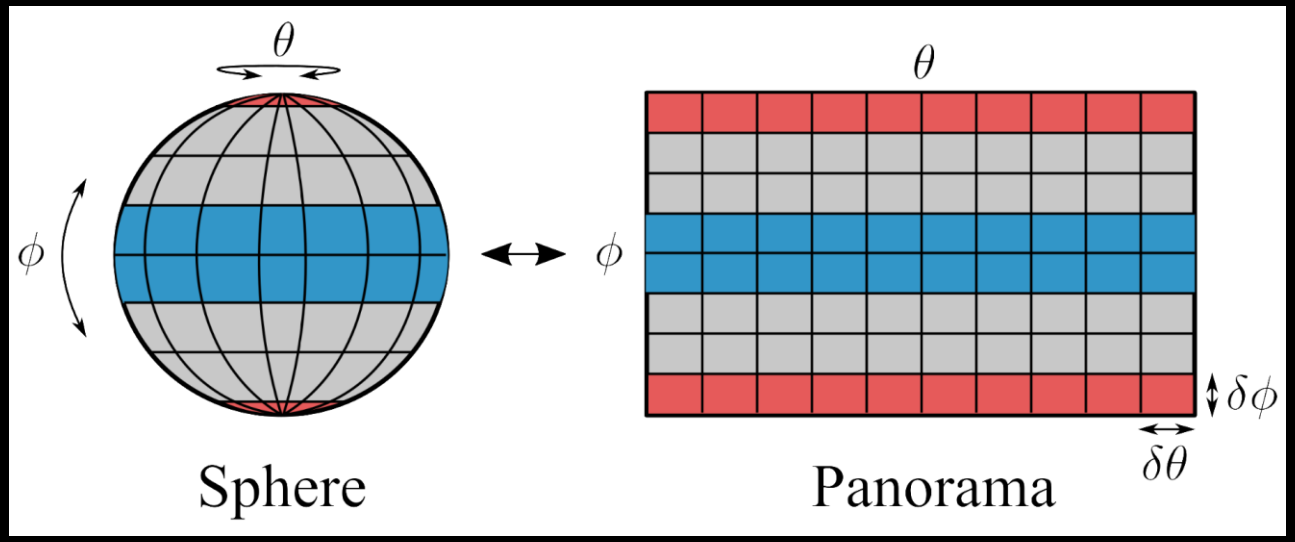
does current saliency map computation method generalize to 360° images?

[John et al., 2018]

John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- 2D isotropic Gaussian kernel on the equirectangular image [Sitzmann et al., 2018]:



Sphere and equirectangular representations under equiangular sampling.

Red and blue regions correspond to the pole and equator regions, respectively.

Horizontal angle (longitude) $\theta \in [-\pi, \pi]$, Vertical angle (latitude) $\phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

Sitzmann et al. (2018). Saliency in VR: How do people explore virtual environments?. *IEEE transactions on visualization and computer graphics*, *24*(4), 1633-1642.

Generating saliency maps from eye-tracking data in VR

- 2D isotropic Gaussian kernel on the equirectangular image [Sitzmann et al., 2018]:

Sitzmann et al. (2018). Saliency in VR: How do people explore virtual environments?. *IEEE transactions on visualization and computer graphics*, *24*(4), 1633-1642.

Generating saliency maps from eye-tracking data in VR

- 2D isotropic Gaussian kernel on the equirectangular image [Sitzmann et al., 2018]:

  1. Convert the normalized gaze tracker coordinates to longitude θ and latitude ϕ in the 360° panorama

  2. Convolve the fixation maps with a Gaussian with a standard deviation of 1° of visual angle

Sitzmann et al. (2018). Saliency in VR: How do people explore virtual environments?. *IEEE transactions on visualization and computer graphics*, *24*(4), 1633-1642.

Generating saliency maps from eye-tracking data in VR

- 2D isotropic Gaussian kernel on the equirectangular image [Sitzmann et al., 2018]:

  - ***F*=fixation map** and σ the standard deviation of the Gaussian function

From [John et al., 2018]

procedure ISOTROPIC(F,σ)

$G_\sigma \leftarrow 1d\_gaussian(\sigma, kernel\_size)$

$K \leftarrow G_\sigma \cdot G_\sigma^T$

$S \leftarrow F \circledast K$ ▷ $\circledast$ indicates 2D convolution

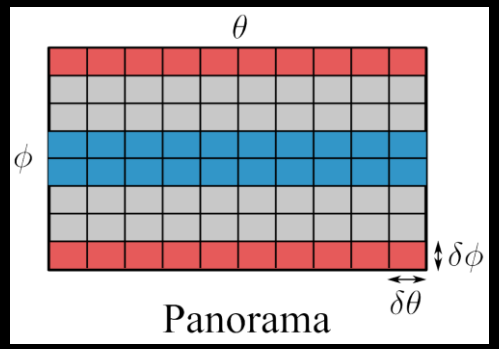
Sitzmann et al. (2018). Saliency in VR: How do people explore virtual environments?. *IEEE transactions on visualization and computer graphics*, *24*(4), 1633-1642.
John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- 2D isotropic Gaussian kernel on the equirectangular image [Sitzmann et al., 2018]:

  - This technique is fast and simple ⇒ straightforward extension of 2D

  - It does not account for distortions that are present in panorama images.
  - The equiangular sampling is not uniform.
  - Although much more smaller, the poles (red) are described the same number of pixel than the equator (blue).



Sitzmann et al. (2018). Saliency in VR: How do people explore virtual environments?. *IEEE transactions on visualization and computer graphics*, *24*(4), 1633-1642.

Generating saliency maps from eye-tracking data in VR

- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017; De Abreu et al., 2017]:
  - To account for distortions near the poles ⇒ Modified 2D Gaussian Kernel K :

$$K = G_{\sigma_y} \times G_{\sigma_x}^T$$

Where, G_{σ_n} is a column vector representing a 1D Gaussian with standard deviation σ_n.

The main idea is to adapt σ_x according to the vertical angle (e.g. elevation).

Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.
De Abreu, A., Ozcinar, C., & Smolic, A. (2017, May). Look around you: Saliency maps for omnidirectional images in VR applications. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE.

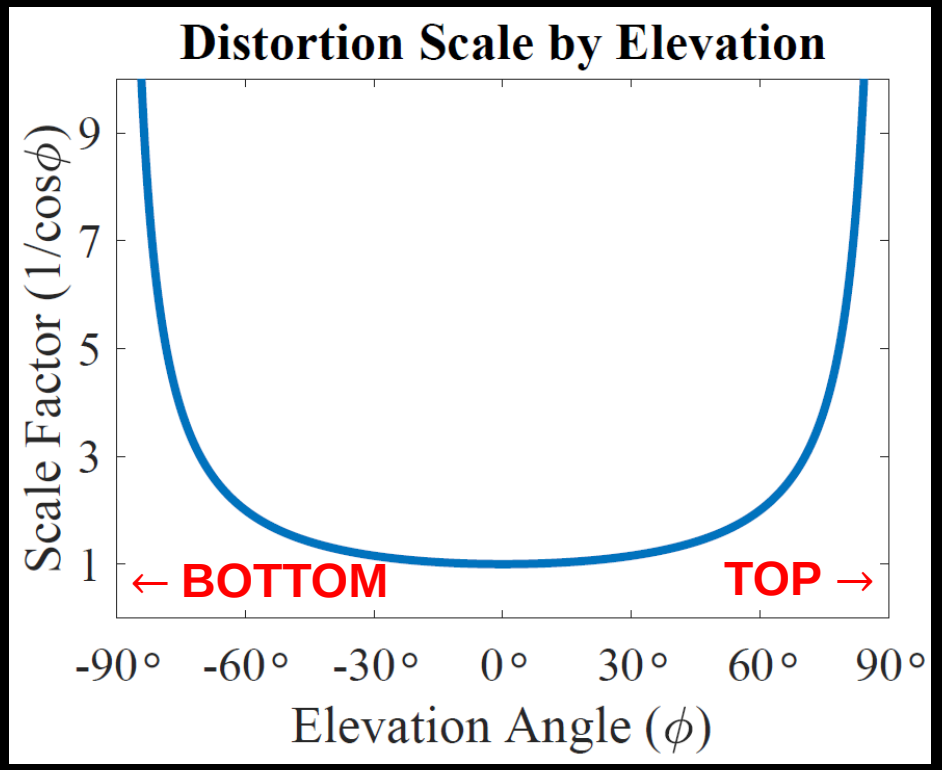
Generating saliency maps from eye-tracking data in VR

- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017]:

$$K = G_{\sigma_y} \times G_{\sigma_x}^T$$

With a scale factor to stretch horizontally the Gaussian kernel

$$\sigma_x = \frac{1}{cos(\phi)} \times \sigma_y$$



Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.

Generating saliency maps from eye-tracking data in VR

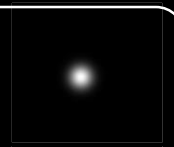
- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017]:

$$K = G_{\sigma_y} \times G_{\sigma_x}^T$$

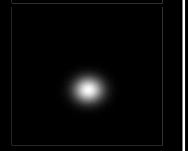
With a scale factor to stretch horizontally the Gaussian kernel

$$\sigma_x = \frac{1}{cos(\phi)} \times \sigma_y$$

$\phi = 0^\circ, \Rightarrow \sigma_x = \sigma_y$, 2D isotropic Gaussian Kernel

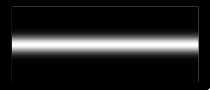


$\phi = 30^\circ, \Rightarrow \sigma_x = 1.2 \times \sigma_y$, 2D anisotropic Gaussian Kernel



$\phi = 60^\circ, \Rightarrow \sigma_x = 2.0 \times \sigma_y$, 2D anisotropic Gaussian Kernel

$\phi = 90^\circ, \Rightarrow \sigma_x \rightarrow +\infty$, 2D anisotropic Gaussian Kernel



Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.

Generating saliency maps from eye-tracking data in VR

- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017]:

From [John et al., 2018]

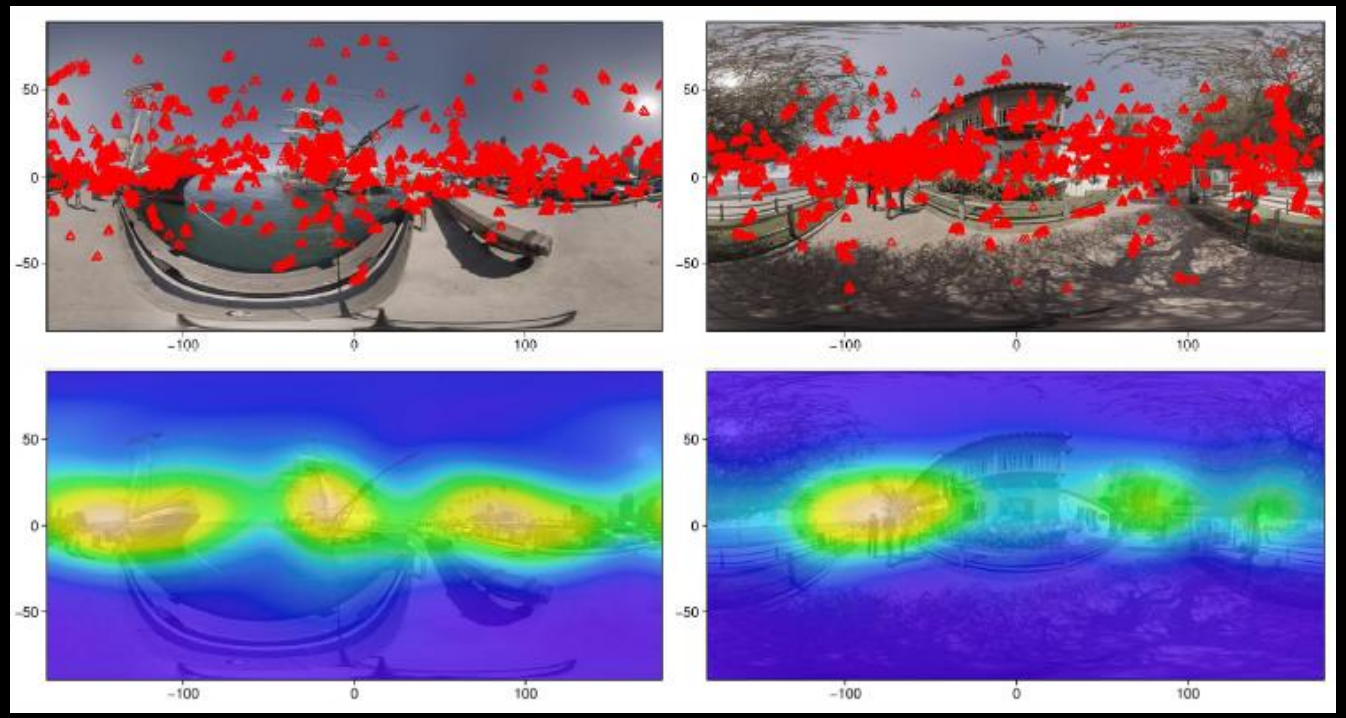
procedure MODIFIEDGAUSSIAN(F,σ)
  $S \leftarrow zeros(num\_rows, num\_cols)$
  for $r = 1$ to num_rows **do** ▷ Parallelizable loop
    $\phi \leftarrow \pi|r/num\_rows - 0.5|$
    $G_{\sigma_y} \leftarrow 1d\_gaussian(\sigma, kernel\_size)$
    $G_{\sigma_x} \leftarrow 1d\_gaussian(\sigma/cos\phi, kernel\_size)$
    $K \leftarrow G_{\sigma_y} \cdot G^T_{\sigma_x}$
    $S_{row} \leftarrow F \circledast K$ ▷ Optimized to only output row r
    $S(r) = S_{row}$

Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.

John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017]:



Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.

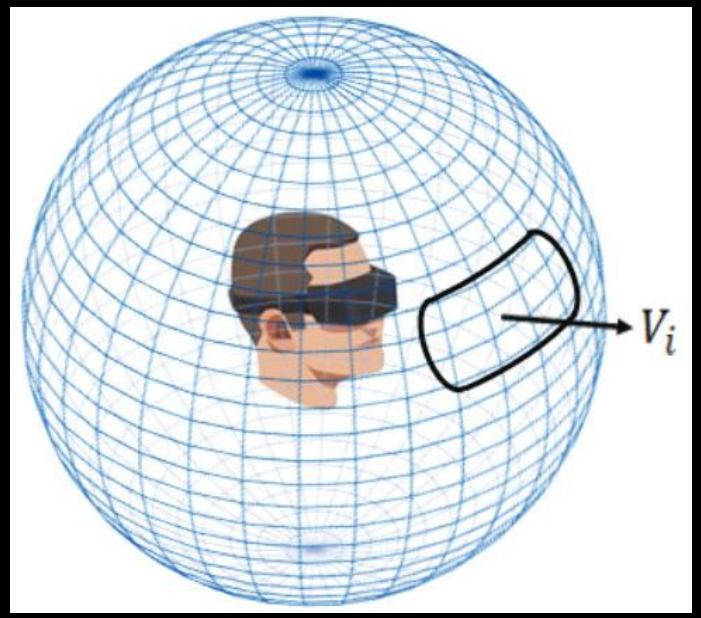
Generating saliency maps from eye-tracking data in VR

- Modified 2D isotropic Gaussian kernel on the equirectangular image [Upenik & Ebrahimi, 2017]:
  - This technique is simple to implement.
  - Time-consuming, since it requires many 2D convolutions (i.e. one per row)

Upenik, E., & Ebrahimi, T. (2017, July). A simple method to obtain visual attention data in head mounted virtual reality. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 73-78). IEEE.

Generating saliency maps from eye-tracking data in VR

- Viewport method [Rai et al., 2017]:
  - Viewport or Field of View (FoV) is a fraction of the omnidirectional image displayed on the 2D screen:
    - Head rotation + field of view (fov_x, fov_y)
  - For each fixation:
    1. a viewport is created and convolved with a Gaussian kernel.
    2. viewport is projected onto an equirectangular image.
    3. aggregation for equirectangular images.



Adapted from [Chakareski et al., 2018]

Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.
Chakareski et al., V. (2018, May). Viewport-Driven Rate-Distortion Optimized 360º Video Streaming. In *2018 IEEE International Conference on Communications (ICC)* (pp. 1-7). IEEE.

Generating saliency maps from eye-tracking data in VR

- Viewport method [Rai et al., 2017]:

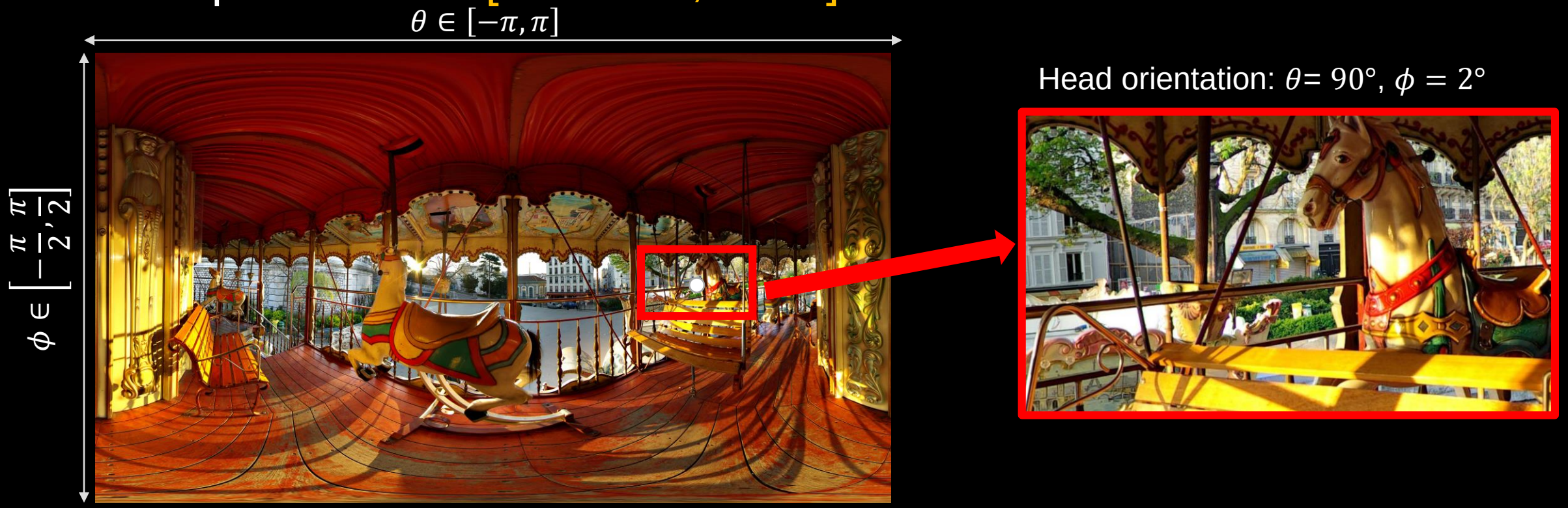
```
procedure VIEWPORT(FIXATIONS, ROTATIONS, σ_x, σ_y)
    S ← zeros(num_rows, num_cols)
    for i = 1 to M fixations do
        viewport ← zeros(viewport_size)
        x, y ← fixation_i
        R ← rotation_i
        viewport(y, x) = 1
        G_{σ_y} ← 1d_gaussian(σ_y, kernel_size)
        G_{σ_x} ← 1d_gaussian(σ_x, kernel_size)
        K ← G_{σ_y} · G^T_{σ_x}
        S_viewport ← viewport ⊛ K
        S_equirect ← vp2sphere(S_viewport, R, fov_x, fov_y)
        S += S_equirect
```

From [John et al., 2018]

Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.
John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- Viewport method [Rai et al., 2017]:

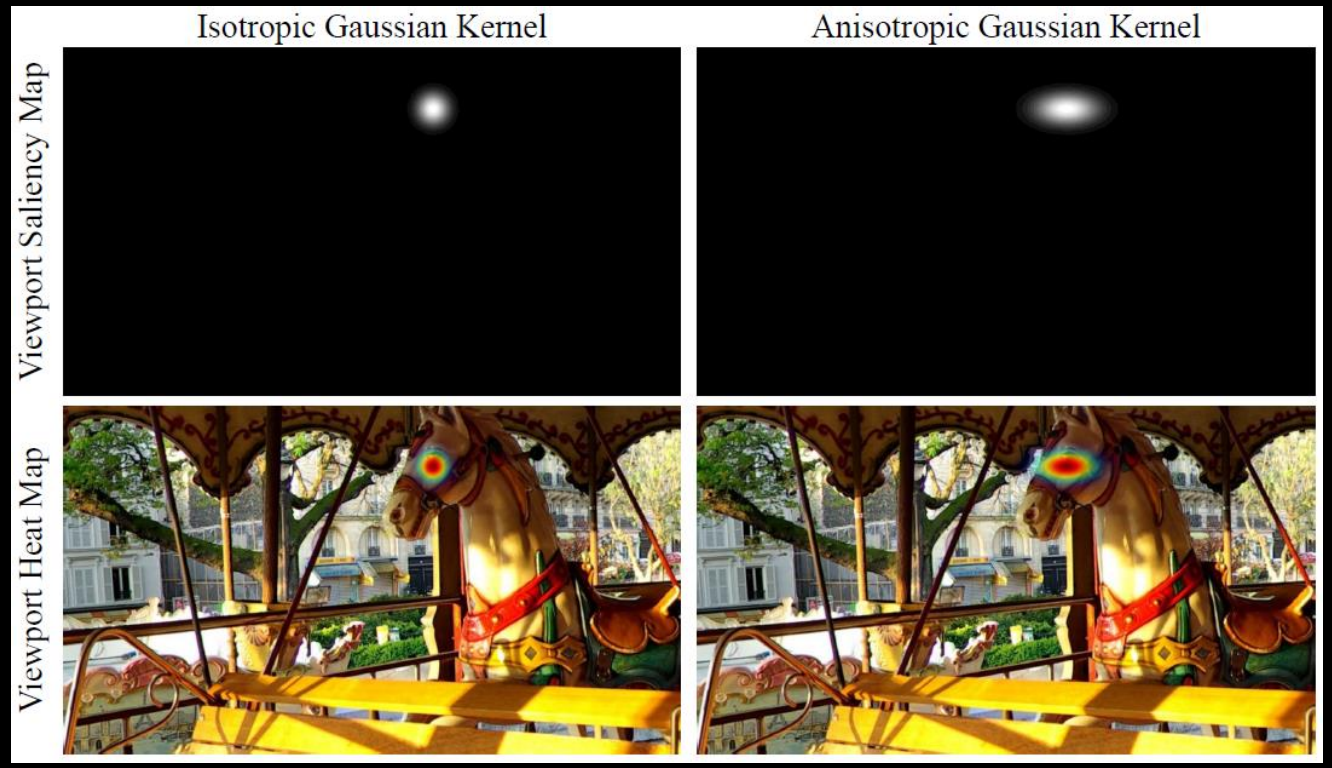


Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.

Generating saliency maps from eye-tracking data in VR

- Viewport method [Rai et al., 2017]:

Head orientation: $\theta = 90.1°$, $\phi = 1{,}7°$



Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.

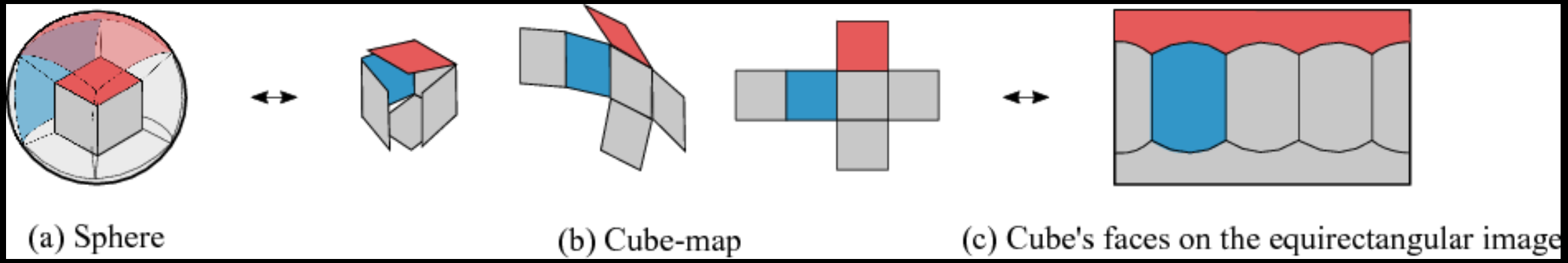
Generating saliency maps from eye-tracking data in VR

- Viewport method [Rai et al., 2017]:
  - Time-consuming methods (viewport convolution and projection, isotropic or anisotropic Gaussian kernel)
  - Does not scale well, depends on the number of fixations to process
  - Require to know the HMD's horizontal and vertical field of view

Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.

Generating saliency maps from eye-tracking data in VR

- Cube map projection [Facebook, Maugey et al., 2017]:



  - Projecting the spherical image onto cube faces
    - Cube maps' pixels are well-distributed — each face is equally important (no more equator or poles)
    - Radial distortion are less noticeable

Maugey, T. et al. (2017, October). Saliency-based navigation in omnidirectional image. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.
Facebook. Under the hood: building 360 video. https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/.

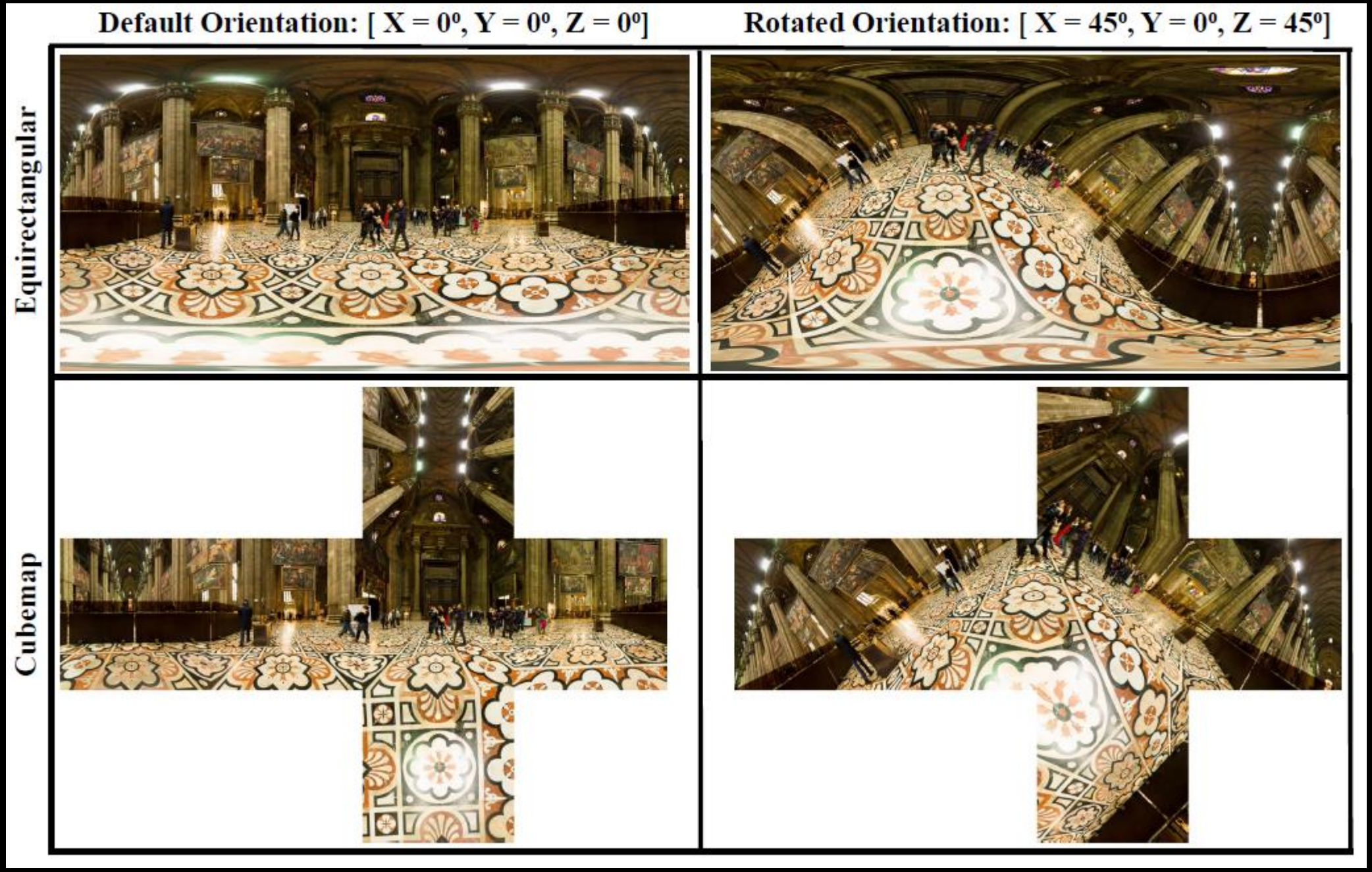
Generating saliency maps from eye-tracking data in VR

- Cube map projection [Facebook, Maugey et al., 2017]:

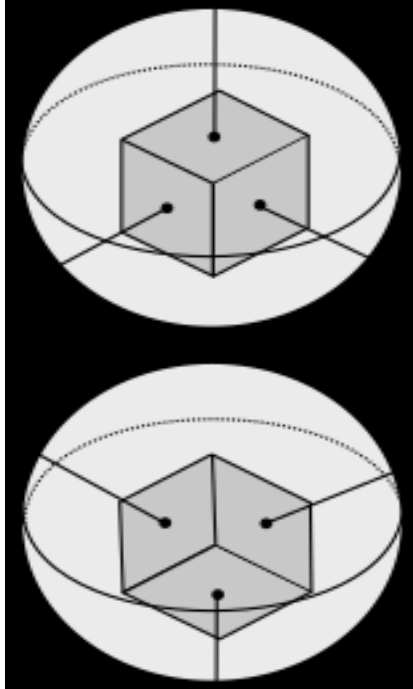
Radial distortion are less but still noticeable

Artificial borders are created between faces

Maugey, T. et al. (2017, October). Saliency-based navigation in omnidirectional image. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.



Double cube projection



Generating saliency maps from eye-tracking data in VR

- Cube map projection [Facebook, Maugey et al., 2017]:

Double cube projection

From [John et al., 2018]

procedure CUBEMAP(F,σ)

$F_{rot} \leftarrow rotatesphereXYZ(F, \frac{pi}{4}, 0, \frac{pi}{4})$

$G_\sigma \leftarrow 1d\_gaussian(\sigma, kernel\_size)$

$K \leftarrow G_\sigma \cdot G_\sigma^T$

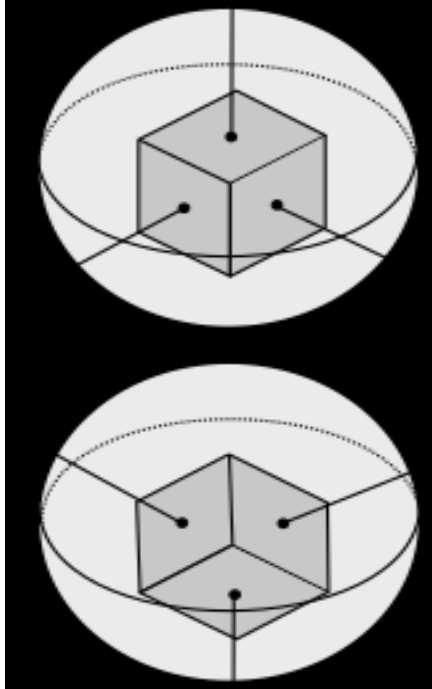
$Cube_1 \leftarrow equirect2cube(F) \circledast K$

$Cube_2 \leftarrow equirect2cube(F_{rot}) \circledast K$

$S_1 \leftarrow cube2equirect(Cube_1)$

$S_{\leftarrow} rotatesphereXYZ(cube2equirect(Cube_2), -\frac{pi}{4}, 0, -\frac{pi}{4})$

$S \leftarrow W_1 .* S_1 + W_2 .* S_2$



Maugey, T. et al. (2017, October). Saliency-based navigation in omnidirectional image. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.
John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- Cube map projection [Facebook, Maugey et al., 2017]:
  - It can be extended to use more than two orientations for a better approximation.
  - Transforming high resolution images between formats is time consuming.
  - It does not completely remove border discontinuities.

Maugey, T. et al. (2017, October). Saliency-based navigation in omnidirectional image. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.
Facebook. Under the hood: building 360 video. https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/.

Generating saliency maps from eye-tracking data in VR

- However, natural method to generate 360° saliency map is to process the fixation data directly on the sphere:
  - 2D Gaussian function cannot be used on the sphere
  - Kent distribution is the extension of the Gaussian distribution on the sphere. The isotropic form of the pdf is:

$$f(\boldsymbol{x}, k, \boldsymbol{\gamma}) = \frac{k}{4\pi \times \sinh(k)} exp(k \times (\boldsymbol{\gamma} \cdot \boldsymbol{x})),$$

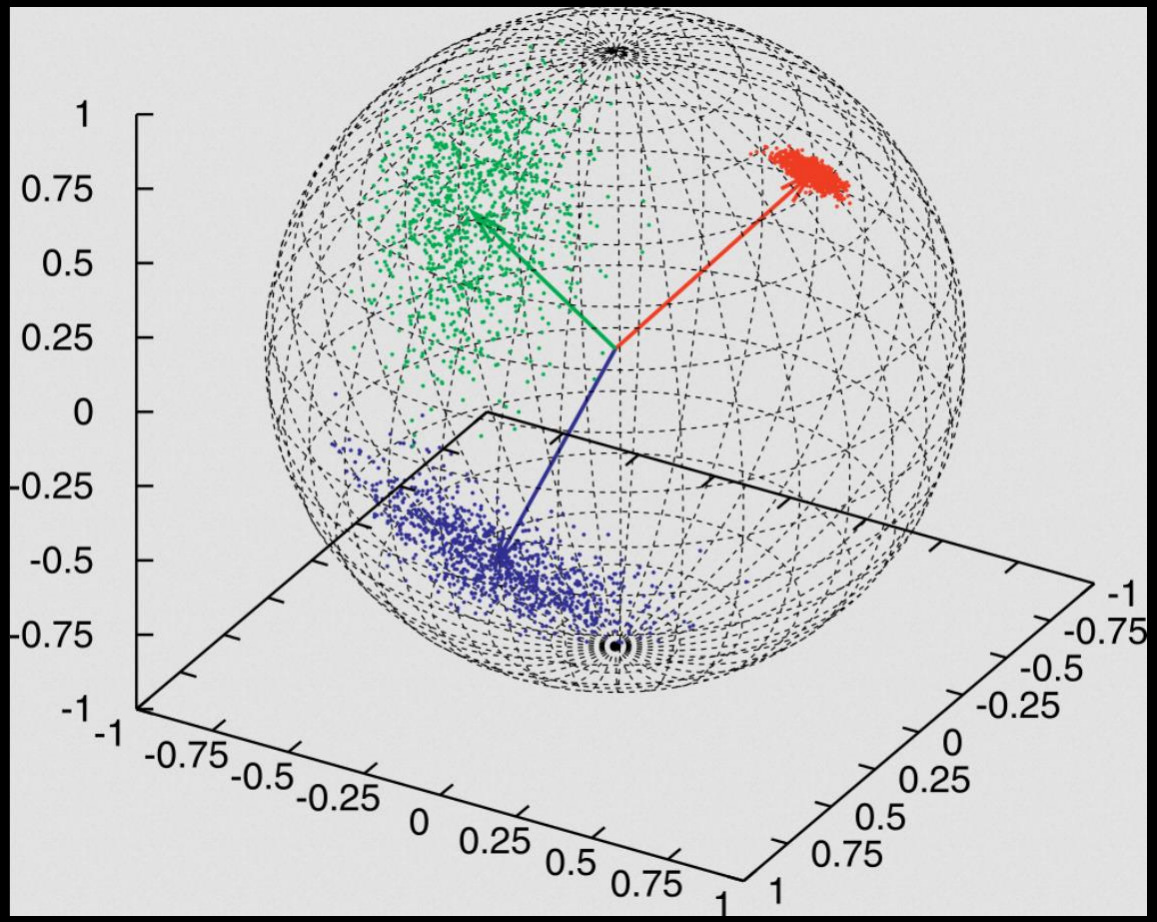
$$x, \gamma \in R^3$$

Generating saliency maps from eye-tracking data in VR

- Isotropic form of Kent distribution :

$$f(\boldsymbol{x}, k, \boldsymbol{\gamma}) = \frac{k}{4\pi \times \sinh(k)} exp(k \times \boldsymbol{\gamma} \times \boldsymbol{x}),$$
$$x, \gamma \in R^3$$

  - γ the mean direction of the distribution, around which points are normally distributed.
  - k (>0) represents the concentration of the probability density function



Three distributions, red is the most compact

Hamelryck et al. (2006). Sampling realistic protein conformations using local structural bias. *PLoS Computational Biology*, *2*(9), e131.

Generating saliency maps from eye-tracking data in VR

- Isotropic form of Kent distribution :

From [John et al., 2018]

```
procedure KENT(F, κ)
    S ← zeros(num_rows, num_cols)
    for r = 1 to num_rows do                    ▷ Parallelizable loop
        φ ← π|r/num_rows − 0.5|
        θ ← 0                                   ▷ θ is constant
        c = num_cols/2                          ▷ c is constant
        γ⃗ ← sph2cart(θ, φ, 1)
        K ← f(N(c, r, kernel_size), κ, γ⃗)
        Normalize K                             ▷ Kernel weights sum to 1
        S_row ← F ⊛ K                           ▷ Optimized to only output row r
        S(r) = S_row
```

John, B., Raiturkar, P., Le Meur, O., & Jain, E. (2018, December). *A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 136-139). IEEE.

Generating saliency maps from eye-tracking data in VR

- Isotropic form of Kent distribution :
  - Due to unequal sampling of the sphere in equirectangular images, the kernel must be recomputed for each row.
  - Required high precision floating point arithmetic, more than 64bits…
  - Very time consuming.

Generating saliency maps from eye-tracking data in VR

- The ground truth saliency maps are computed with Kent method

Which of the four methods best approximate Kent method?

Equirectangular image and Gaussian convolution
Equirectangular image and modified Gaussian convolution
Viewport and Gaussian convolution
Cubemaps and Gaussian convolution

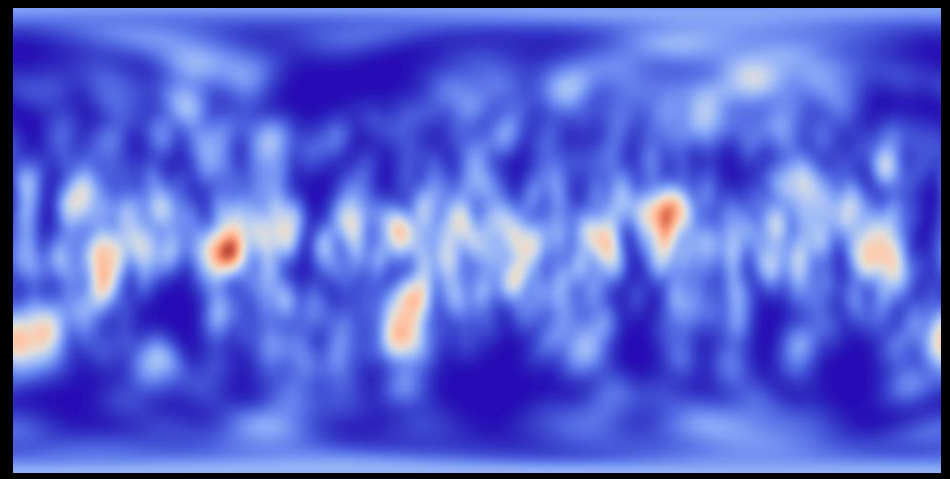
Generating saliency maps from eye-tracking data in VR

- Similarity metrics between two saliency maps [Le Meur & Baccino, 2013; Bylinskii et al., 2019]
  - Root Mean Square Error (RMSE): $RMSE = \sqrt{\frac{\sum_{\boldsymbol{x}}\left(Kent(\boldsymbol{x}) - approx(\boldsymbol{x})\right)^2}{N}}$
  - Correlation Coefficient (CC): $CC = \frac{cov(Kent, approx)}{\sigma_{Kent}\sigma_{approx}}$
  - Kullback-Leibler Divergence (KLD): $KLD = \sum_{\boldsymbol{x}} approx(\boldsymbol{x}) \times log\left(\frac{approx(\boldsymbol{x})}{Kent(\boldsymbol{x})}\right)$, assuming $\sum_{\boldsymbol{x}} approx(\boldsymbol{x}) = \sum_{\boldsymbol{x}} Kent(\boldsymbol{x}) = 1$.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, *45*(1), 251-266.
Bylinskii et al. (2019). What do different evaluation metrics tell us about saliency models?. *IEEE transactions on pattern analysis and machine intelligence*, *41*(3), 740-757.

Generating saliency maps from eye-tracking data in VR

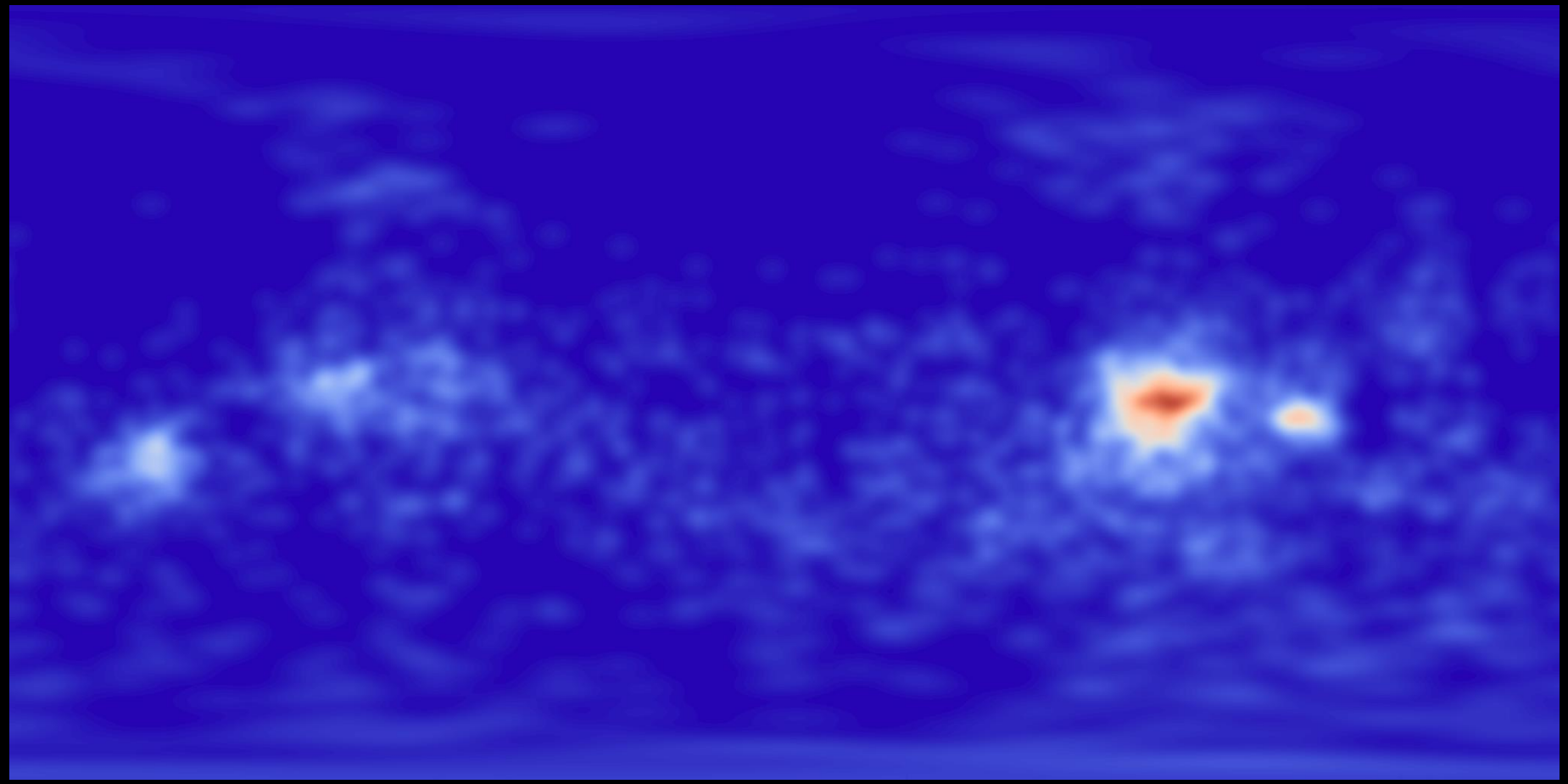
- Eye tracking dataset:
  - Salient360! Dataset [Rai et al., 2017]
  - More than 40 observers
  - 60 (360°) images (equirectangular images cover up to 18k by 9k resolution)
  - + saliency maps generated using the viewport method (isotropic Gaussian)



Rai et al. (2017, June). A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 205-210). ACM.

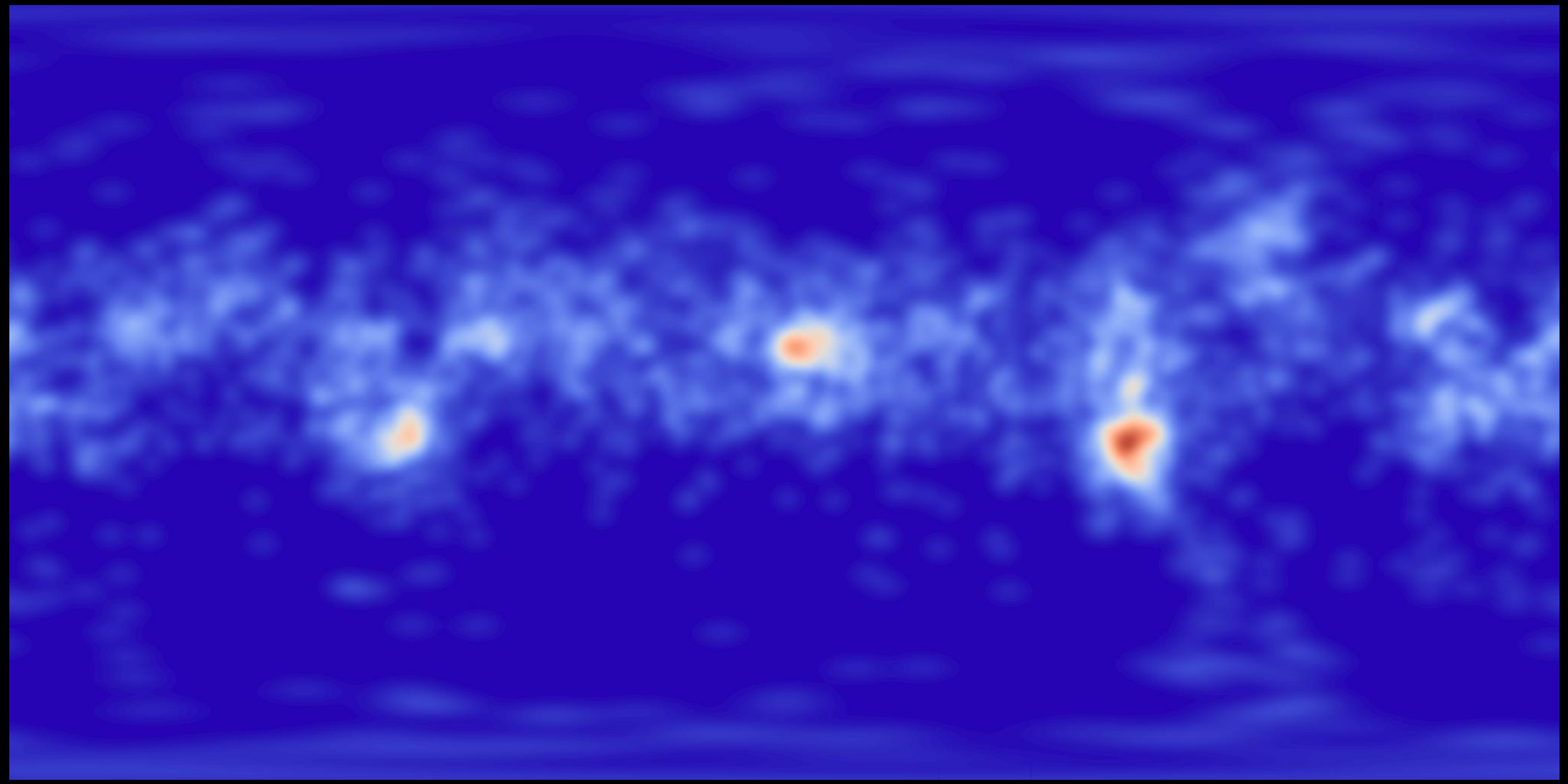
Generating saliency maps from eye-tracking data in VR

Generating saliency maps from eye-tracking data in VR



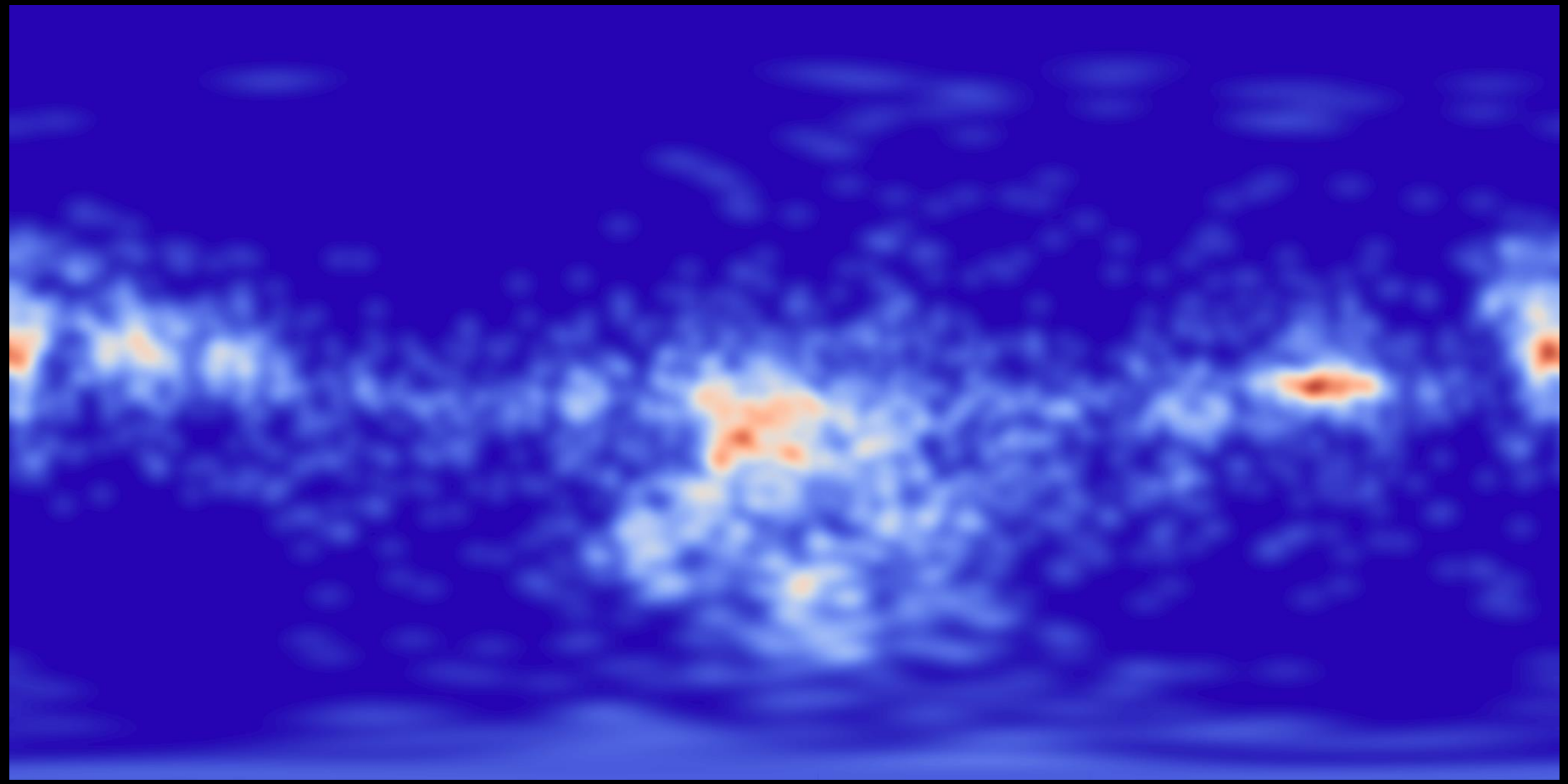
Generating saliency maps from eye-tracking data in VR

Generating saliency maps from eye-tracking data in VR



Generating saliency maps from eye-tracking data in VR

Generating saliency maps from eye-tracking data in VR



Generating saliency maps from eye-tracking data in VR

- Results

Method	Modified Gaussian [Upenik et al., 2017]	Isotropic Gaussian [Sitzmann et al., 2018]	Cubemap [Maugey et al., 2017]	Viewport [Rai et al., 2017]
Dataset	[Rai et al., 2017]	[Rai et al., 2017]	[Rai et al., 2017]	[Rai et al., 2017]
CC (↑)	0.994	0.988	0.984	0.408
KLD (↓)	0.011	0.097	0.108	0.865
RMSE (↓)	0.018	0.022	0.024	0.145
Time (s)	5.16	8.80	280	X

Generating saliency maps from eye-tracking data in VR

- Results

Kent

Viewport

[Rai et al., 2017]

Generating saliency maps from eye-tracking data in VR

- Our dataset:
  - Replicated data collection of [Rai et al. 2017]
  - Same equipment:
    - Oculus DK2 with SMI eye tracker (60 Hz)
  - Same set of images (four different starting orientations)
  - 20 observers viewed each image (25s, 6s of a gray screen between images)
  - Natural free viewing

Generating saliency maps from eye-tracking data in VR

- Our dataset:

isotropic gaussian

anisotropic gaussian

Method	Modified Gaussian [Upenik et al., 2017]	Isotropic Gaussian [Sitzmann et al., 2018]	Cubemap [Maugey et al., 2017]	Viewport [Rai et al., 2017]	Viewport Anisotropic Gaussian
Dataset	[Rai et al., 2017]	[Rai et al., 2017]	[Rai et al., 2017]	[Rai et al., 2017]	Ours
CC (↑)	0.994	0.988	0.984	0.408	0.930
KLD (↓)	0.011	0.097	0.108	0.865	0.273
RMSE (↓)	0.018	0.022	0.024	0.145	0.044
Time (s)	5.16	8.80	280	X	1031

Generating saliency maps from eye-tracking data in VR

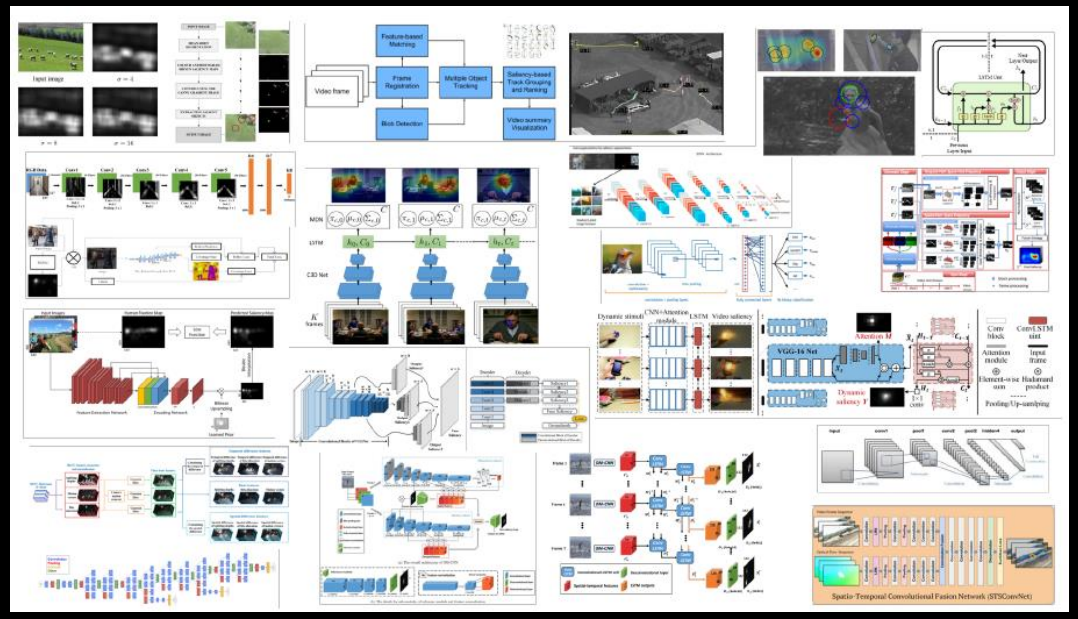
- Our conclusion & recommendation (**take-home message**):

Saliency map should be computed with the modified 2D

isotropic Gaussian kernel on the equirectangular image

Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - From an input image, models aims to compute a saliency map
  - A number of models for 2D images
    - A new breakthrough with deep architecture

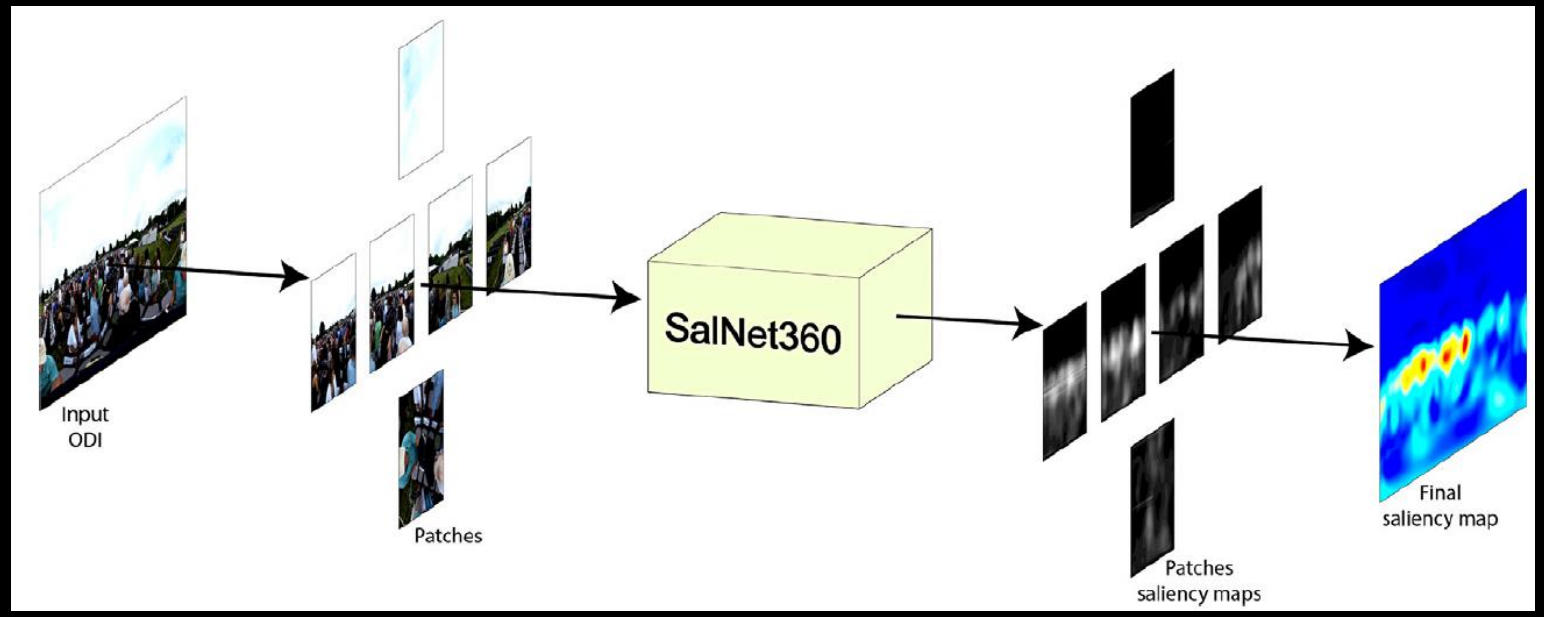


Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - Few models for 360° images:
    - SalNet 360 [Monroy et al., 2018]
    - GBVS360 and BMS360 [Lebreton & Raake, 2018]
    - SalGAN 360 [Chao et al., 2018]

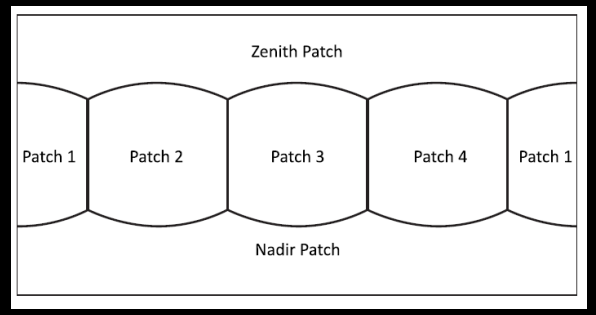
Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - SalNet360 [Monroy et al., 2018]



Key idea:

- 6 patches covering the whole sphere are extracted from the 360° image
- Two CNN (CNN 2D) + Refinement (CNN Saliency map + spherical coordinates per pixel)



Monroy, R., Lutz, S., Chalasani, T., & Smolic, A. (2018). Salnet360: Saliency maps for omni-directional images with cnn. *Signal Processing: Image Communication*, *69*, 26-34.

Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - SalNet360 [Monroy et al., 2018]

Top:
Stimulus / Ground truth

Bottom:
Base CNN
Base CNN + patches
Base CNN + patches + spherical coordinate

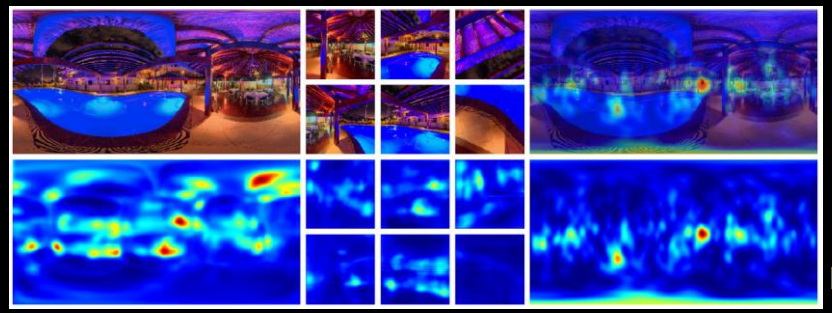
Monroy, R., Lutz, S., Chalasani, T., & Smolic, A. (2018). Salnet360: Saliency maps for omni-directional images with cnn. *Signal Processing: Image Communication*, *69*, 26-34.

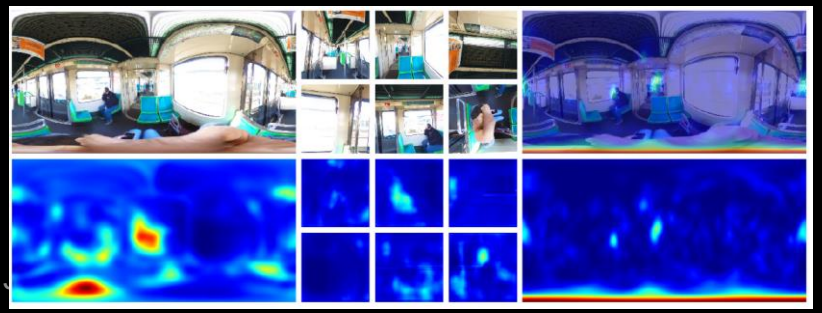
Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - SalNet360 [Monroy et al., 2018]

Comparison of the three experimental scenarios.

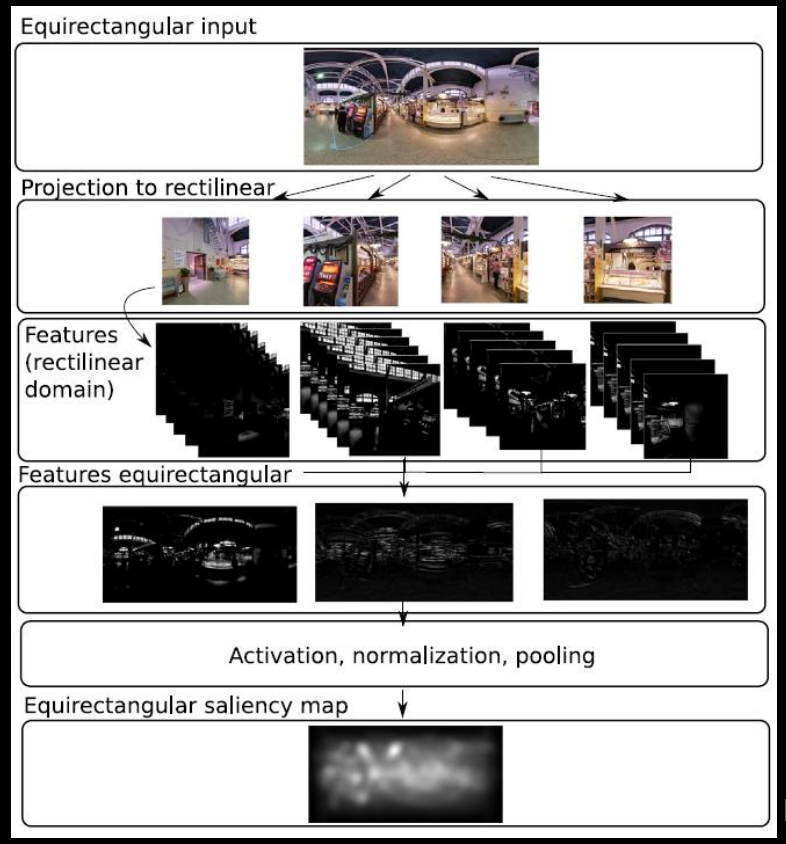
	KL	CC	NSS	AUC
Base CNN	1.597	0.416	0.630	0.648
Above + Patches	0.625	0.474	0.566	0.659
Above + Spherical coords	0.487*	0.536*	0.757*	0.702*





Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - GBVS360 [Lebreton & Raake, 2018]



Key ideas:

- Input equirectangular image is projected into several rectilinear images corresponding to different orientations of the user's viewport
- Gabor-based feature extraction
- Feature maps are then back-projected to the equirectangular domain

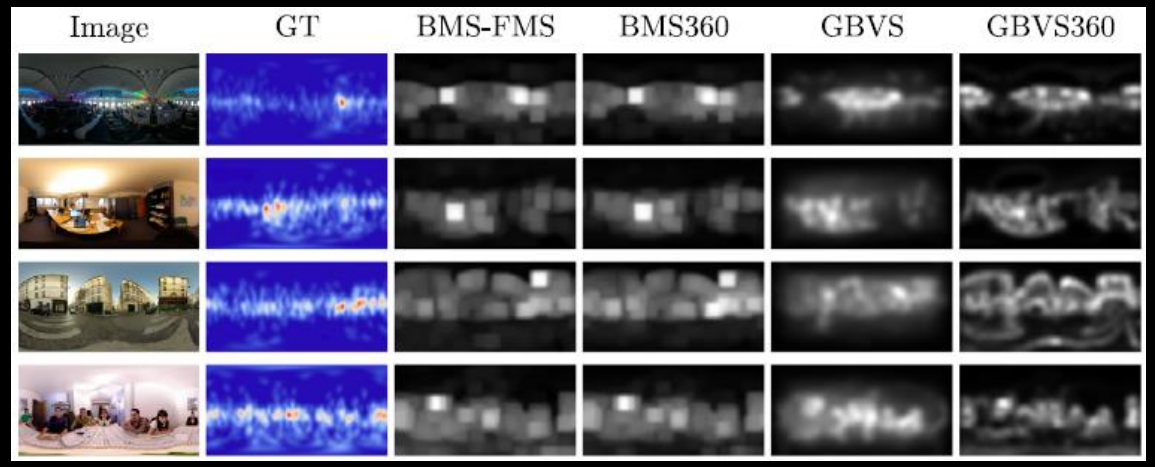
Lebreton, P., & Raake, A. (2018). GBVS360, BMS360, ProSal: Extending existing saliency prediction models from 2D to omnidirectional images. *Signal Processing: Image Communication*, *69*, 69-78.

Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - GBVS360 [Lebreton & Raake, 2018]

Model	Head/Eye KL	CC	NSS	AUC	Rank
TU Munich (3) [36]	0.44892	0.57862	0.80527	0.72598	1
TU Munich (6) [36]	0.42050	0.61478	0.80697	0.72209	1
SJTU [38]	0.48060	0.53244	0.91810	0.73478	2.25
Wuhan Univ. [35]	0.50825	0.53836	0.93588	0.73630	3.25
TU Munich (1) [36]	0.50084	0.55356	0.91531	0.74677	3.25
TU Munich (5) [36]	0.47503	0.55859	0.69606	0.71286	4
BMS360-Eq	0.59873	0.55365	0.93629	0.73556	
BMS360	0.52546	0.47202	0.85887	0.73473	
BMS-Eq	0.54289	0.54452	0.87607	0.72247	
GBVS360-Eq	**0.6980**	**0.52704**	**0.85053**	**0.71402**	**5.75**
BMS	0.53085	0.44963	0.76229	0.71104	

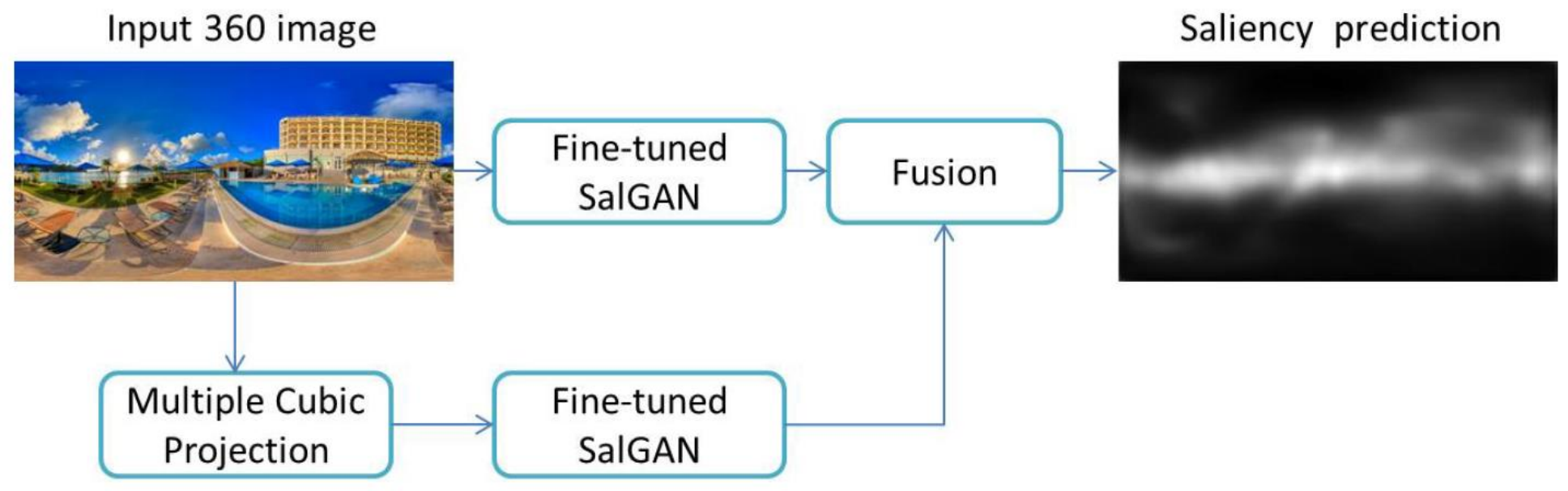
Eq = Equatorial prior based on a Gaussian mixture of 3 Gaussians (2 for the pole regions and one for the central region)



Lebreton, P., & Raake, A. (2018). GBVS360, BMS360, ProSal: Extending existing saliency prediction models from 2D to omnidirectional images. *Signal Processing: Image Communication*, *69*, 69-78.

Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - SalGAN 360 [Chao et al., 2018]



Key ideas:

- Global saliency map SalGan
- Local saliency maps by considering 9 rotations (cube maps) and saliency maps computed over cube faces (6x9=54)

Chao, F. Y., Zhang, L., Hamidouche, W., & Deforges, O. (2018, July). Salgan360: Visual Saliency Prediction On 360 Degree Images With Generative Adversarial Networks. In *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 01-04). IEEE.

Predicting saliency maps in VR, a brief review

- Computational models of visual attention
  - SalGAN 360 [Chao et al., 2018]

Method	KL↓	CC↑	NSS↑	AUC↑
SalGAN	1.236	0.452	0.810	0.708
SalGAN&FSM [15]	0.896	0.512	0.910	0.723
Maugey *et al.*	0.585	0.448	0.506	0.644
SalNet360[13]	0.458	0.548	0.755	0.701
GBVS360 [12]	0.698	0.527	0.851	0.714
SalGAN360	**0.431**	**0.659**	**0.971**	**0.746**

Chao, F. Y., Zhang, L., Hamidouche, W., & Deforges, O. (2018, July). Salgan360: Visual Saliency Prediction On 360 Degree Images With Generative Adversarial Networks. In *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 01-04). IEEE.

Conclusion

Topic	Learning Objectives
Part 3: Methods to generate saliency maps from eye tracking data [Le Meur]	1. Explain why 2D saliency map methods need to be generalized for omnidirectional viewing 2. Discuss the pros and cons of the selected methods 3. Compare the performance of different methods using standard metrics 4. Computational saliency models for 360° images

THANKS!!