Quality of Experience and user study: direct and indirect approaches

Patrick LE CALLET

Université de Nantes IPI (IMAGE PERCEPTION & INTERACTION) /LS2N

De l'Expérience à l'Algorithmie cognitive

Des algorithmes « bio inspirés »

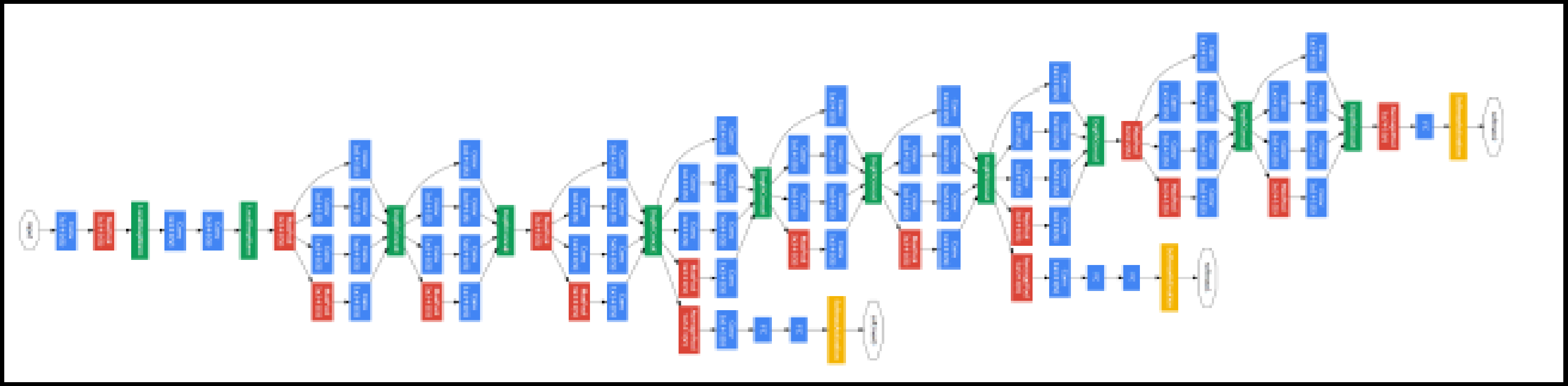
⇒Des algos d'IA

(dont Apprentissage profond)

Modélisation (et réutilisation) de fonctions cognitives/perceptuelles de haut niveau

Approche hybride:

ni Blanc ni Noir

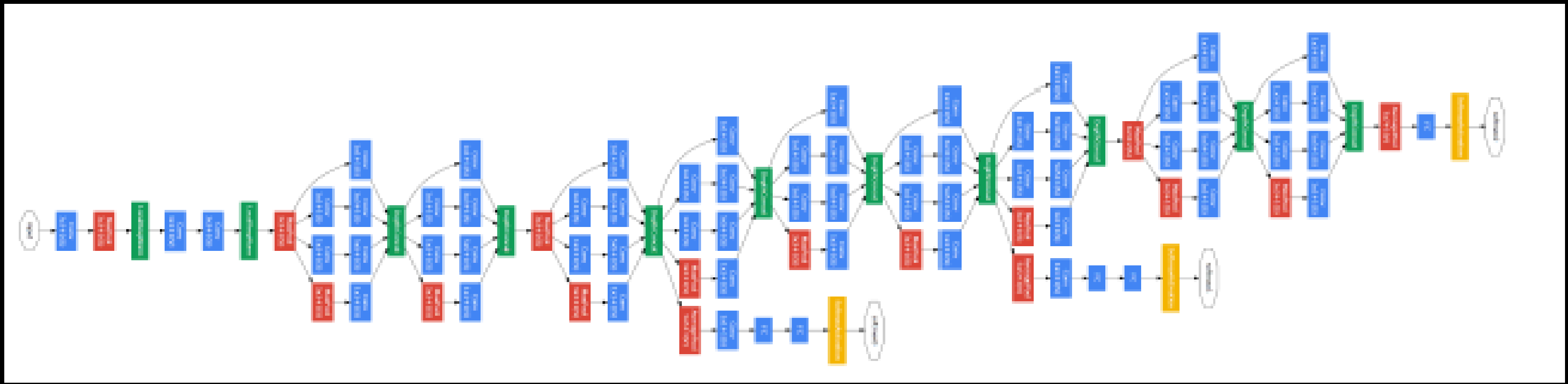


Prédire la mémorabilité des images

Modélisation (et réutilisation) de fonctions cognitives/perceptuelles de haut niveau

Fine tuning à partir de réseaux « sémantiques» + « émotion»

Cohendet and al. « **Deep Learning for Image Memorability Prediction: the Emotional Bias »**
ACM MM 2016 (best paper award)



Definition of Quality of Experience

Qualinet White Paper on Definitions of Quality of Experience

P. Le Callet, S. Möller and A. Perkis, eds.

Available at http://www.qualinet.eu

⇒A collective effort of more than 40 authors/contributors (COST Action IC 1003)
European Network on Quality of Experience in Multimedia Systems and Services

Quality of Experience (QoE) is the degree of **delight or annoyance of the user of an application or service.**

It results from the **fulfillment of his or her expectations** with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state.

Multimedia QoE: Apprécier la « qualité » comme un humain

Avec Netflix, l'université de Nantes repense les codes de la vidéo en ligne

En juin 2017, le système développé par l'université de Nantes a été présenté au Mobile World Congress de Barcelone. // © JORDI BOIXAREU/ZUMA/REA

Pour améliorer la compression de ses vidéos et permettre une diffusion plus fluide de son contenu sur smartphone, Netflix a fait appel à une équipe de recherche de l'université de Nantes. Ce partenariat prend désormais la forme d'un mécénat. Seuls trois établissements dans le monde sont ainsi soutenus par le géant américain.

QoE: Typology of influencing factors

System IF: *"properties and characteristics that determine the technically produced quality of an application or service. They are related to media capture, coding, transmission, storage, rendering, and reproduction/display, as well as to the communication of information itself from content production to user"*

QoE: Typology of influencing factors

System IF

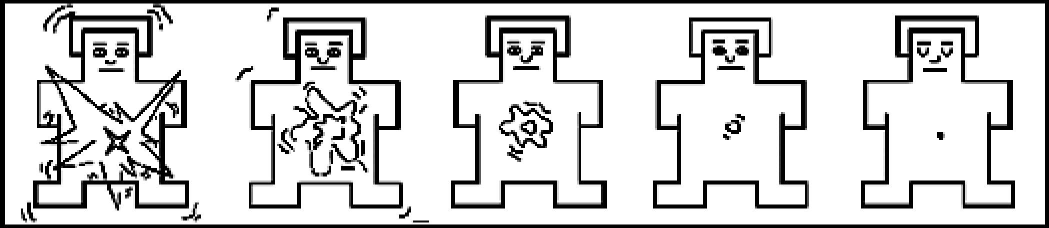
Context IF *"factors that embrace any situational property to describe the user's environment in terms of physical, temporal, social, economic, task, and technical characteristics"*

QoE: Typology of influencing factors

System IF

Context IF

Human IF *"any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socio-economic background, the physical and mental constitution, or the user's emotional state"*



Emotional & individual (human IF): how to?

Human IF *"any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socio-economic background, the physical and mental* constitution, or the *user's emotional state*"

From usual user studies …2 tracks:

In the Field:

demographic and socio-economic background

In Lab:

the physical and mental constitution, or the user's emotional state

Comprendre/mesurer/quantifier l’expérience

Approches Directes et Indirectes

méthodes/protocoles

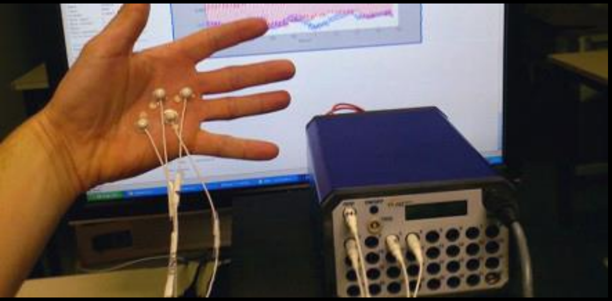
analyses statistiques

traitement signal/data mining

Standardisation de protocoles :

IEEE: booster des approches directes (même précision avec moins de panelistes)

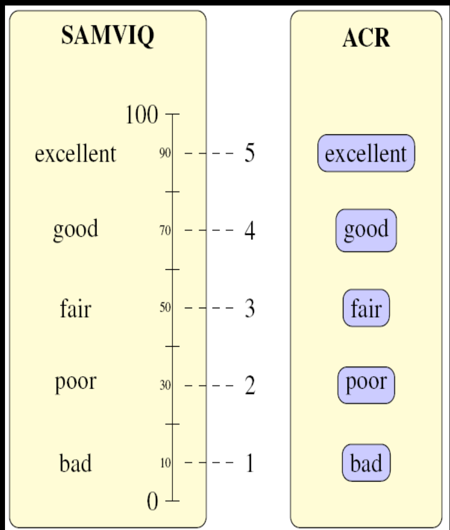
ITU: mesure et prédiction de l’inconfort et fatigue en environnement Immersif



QoE: approches directes (questionnaires)

subjectivité et incertitudes

Improving the discriminability of standard subjective quality assessment methods



Ground Truth

Video	MOS	Std
1	4.12	0.51
2	3.87	0.58
3	2.21	0.61
4	3.56	0.59
5	1.93	0.49
...		

Uncertainty

Traditional Solution: Increase sample size

MLE Quality Recovery Model

$$X_{e,s} = X_e + B_s$$

$$B_s \sim \mathrm{N}\left(b_s, v_s^2\right)$$

- **Observed score of Video *e* from subject *s***: $X_{e,s}$
- **True score of Video *e***: X_e
- **Influence from subject *s***: B_s
- **Subject's inconsistency**: v_s^2
- **Subject's bias**: b_s

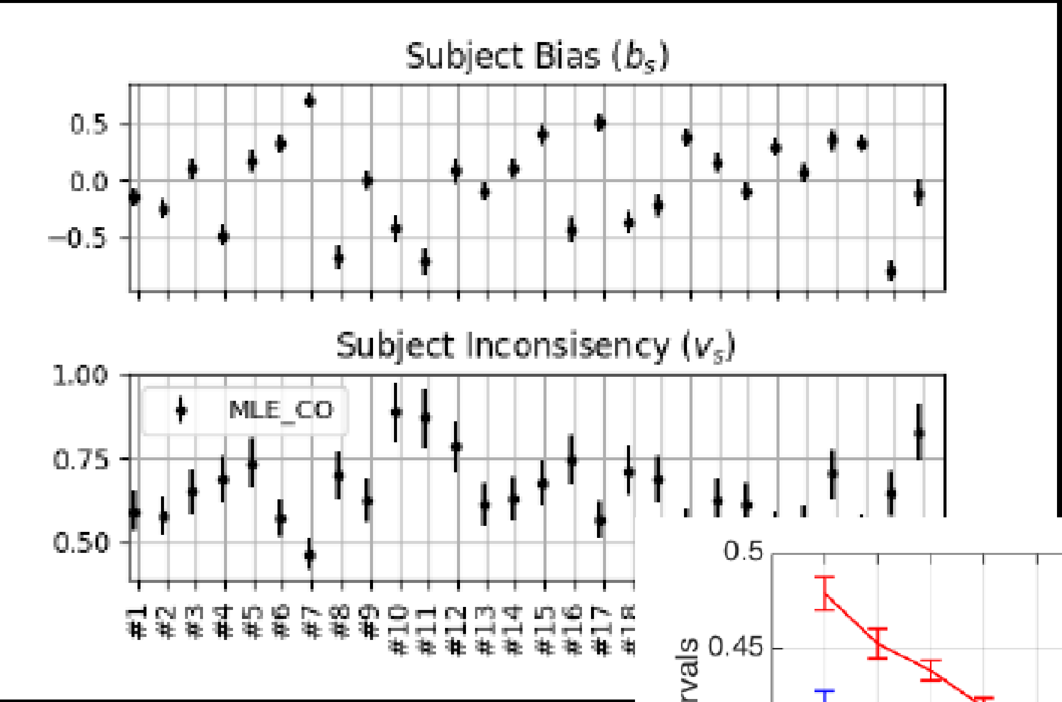
By using Maximize Likelihood Estimation (MLE)

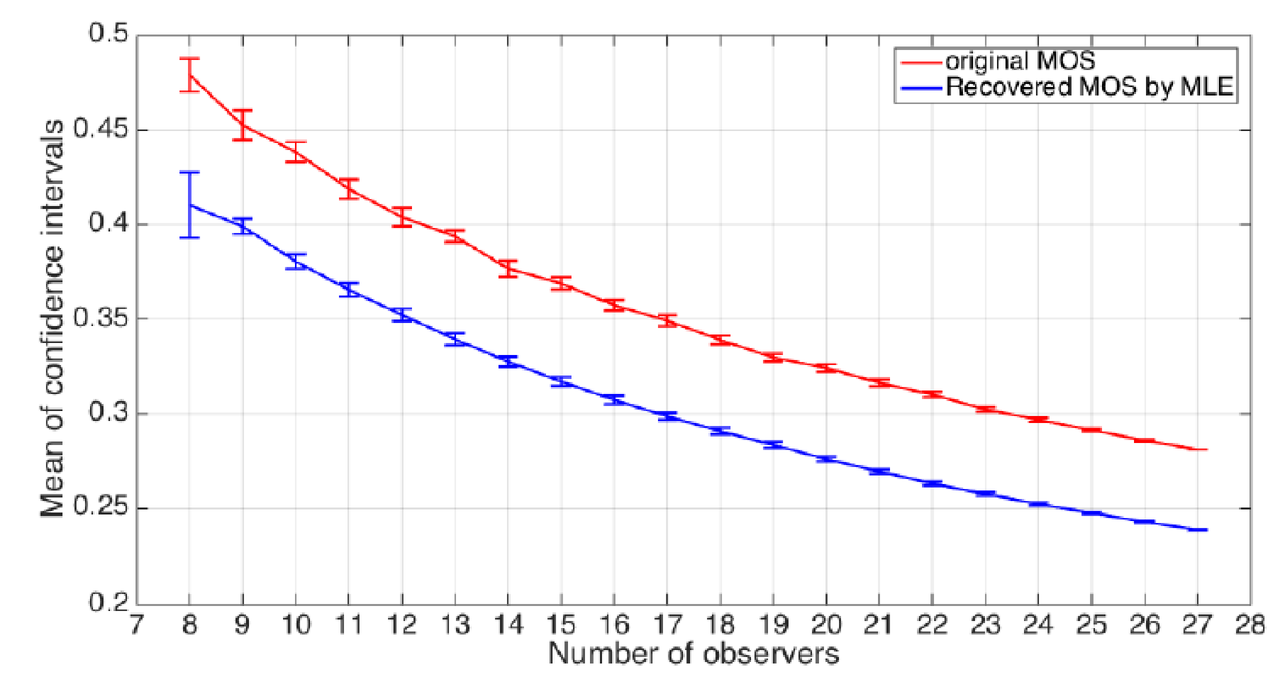
Likelihood function: $L = \log P(X_{e,s} \mid \theta), \theta = (X_e, b_s, v_s)$

MLEs: $\hat{\theta} = \mathrm{argmax}_\theta L$

Confidence Interval (CI): $\hat{\theta} \pm 1.96 \dfrac{1}{\sqrt{-\partial^2 L / \partial \theta^2}}$

Mieux qualifier les panélistes et diminuer le nombre de panélistes





QoE: approches directes

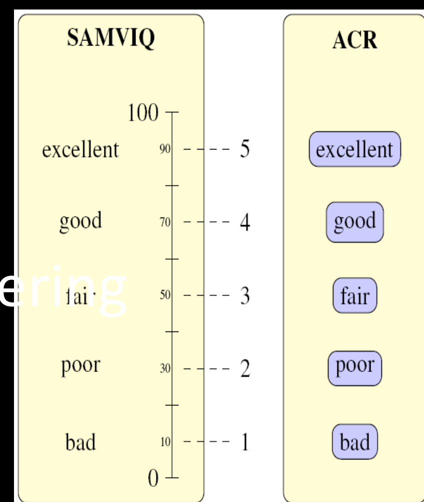
Muldimensionnalité,
Pair comparison &
crowdsourcing

Rating Scale?

Subjects are not always capable of expressing their perceptions, impressions by means of an exact numerical value

naturalness

Visual comfort



Visual Quality

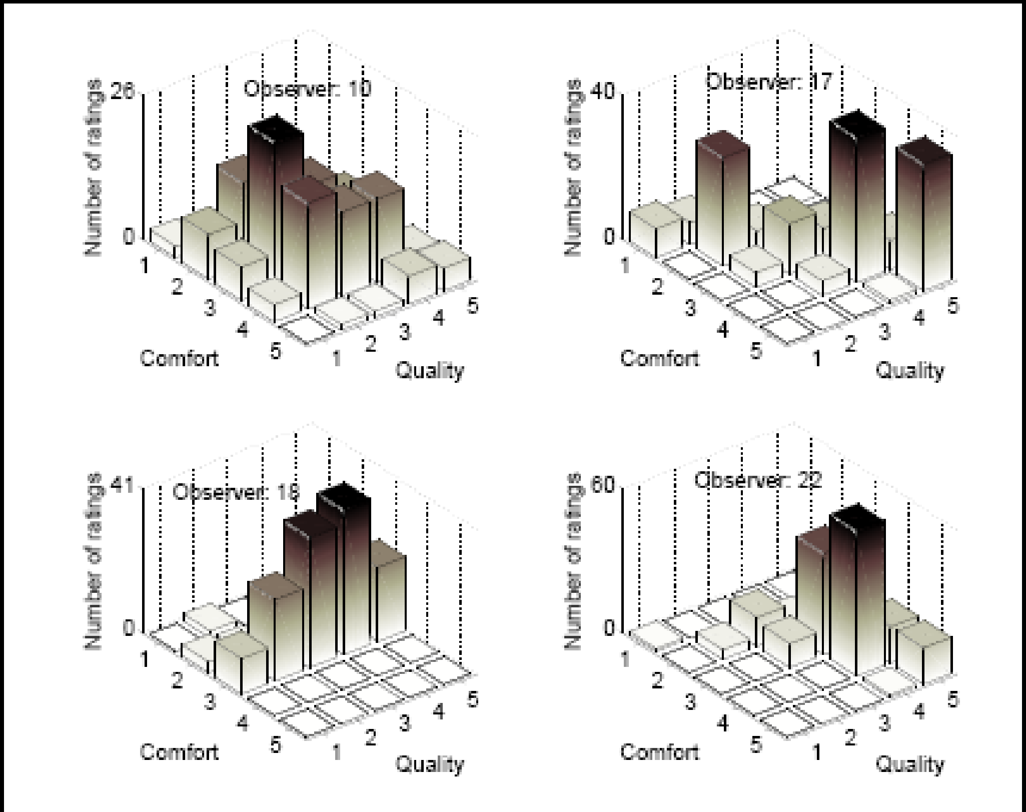
Depth quantity

Depth rendering

Visual experience

Scale interpretation & observer variability

Co joint quality and comfort ratings for 4 observers

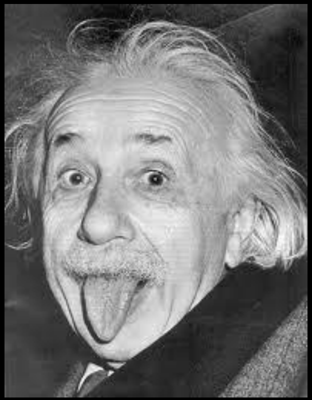


*TOWARDS A FRAMEWORK OF INTER-OBSERVER ANALYSIS IN MULTIMEDIA QUALITY ASSESSMENT- QOMEX 2011 - Ulrich Engelke, Yohann Pitrey, Patrick Le Callet

Pair comparison test

Comparing two stimuli based on an ad hoc question:

Ex: « choose the person with whom you would rather spend an hour discussing a topic of you choosing » (adapted from Rumelhart and Greeno 1971)



What do we get ? => a preference matrix

	A1	A2	A3	A4
A1	-	46	29	48
A2	44	-	34	43
A3	61	56	-	50
A4	42	47	40	-

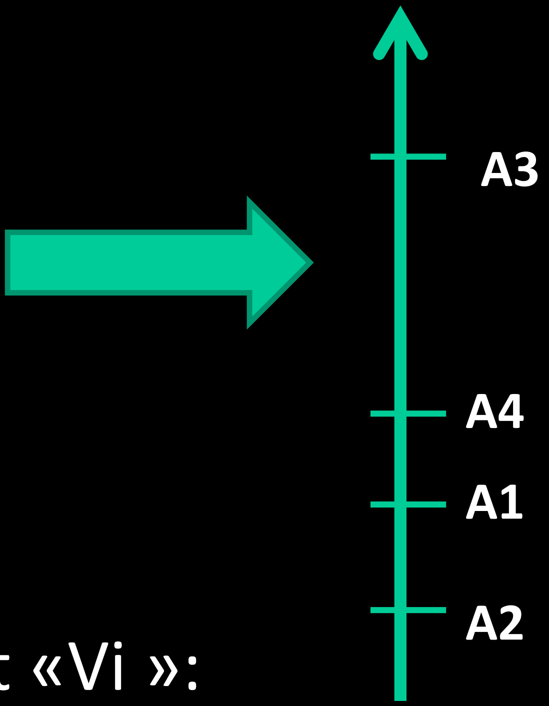
Analysing PC data

- **Conversion to scale values possible using Bradley-Terry or Thurstone-Mosteller models**

The goal: Mapping **probabilities of preference** to a **scale**

=> Linear models of paired comparisons

	A1	A2	A3	A4
A1	-	46	29	48
A2	44	-	34	43
A3	61	56	-	50
A4	42	47	40	-



Each stimulus Ai has a merit «Vi »:

in psychophysics, a sensation magnitude on a scale

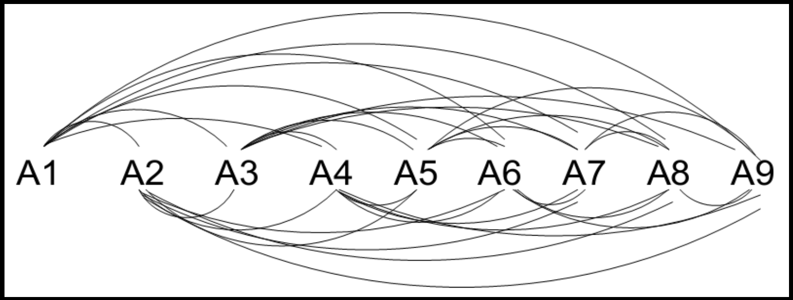
PAIR COMPARISON METHODOLOGY

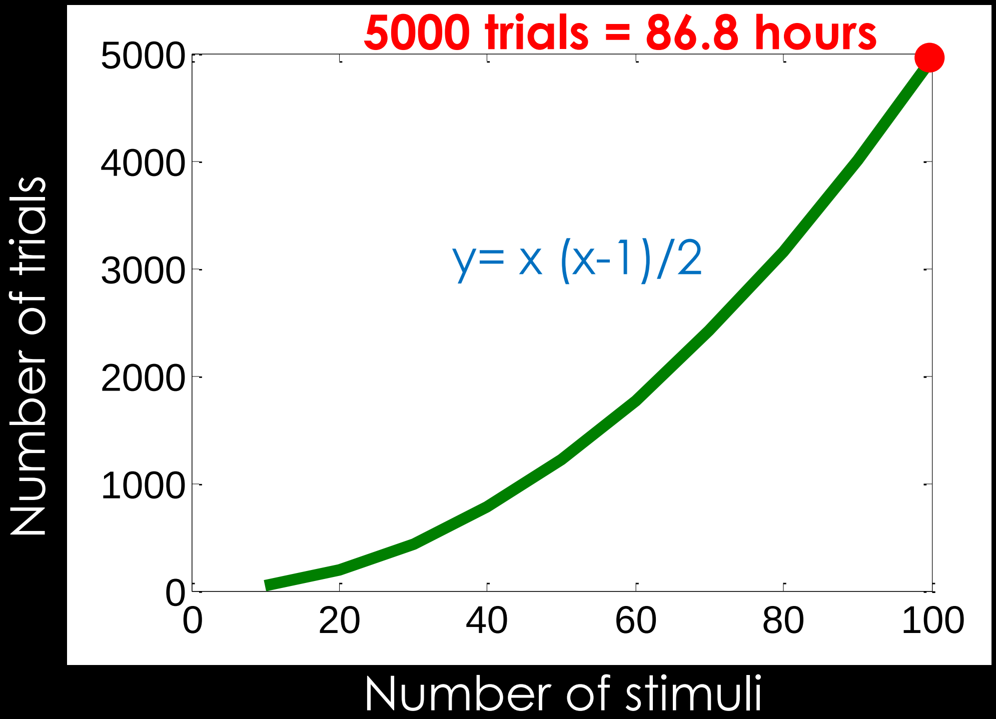
- Full Paired comparison (FPC):

Every pair is compared:

(A, B, C)→ AB, AC, BC

m stimuli → *m(m-1)/2* pairs





Advantages:
Easy, reliable, discriminability on closer pairs

Disadvantage:
Infeasible when the number of stimuli Is large

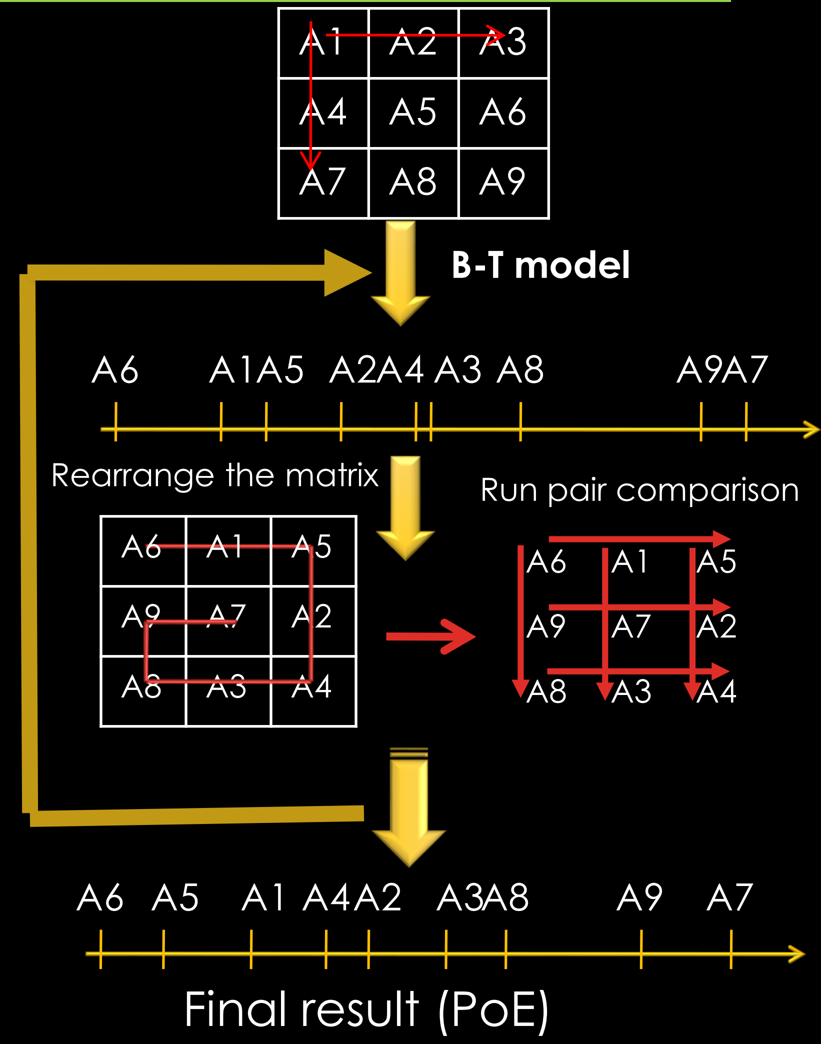
ADAPTIVE SQUARE DESIGN (ASD)

For the scenario that the ranking order of the test stimuli is not available

1. **Initialize** the square matrix randomly
2. **Run paired comparisons** according to the rules of square design.
3. **Calculate** the estimated scores.

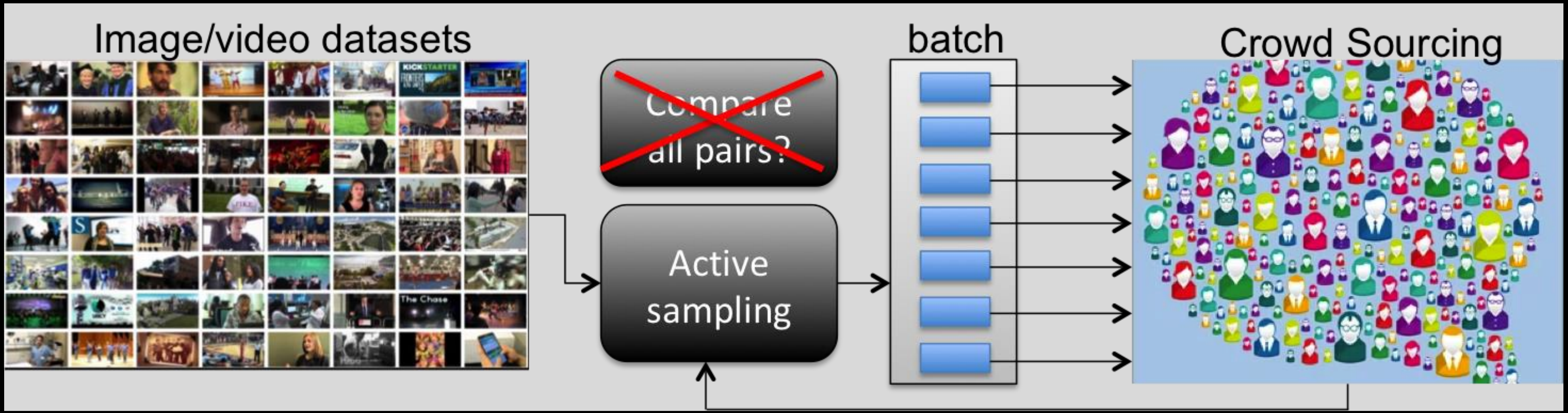
   According to current paired comparison results calculate the scores and sort them.
4. **Update** the square matrix.

   The adjacent pairs could be arranged according to this spiral
5. **Repeat** step 2 and 4, until certain conditions are satisfied (e.g., 40 observers)



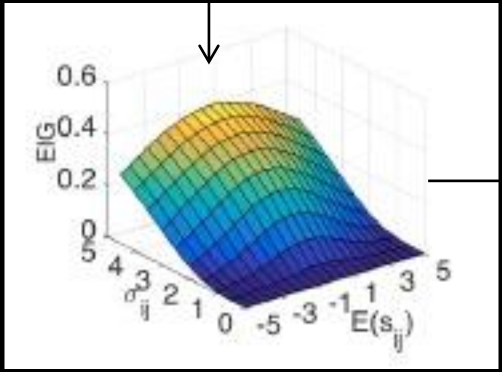
Active sampling for pairwise comparison

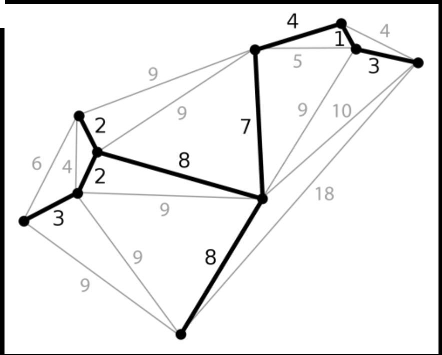
(NIPS 2018)



Batch selection:
Active learning according to
Bayesian theory, **KL divergence**,
Expected Information Gain (EIG),
Minimum Spanning Tree (MST)

$$U_{ij} = \int \sum_{y_{ij}} log \left\{ \frac{p(s_{ij}|y_{ij})}{p(s_{ij})} \right\} p(s_{ij}|y_{ij}) p(y_{ij}) ds_{ij}$$





A minimum spanning tree (selection of batch)

J. Li, et.al., Hybrid-MST: A Hybrid Active Sampling Strategy for Pairwise Preference Aggregation, NIPS 2018

QoE: approches indirectes

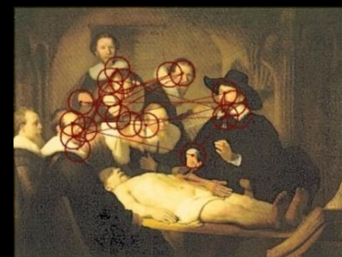
Oculométrie

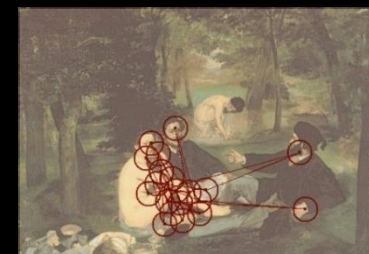
Occulométrie : mesure de localisation du regard

Types de vérité terrain (et modélisations)

Prédictions possibles :

– Chemin visuel





– Région d’intérêt perçu

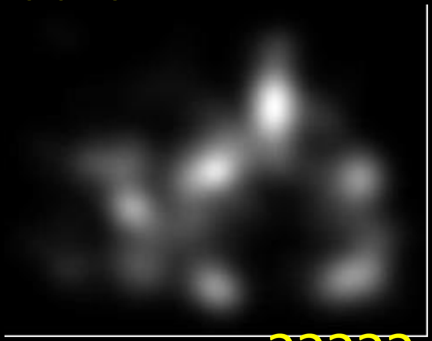
– Carte de saillance

Des données oculométriques aux modèles computationnels

La vérité « terrain »

?????

pas de standard !



?????

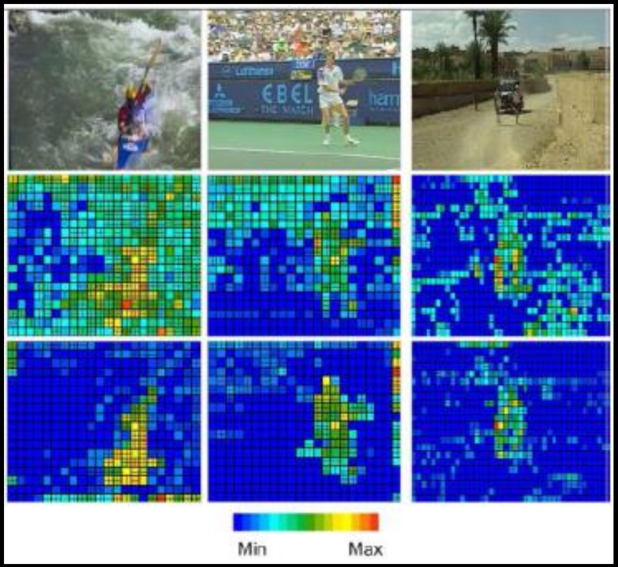
pas de standard !

L’objectif (sortie du modèle):

la carte de saillance

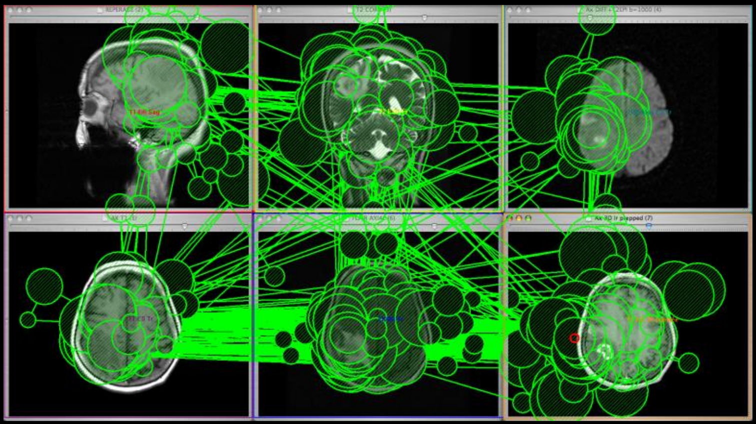
Quelques Applications MM

Codage d'images



Reformatage de contenus

Ergonomie visuelle



Réalité virtuelle

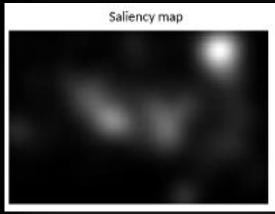
Focus intelligent et adaptatif

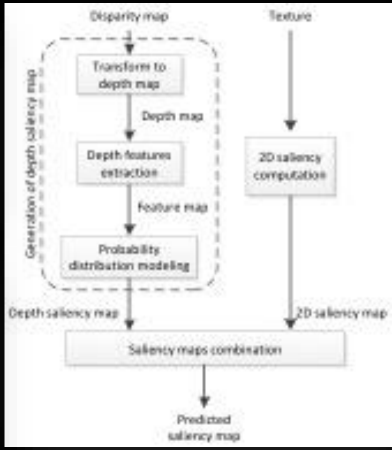
Gestion du conflit vergence/ accomodation

Le Callet and E. Niebur Proc. Of IEEE 2013 «Visual Attention and Applications in Multimedia Technologies»

Quelques Applications MM (2)

Sous titrage « ergonomique »

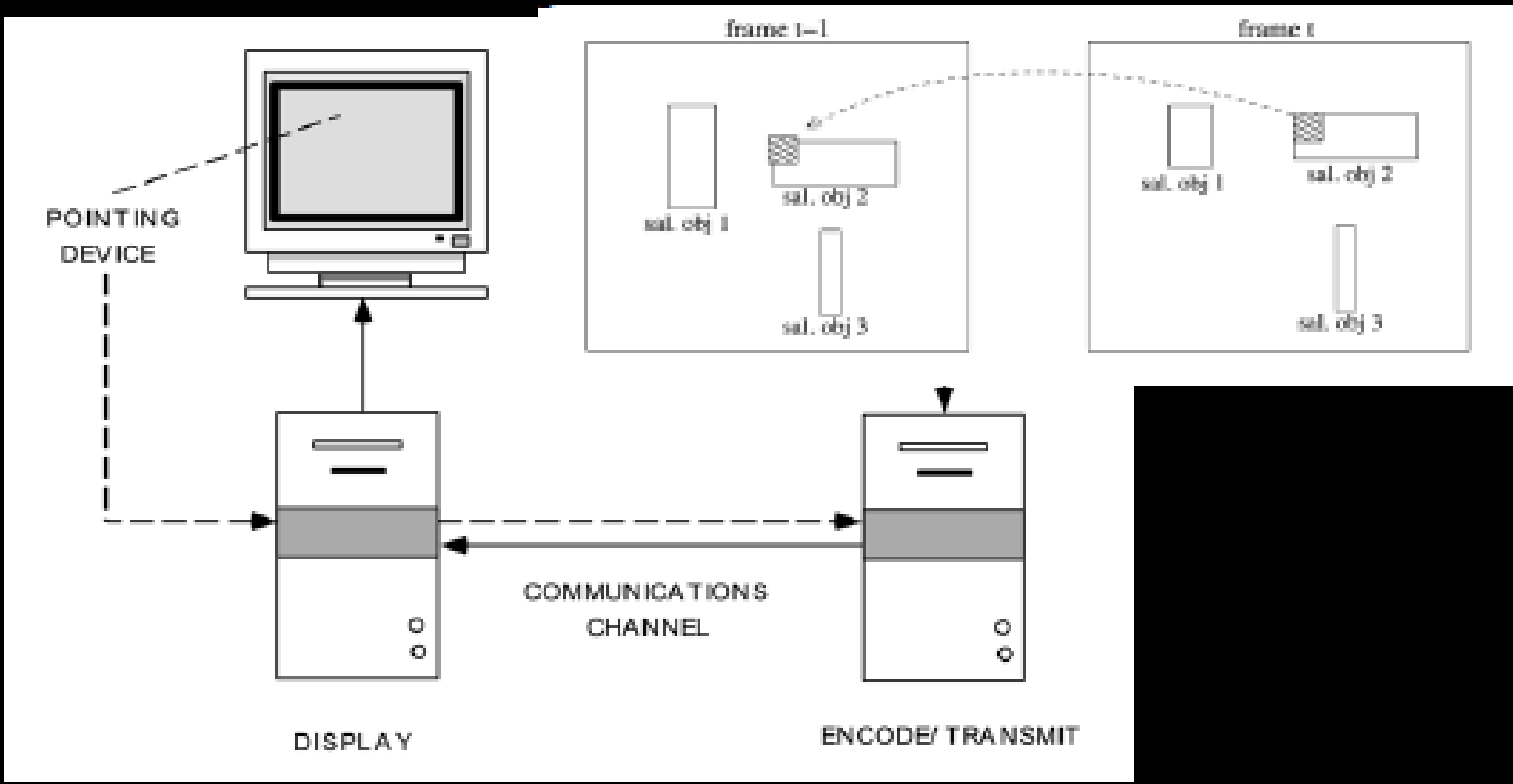




Visualisation 3D stéréo : Modèle computationnel *JEMR12, IEEE TIP13, IEEE TIP14*

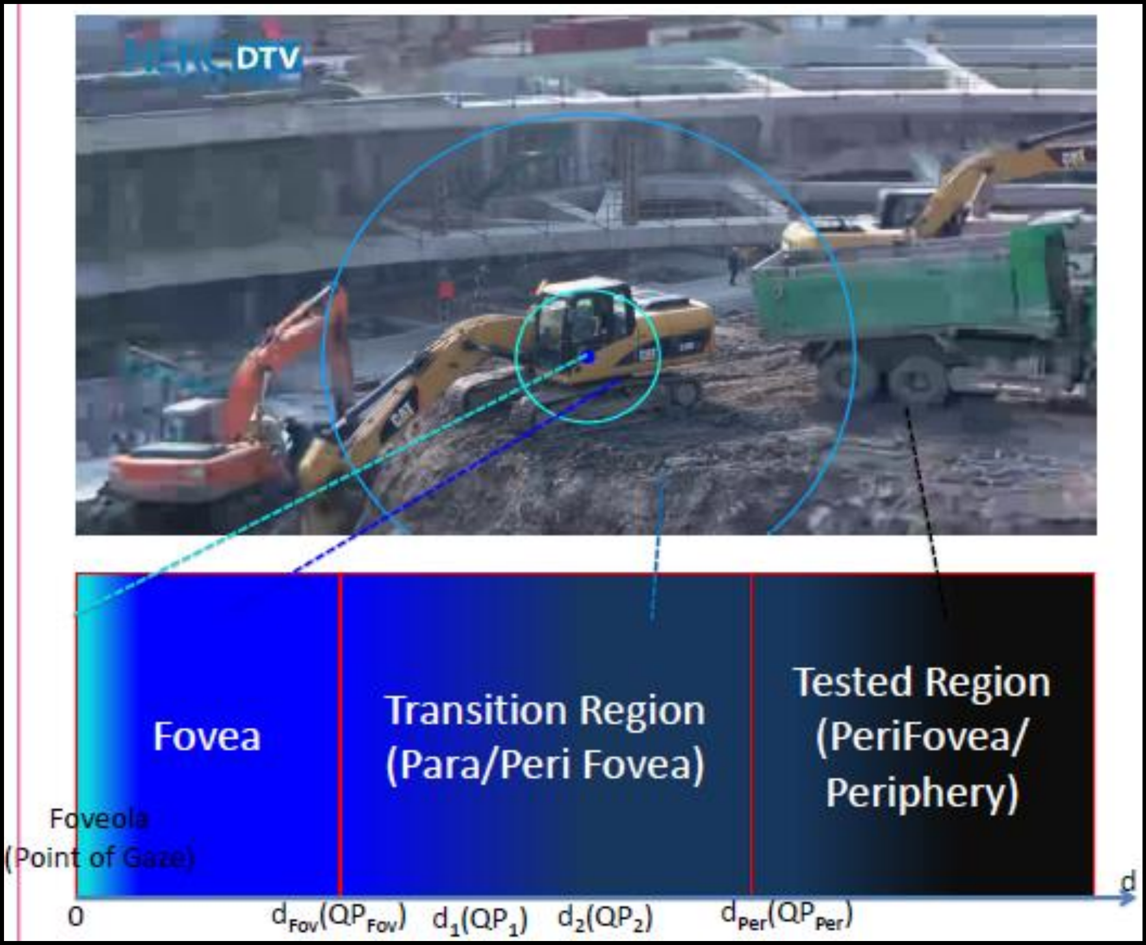
=> 3D retargeting, 3D confortable

streaming interactif ?



Y. Feng, G. Cheung, W. Tan, P. Le Callet, et Y. Ji, « Low-Cost Eye Gaze Prediction System for Interactive Networked Video Streaming », IEEE Transactions on Multimedia, vol. 15, no 8, p. 1865-1879, 2013.

Comment des distorsions périfoéales modifient elles le chemin visuel (scanpath)?

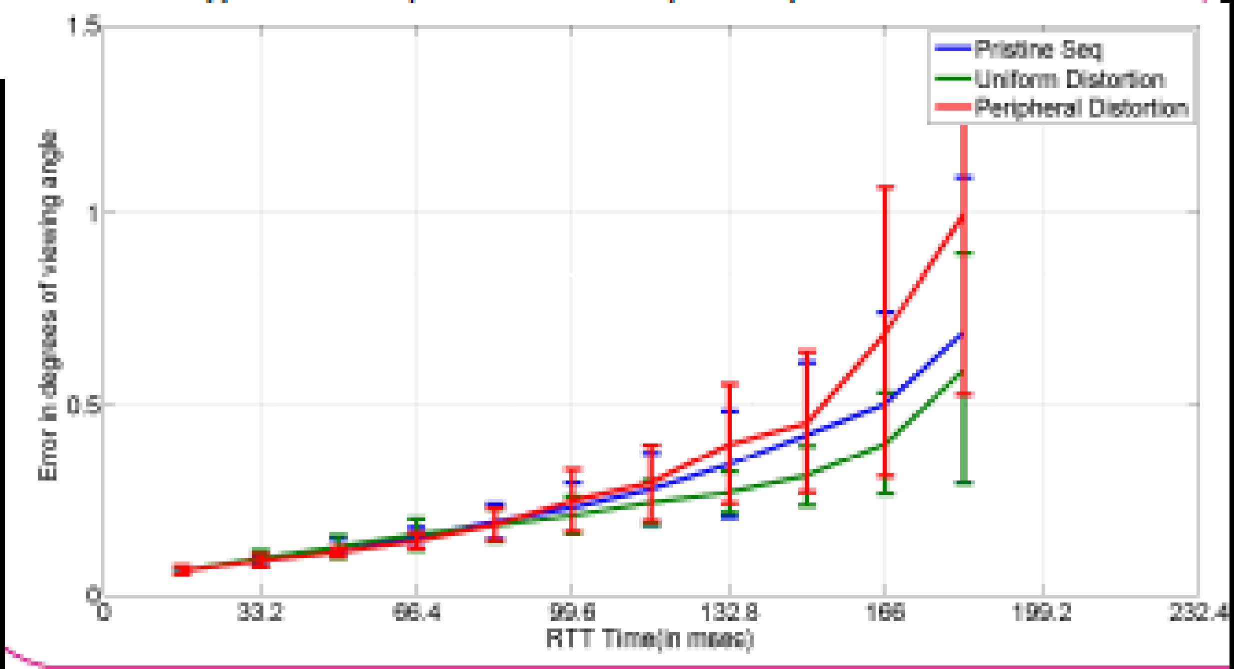


Y. Rai , M. Barkowsky & P. Le Callet « Role of peripheral Spatio-Temporal distortions indisrupting natural attention deploymentzation » HVEI'16 (best student paper)

Un espoir pour le streaming interactif

Can the models predict gaze accurately?

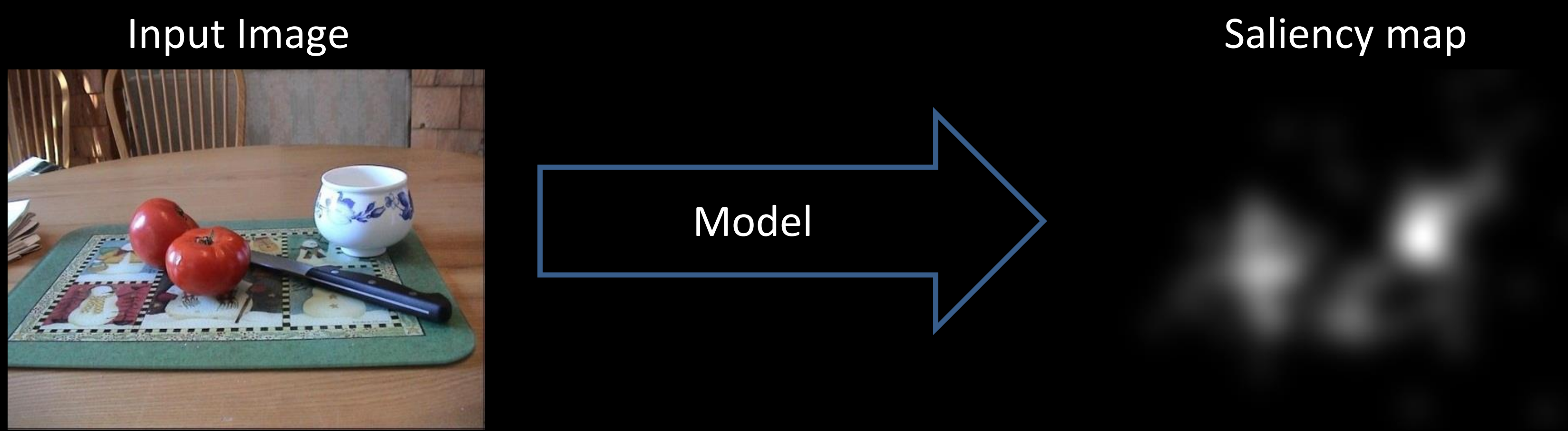
- Prediction error increases with time in case of peripheral distortions: more so if the sequence contains a lot of temporal activity.
- Prediction error can be restricted to a maximum of about 1.5 degrees of viewing angle within 200ms.
- Gaze prediction is not statistic significantly affected by Peripheral or Uniform distortions!



Y. Rai , G. Cheung & P. Le Callet « Role of HEVC Coding artifacts on gaze prediction in interactive video streaming systems» ICIP 16

Modélisation à partir de données oculométriques ?

25 ans de recherche en modélisation algorithmique d'attention visuelle

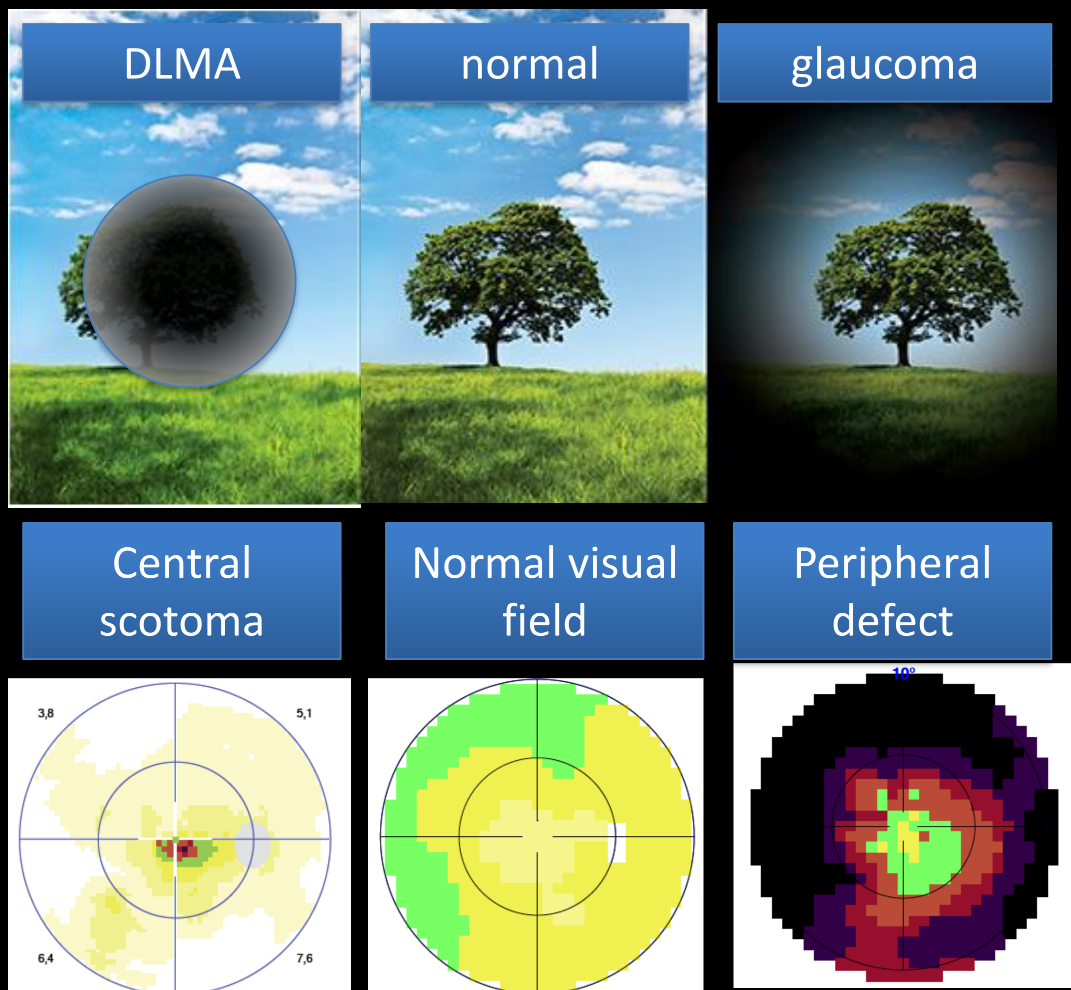


Les modèles actuels sont des modèles fovéaux

L'hypothèse qu'une image est inspectée uniquement avec la fovéa est très discutable
(hypothèse pratique liée à la vérité terrain)

Approche Projet VAM2020 (Atlanstic2020)

comprendre les relations entre vision périfovéale et fovéale au travers d'études de personnes souffrant de pathologies du champ visuel

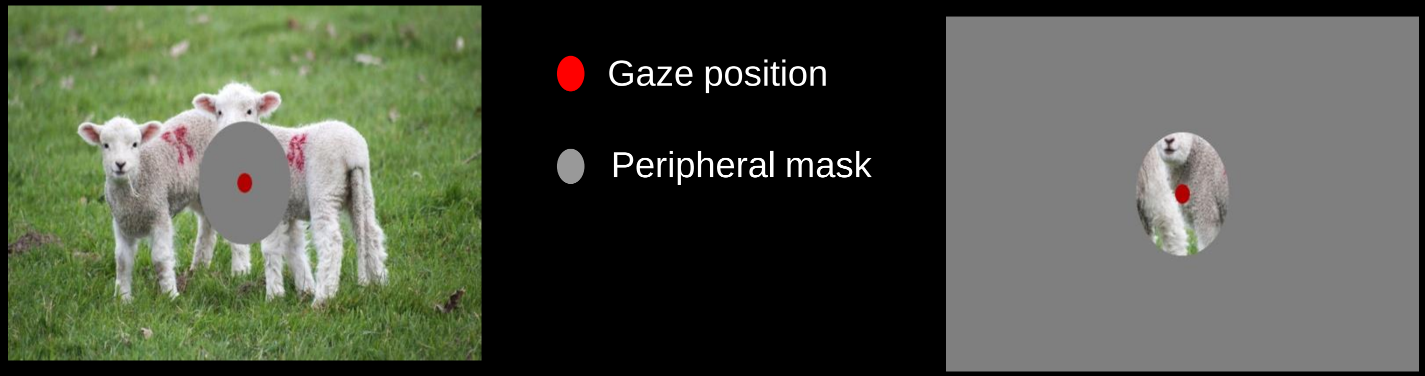


Artificial Visual Field Defects

Mask radius sizes

Mask types

Real-time modification of on-screen stimuli according to gaze position.



Artificial Visual Field Defects - Protocol

Camera sampling 2ms. – 500Hz

Screen update
7ms. – 144Hz

Scene update
1ms.

Subject

EyeTracker

Screen

Display Computer

Gaze processing
2ms.

Network
1ms.

EyeTracking Computer

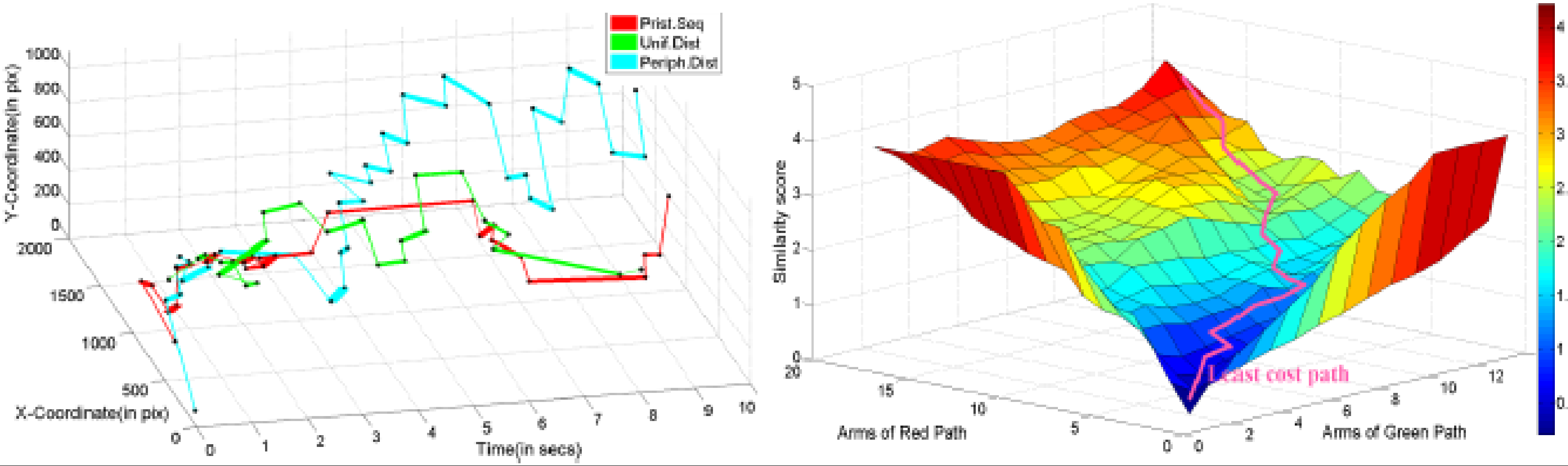
Gaze Data

Eye movement to screen max update latency = 13ms

Analyser, Comparer des « scanpaths »

Comment comparer 2 scan paths?
(…sans considérer le contenu)

- Vector Similarity: Combined analysis of several gaze parameters like Saccade Amplitudes, Fixation duration, Frequency of saccades, Areas of interest



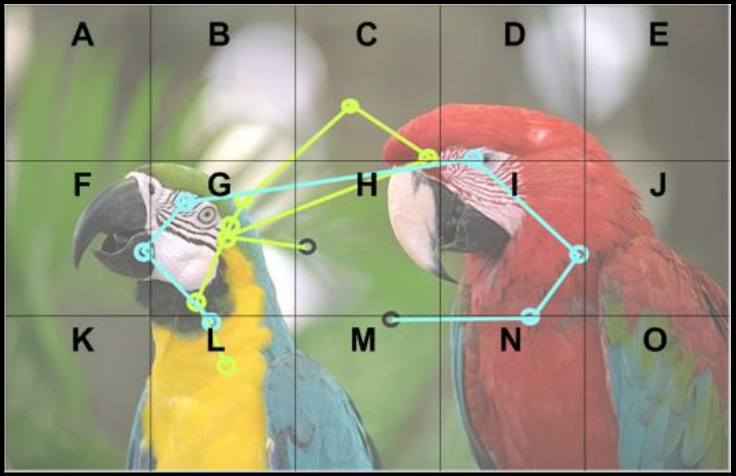
Y. Rai , M. Barkowsky & P. Le Callet « Role of peripheral Spatio-Temporal distortions indisrupting natural attention deploymentzation » HVEI'16 (best student paper)

Comment comparer 2 scan paths?

String edit (initiallement pour mesurer la distance entre deux mots): levenshtein similarity metric

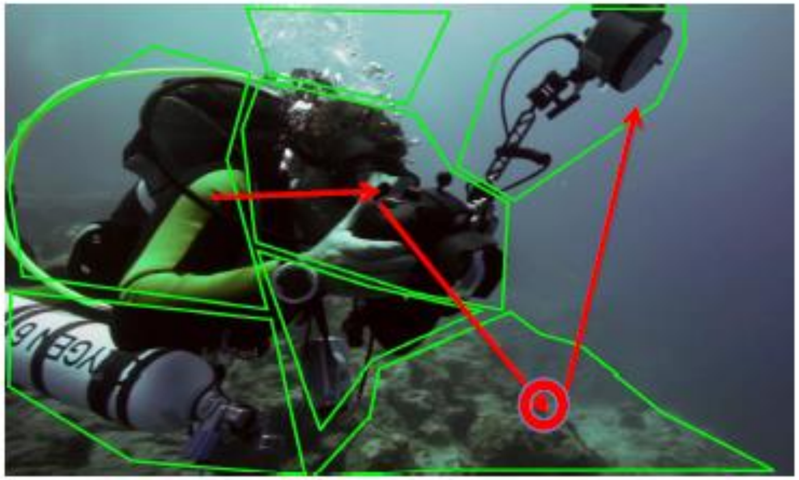
Nombre d'opération minimum pour transformer un chaine de carectère en une autre

Advantages:
+ Easy to compute
+ hold the order of fixation
Drawbacks:
- How many viewing areas of interest should be used (7,12,15,25...)?
- does not take into account fixation duration



Comment comparer 2 scan paths?

- Asking the users, *Which objects did you notice in the presented scene?* : Segmenting the scene manually into regions with fixed semantic meaning.



- Comparing object transitions : **D-B-B-C-C-C-A**, **D-D-C-B-A** followed by Levenshtein similarity of string patterns.

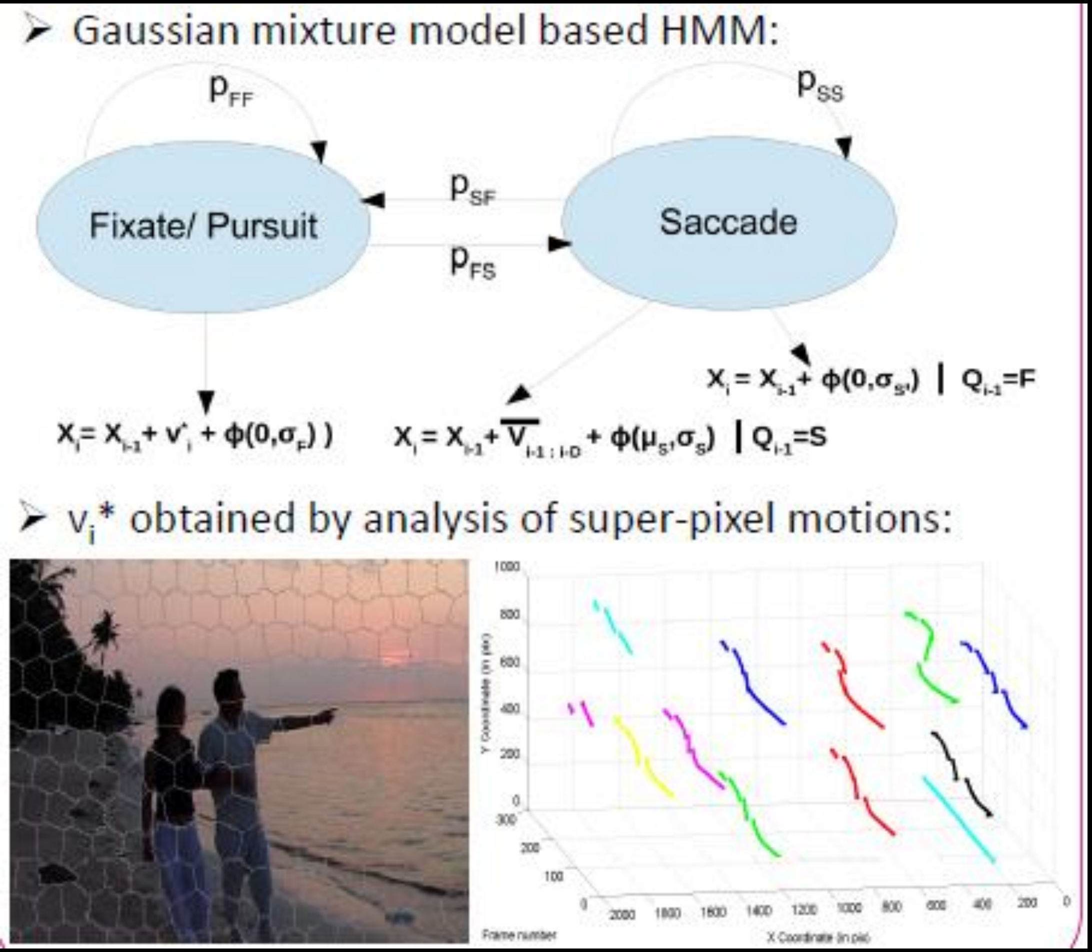
- Comparing the relative shift in attention from ROIs to non-ROIs and vice versa using contingency tables : Mc-Nemar Chi-Square test.

Ref \ Test	ROI	BG
ROI	26	10
BG	11	1

Ref \ Test	ROI	BG
ROI	24	12
BG	6	5

Y. Rai , M. Barkowsky & P. Le Callet « Role of peripheral Spatio-Temporal distortions indisrupting natural attention deploymentzation » HVEI'16 (best student paper)

Comment comparer 2 scan paths?
En considérant le contenu ...vidéo



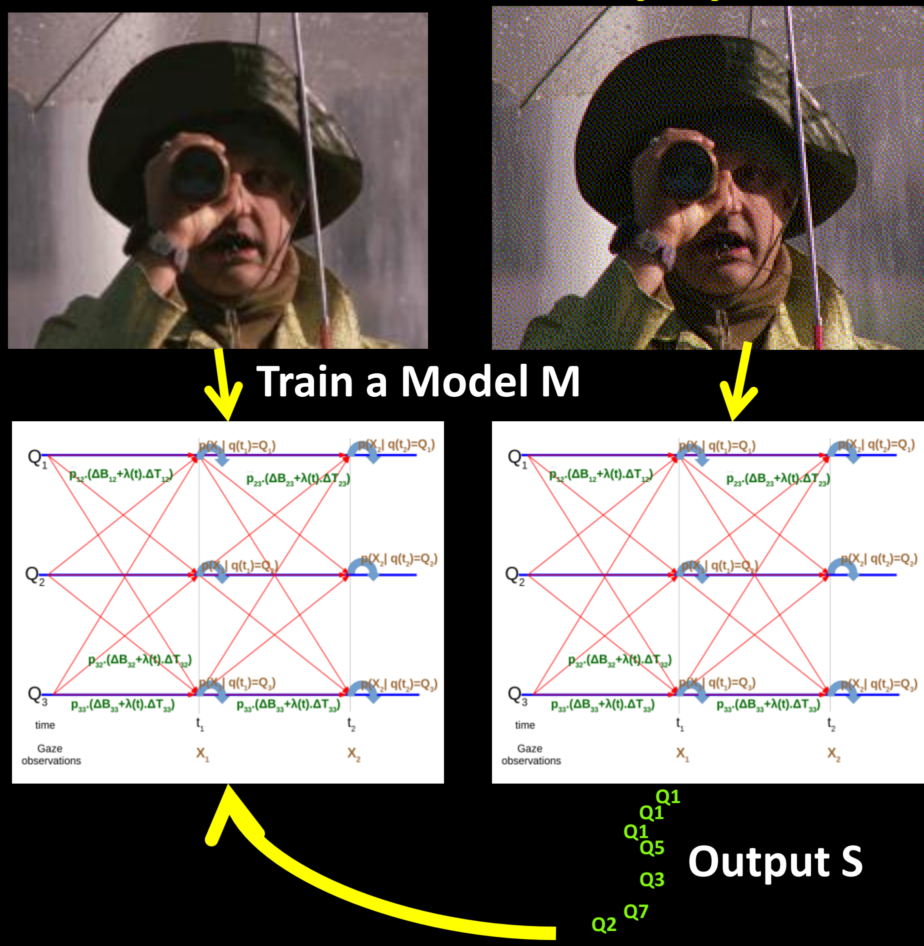
Approche
Markovienne
+ computer vision

Y. Rai , G. Cheung & P. Le Callet « Role of HEVC Coding artifacts on gaze prediction in interactive video streaming systems» ICIP 16

Comment comparer 2 scan paths?
En considérant le contenu …vidéo (2)

- Entrainement à partir de données correspondant à différentes conditions

- Tester la vraisemblance que des sorties d'un modèle correspondent à un autre



Difference=lik ($S_{impaired}$ | $M_{pristine}$) - lik($S_{pristine}$ | $M_{pristine}$)

Top down vs Bottom up

Un autre regard sur la vérité terrain oculométrique

Wang and al. HVEI 11 « Quantifying the relationship between visual salience and visual importance »
Engelke and al. , SPIC 14 « Perceived interest and overt visual attention in natural images"

Context

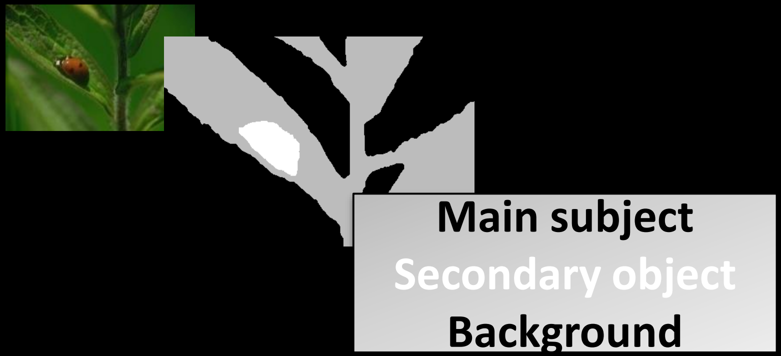
- Two types of ground truths for visual attention
  - Fixation density map (visual saliency, bottom-up)
  - Region of interest (visual importance, top-down)

- Quantitative relationship between visual saliency and visual importance
  - Two psychophysical experiments jointly conducted

Wang and al. HVEI 11 « Quantifying the relationship between visual salience and visual importance »

Experiment I: a scoring experiment

- Task
  - Give a score of "importance" to each object



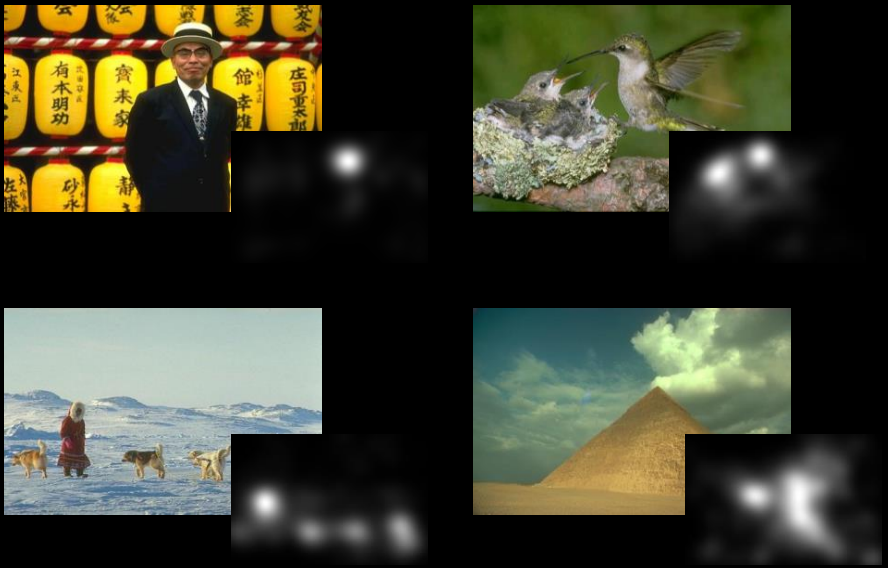
- Post-processing of data
  - Raw data ->
    Classification of objects
    - Main subject
    - Secondary object
    - Background

Wang and al. HVEI 11 « Quantifying the relationship between visual salience and visual importance »

Experiment II: eyetracking

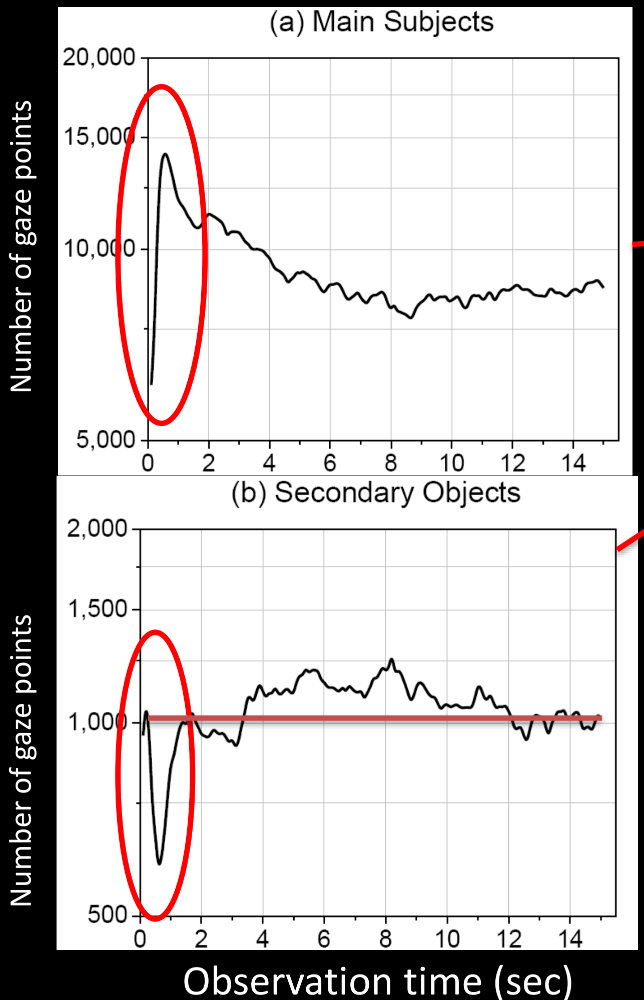
- Task
  - Free-viewing

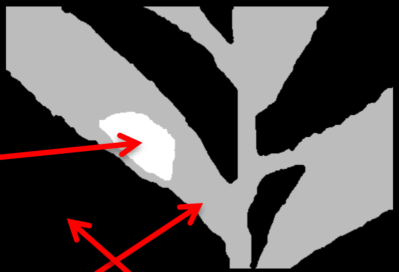
- Post-processing of data
  - Eye-tracking data -> Visual saliency map (i.e. FDM)

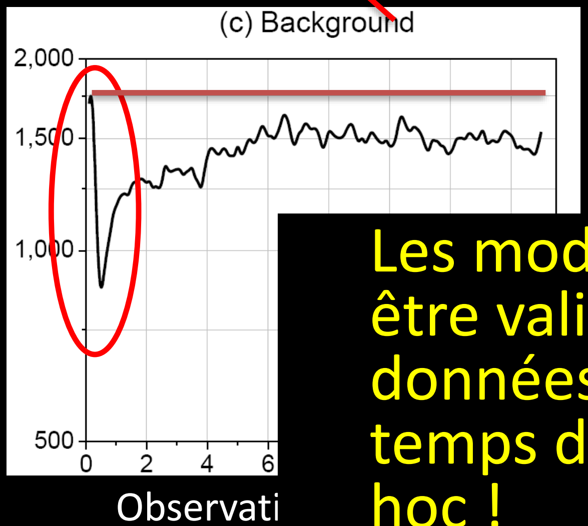


Wang and al. HVEI 11 « Quantifying the relationship between visual salience and visual importance »

Time depency analysis





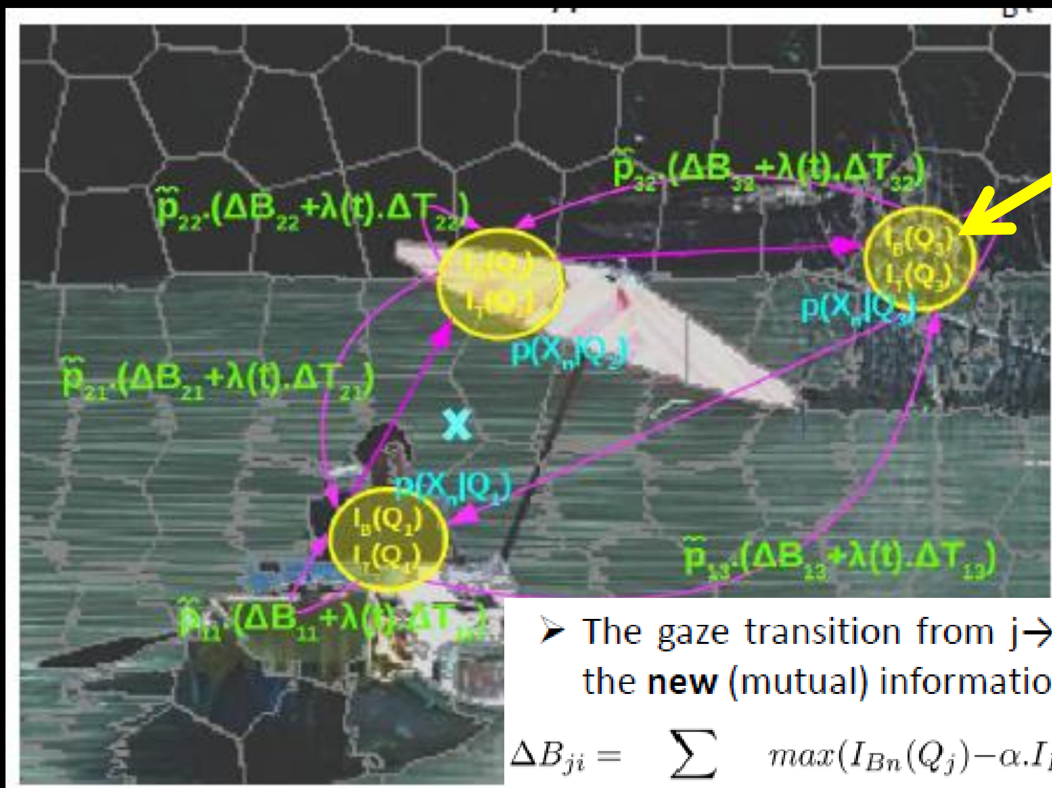


Les modèles devraient être validés avec les données correspondant au temps de visualisation ad hoc !

Wang and al. HVEI 11 « Quantifying the relationship between visual salience and visual importance »

Modélisation des infuences mutuelles Bottom-up top down

Image = HMM avec comme états cachés des superpixels



Super pixel contient :
information top down
Information Bottom-up

- The gaze transition from j→i (Transition probability M_{ji}) is a function of the **new** (mutual) information: ΔB_{ji} , ΔT_{ji} & the occulomotor bias $\tilde{p}_{ji}$

$$\Delta B_{ji} = \sum_{n=\in(Col,Lum,Tex,Mot)} max(I_{Bn}(Q_j) - \alpha.I_{Bn}(Q_i), 0) \qquad \Delta T_{ji} = \sum_{n\in Objects} max(I_{Tn}(Q_j) - \alpha.I_{Tn}(Q_i), 0)$$

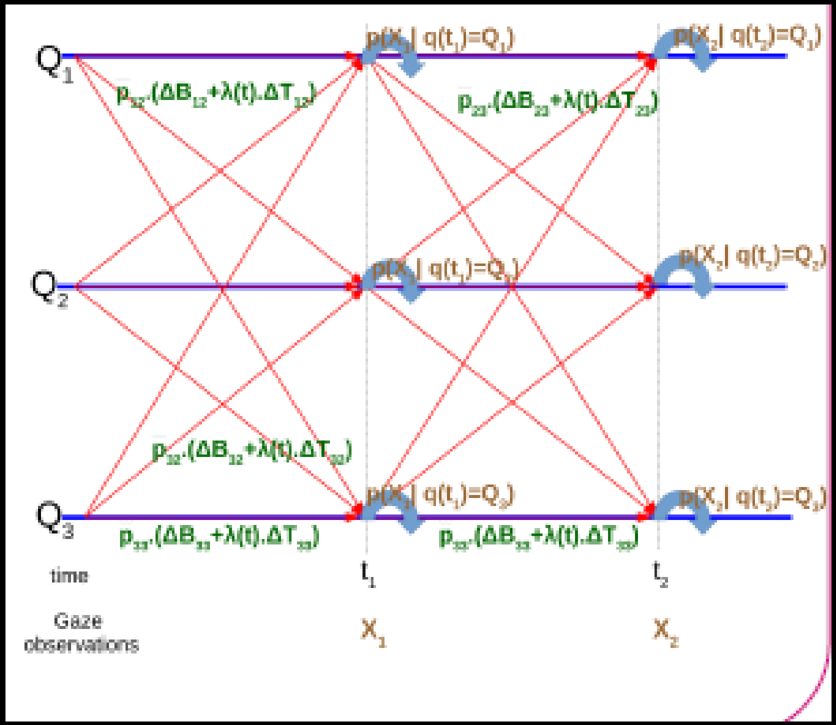
$$M_{ji} = \tilde{p}_{ji}(\Delta B_{ji} + \lambda(t)\Delta T_{ji})$$

Y. Rai , P. Le Callet and G. Cheung« Quantifying the relation between perceived interest and visual salience during free viewing using Trellis based Optimization » IEEE IVMSP16

Modélisation des infuences mutuelles Bottom-up top down

- Given the gaze data X $_{t(m):t(n)}$, we iteratively compute the likelihood over the trellis to converge towards the optimum $\lambda_{t(m):t(n)}$ in this period

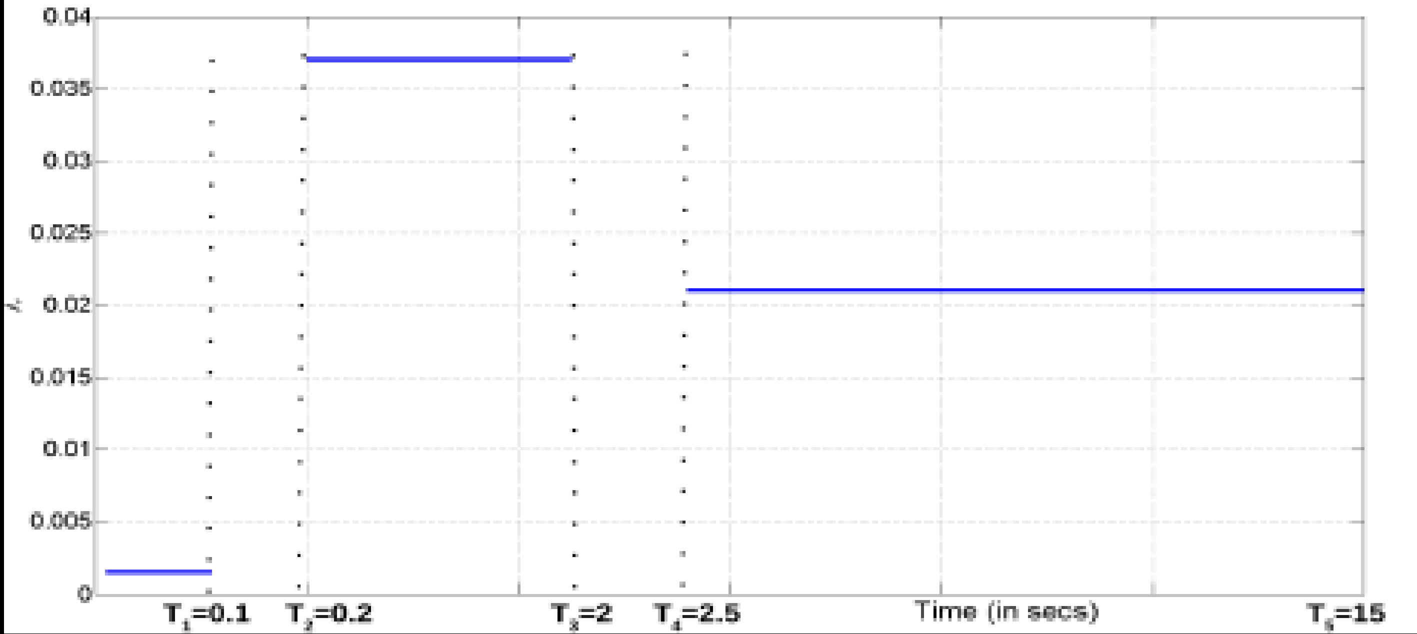
$$(\alpha, \lambda_{t_m - t_n})_{opt} = arg \max_{\alpha,\lambda} p(X_{t_m:t_n} | \mathcal{M}(\alpha, \lambda))$$



Y. Rai , P. Le Callet and G. Cheung« Quantifying the relation between perceived interest and visual salience during free viewing using Trellis based Optimization » IEEE IVMSP16

Modélisation des infuences mutuelles Bottom-up top down

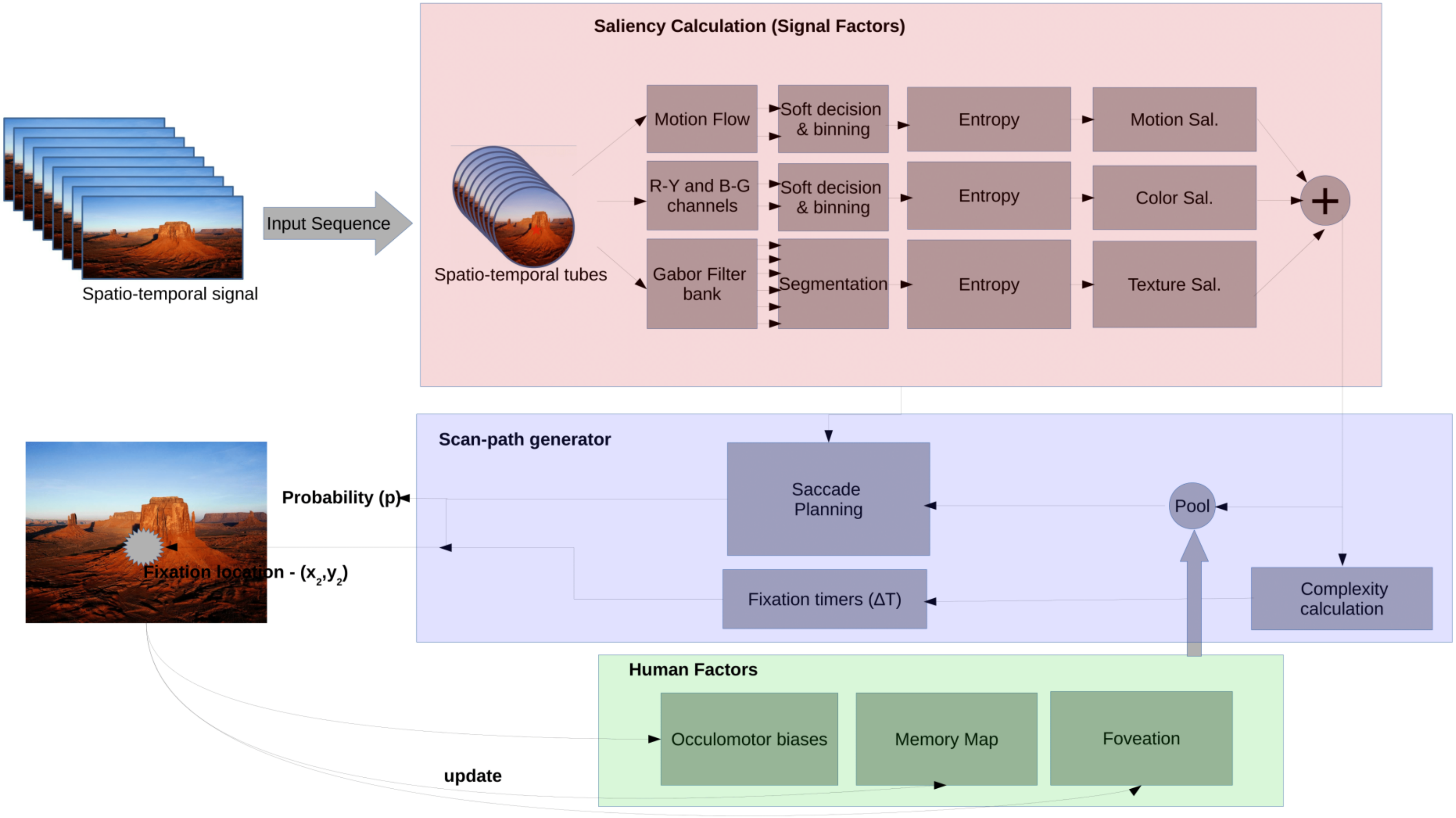
- We determine an optimum λ_t in 3 intervals : Just after the onset of stimuli (<=80ms), Intermediate interval (200ms-2s) and steady state interval >2.5s.



Y. Rai , P. Le Callet and G. Cheung« Quantifying the relation between perceived interest and visual salience during free viewing using Trellis based Optimization » IEEE IVMSP16

Comparer, Analyser … Générer et prédire des « scanpaths »

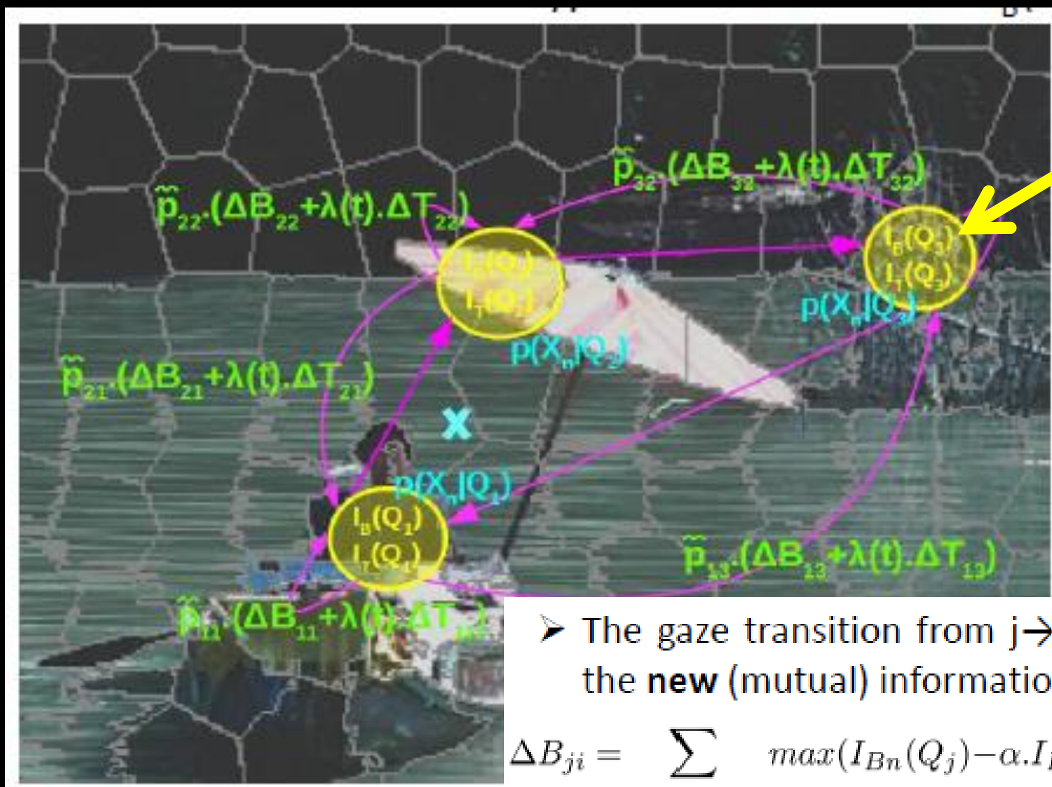
Scanpath model to measure disruption



Vers un modèle de mesure de la «disruption du scanpath

Image = HMM avec comme nœuds cachés des superpixels

Super pixel contient :
information top down
Information Bottom-up



- The gaze transition from j→i (Transition probability M_{ji}) is a function of the **new** (mutual) information: ΔB_{ji} , ΔT_{ji} & the occulomotor bias $\tilde{p}_{ji}$

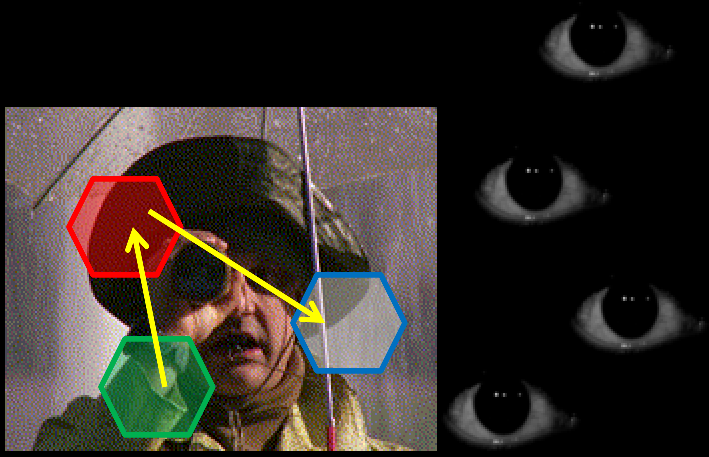
$$\Delta B_{ji} = \sum_{n=\in(Col,Lum,Tex,Mot)} max(I_{Bn}(Q_j) - \alpha.I_{Bn}(Q_i), 0) \qquad \Delta T_{ji} = \sum_{n\in Objects} max(I_{Tn}(Q_j) - \alpha.I_{Tn}(Q_i), 0)$$

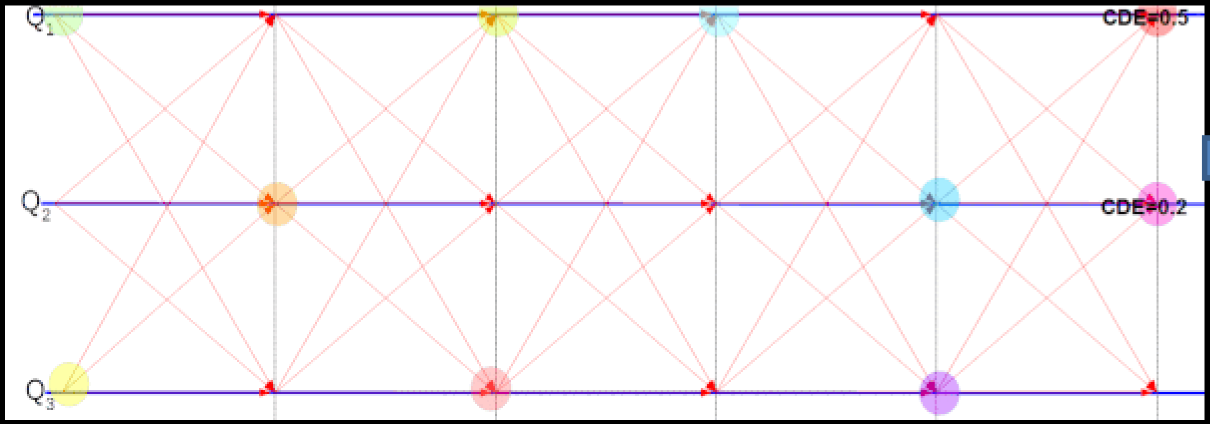
$$M_{ji} = \tilde{p}_{ji}(\Delta B_{ji} + \lambda(t)\Delta T_{ji})$$

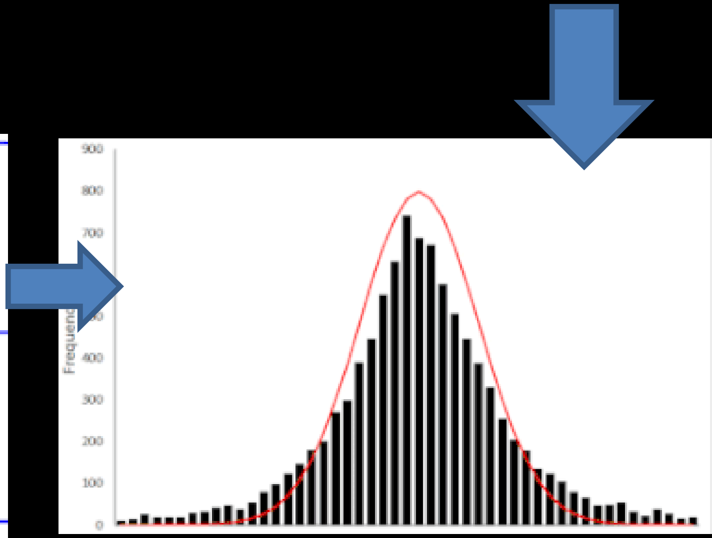
Y. Rai , P. Le Callet and G. Cheung« Quantifying the relation between perceived interest and visual salience during free viewing using Trellis based Optimization » IEEE IVMSP16

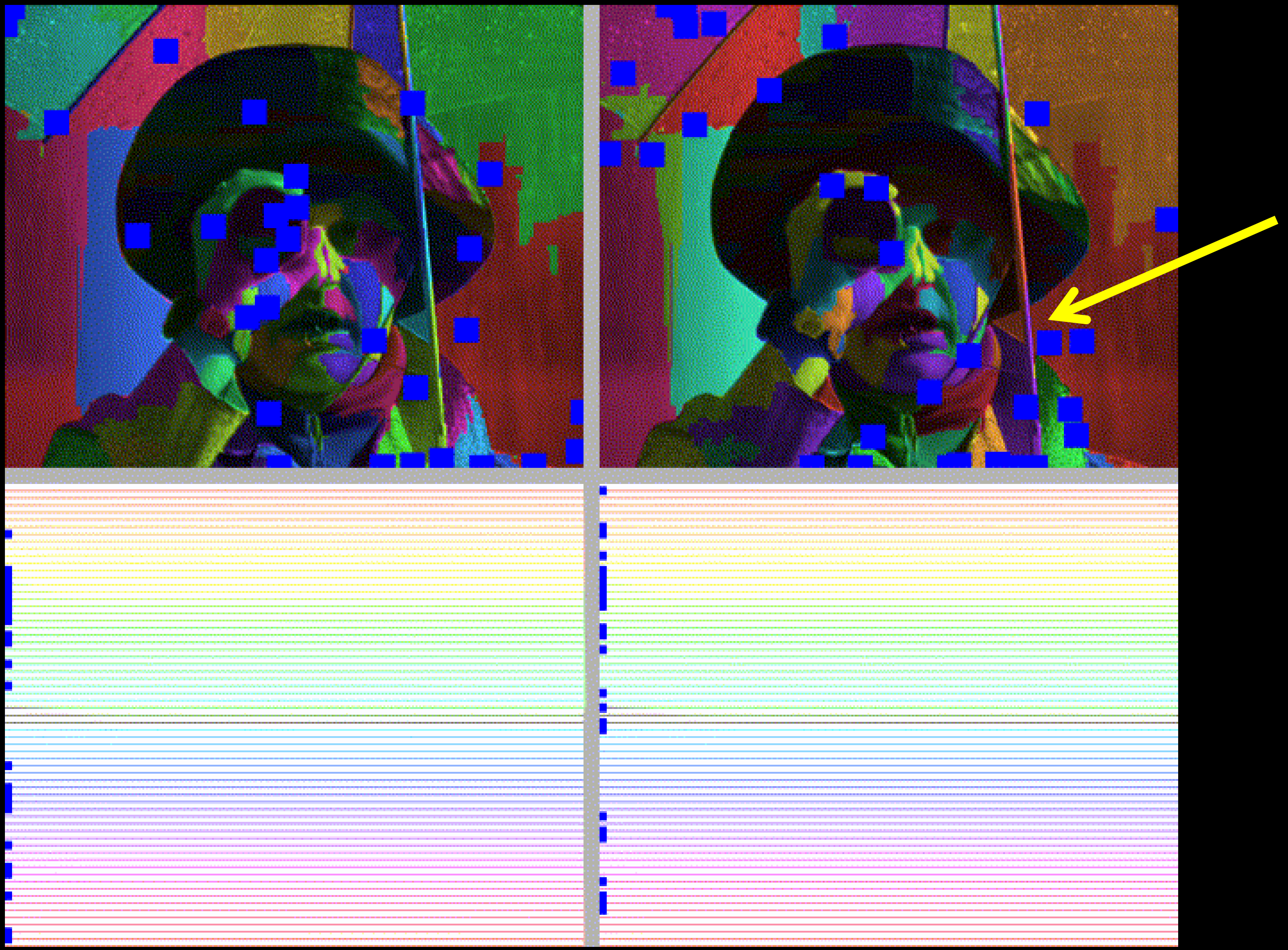
A stochastic model

- N observers who view the video do not have the same perception of quality
- Similarly N simulations of the algorithm should yield a "plausible" but not "deterministic" metric









$$CDE_i = CSE_i\ (S_{impaired} \mid M_{pristine}) - CSE_i\ (S_{pristine} \mid M_{pristine})$$

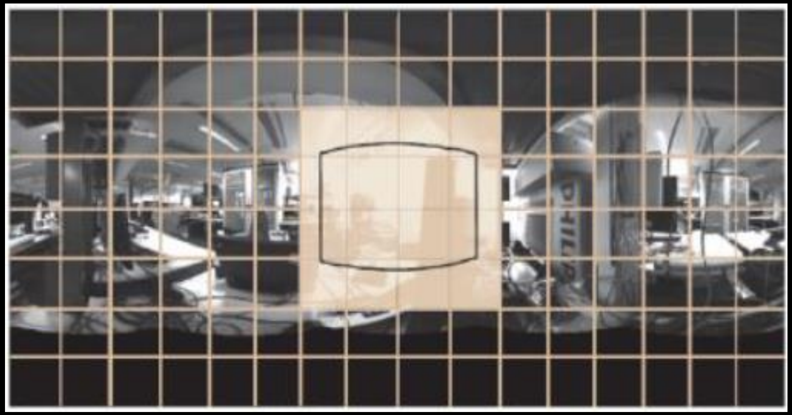
Y. Rai and P. Le Callet, "Do gaze disruptions indicate the perceived quality of non-uniformly coded natural scenes?" in *IS&T/SPIE Electronic Imaging*, 2017

Attention visuelle en 360

360˚ content: Visual attention understanding is useful (Not everything being seen)

Different scenarios compared to conventional image/video viewing:

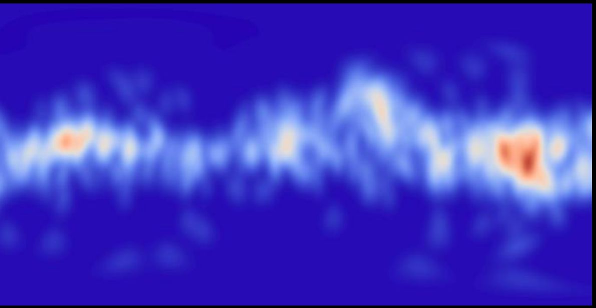
No direct use of 2D VA models for 360˚ content.

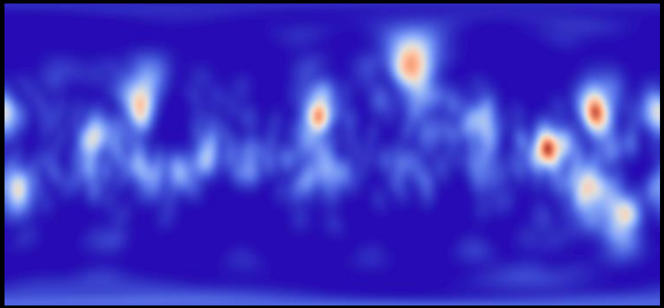


Compétition internationale: prédire l'attention visuelle en 360 (Salient360!)

50 modèles soumis par équipes du monde entier

Given a 360 image





Can you predict the view port location?

Can you predict the view port location and the eye gaze location in the view port?

Can you predict the order of eye gaze fixation?

Visual attention Dataset of 360° images

Equipment:

HMD Oculus Rift DK2

Horizontal and vertical FoV: 100°

Resolution: 960x1080 per eye.

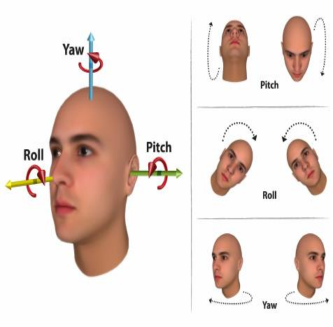
Refresh rate & head-tracking data rate: 75Hz

SMI Eye-tracker

Binocular eye-tracking at 60Hz.

Records

Head Rotation (Pitch, Yaw, Roll), Head Movement (X, Y, Z),
Eye Gaze (Screen x, y), time stamps (display, HMD) per person



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.

Five image classes:

Cityscapes

Naturescapes

Small rooms

Great halls

Scenes containing human faces

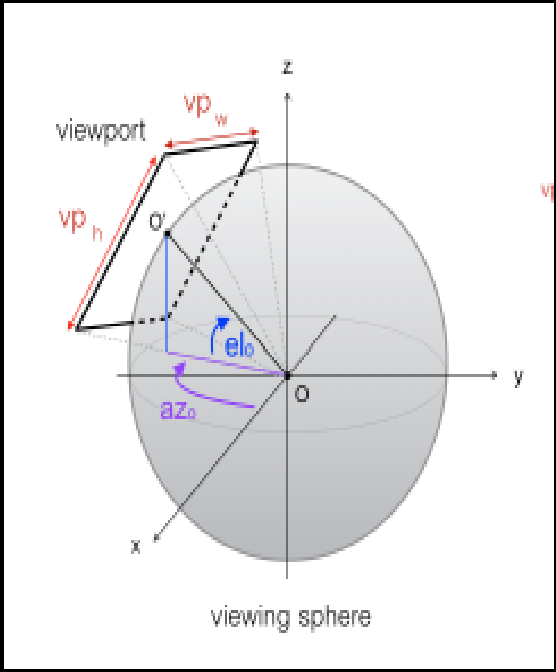
GT gaze to scan-paths and saliency maps

From view port ...

Gaze tracking in the viewport
Classification of gaze data into fixations and saccades
Analysis of the gaze data: location of gaze from center of the viewport

Fixation computation:

- Performed in spherical domain to maintain mathematical precision
- Euclidean distance replaced by orthodromic distances
- Project fixation-points back to equirectangular format for archival and distribution <<Scanpaths track>>

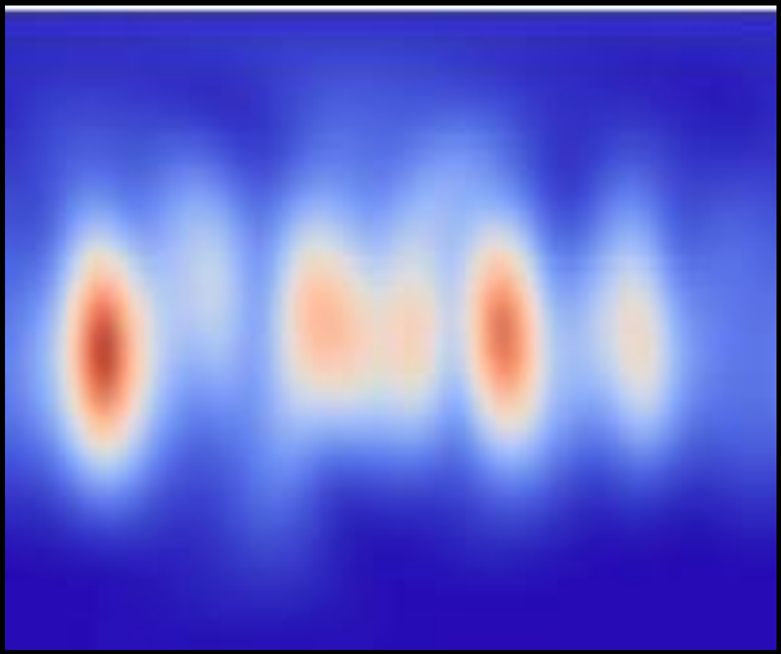


GT gaze to scan-paths and saliency maps

…To saliency map

Back projection on the sphere => equirectangular

Analysis of Saliency maps in equirectangular



Saliency maps :

- Gaussian window foveation performed in viewport domain
- Projected to spherical and subsequently to equirectangular domain

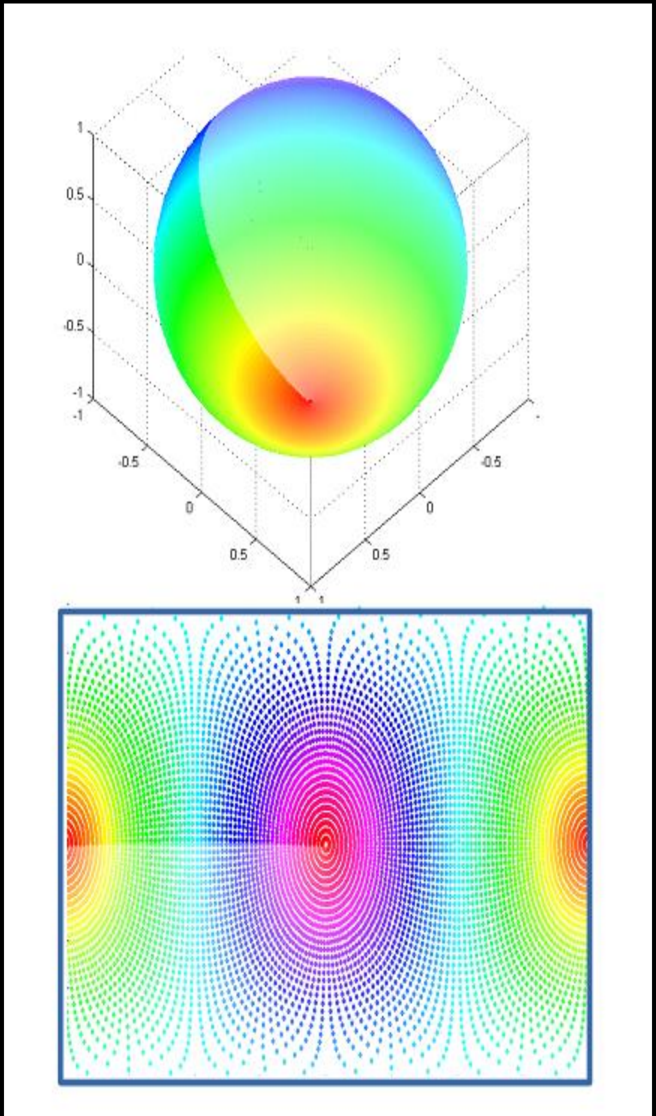
View port in equirectangular

Comparing saliency maps

Cannot compare two equi-rectangular images directly

Uniform sampling on the sphere and re-project to equi-rectangular

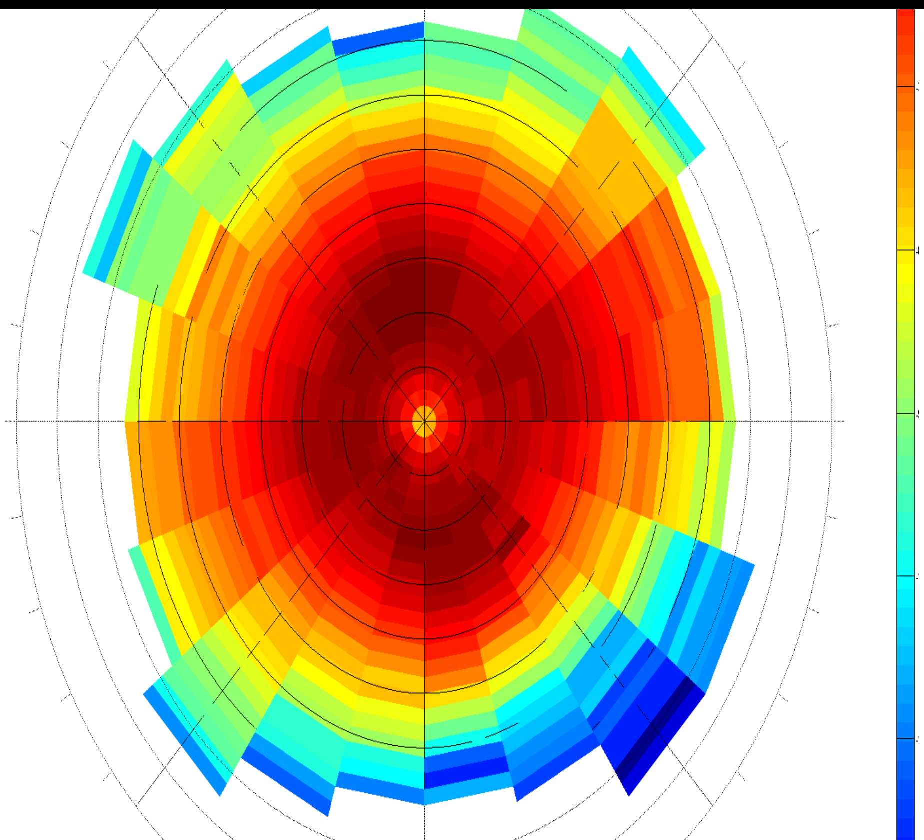
Sample saliency values here only

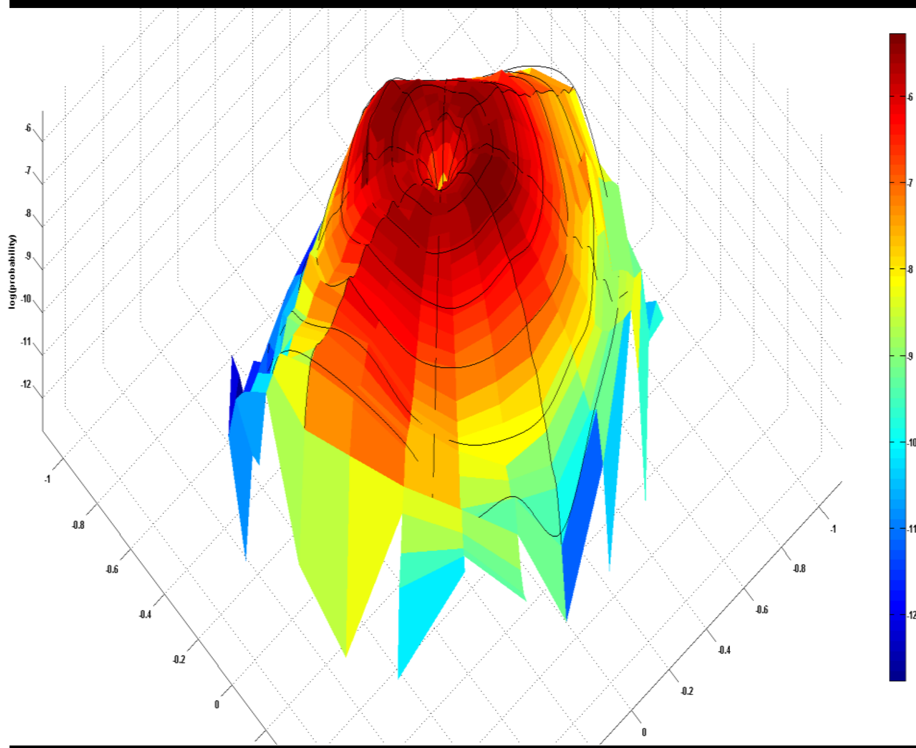


Eye movement statistics

Do people really look at the centre?

Distribution of gaze location form the center of the view port has the shape of a *doughnut*: directionally isotropic





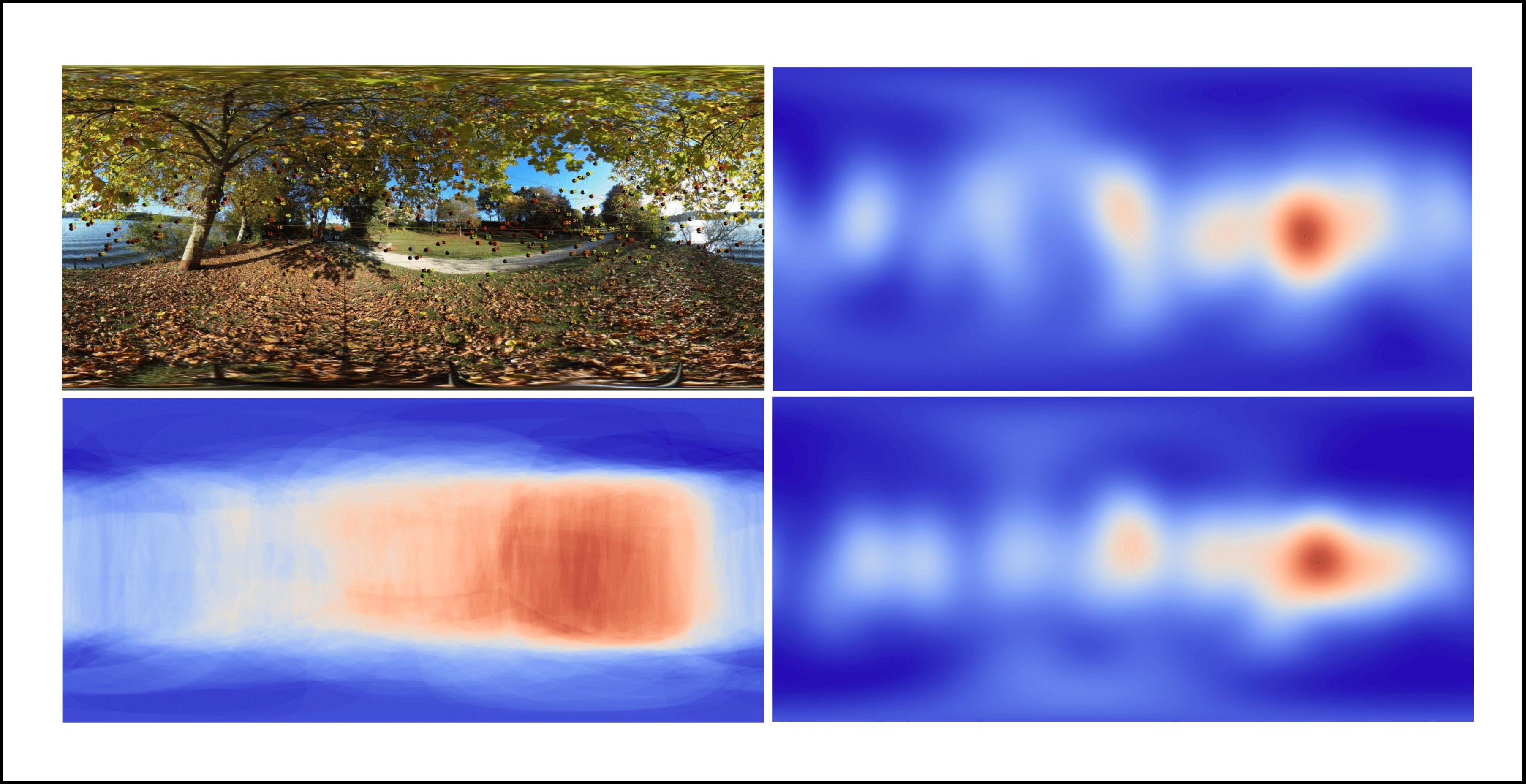
Interpretation

From gaussian to doughnut: An ecological process?

Viewing in HMD is not viewing a fixed display!

- Moving the head more « costly » than moving the eyes

- Motion sickness limitation



[1] M. Yu, H. Lakshman, and B. Girod, “A framework to evaluate omnidirectional video codin schemes,” in *2015 IEEE International Symposium on Mixed and Augmented Reality*, Sept 2015, pp. 31–36
[2] E. Upenik, M. Rerabek, and T. Ebrahimi, “A Testbed for Subjective Evaluation of Omnidirectional Visual Content,” in *32nd Picture Coding Symposium*, 2016

Quality of Experience and user study: direct and indirect approaches

Patrick LE CALLET

Université de Nantes IPI (IMAGE PERCEPTION & INTERACTION) /LS2N

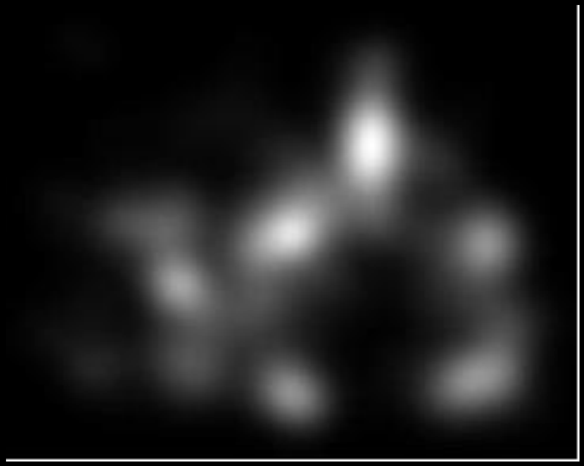
QoE & Artistic intention

Non transparent technology

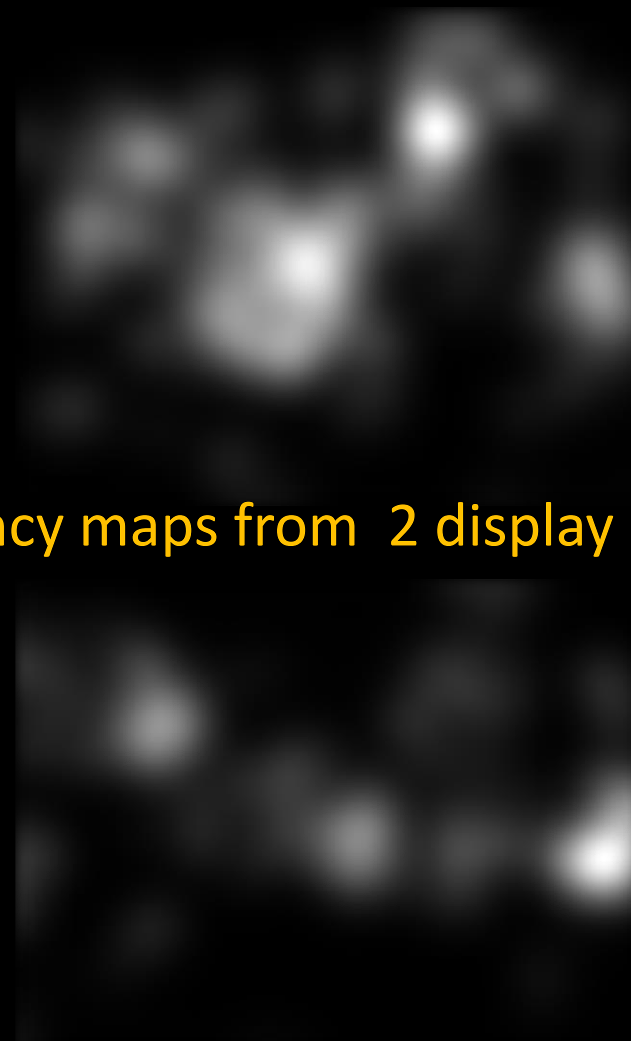
Visual contents can be seen in various conditions

…that can even change the original artistic intention (emotions, image reading …)

Effects on Visual Attention deployement



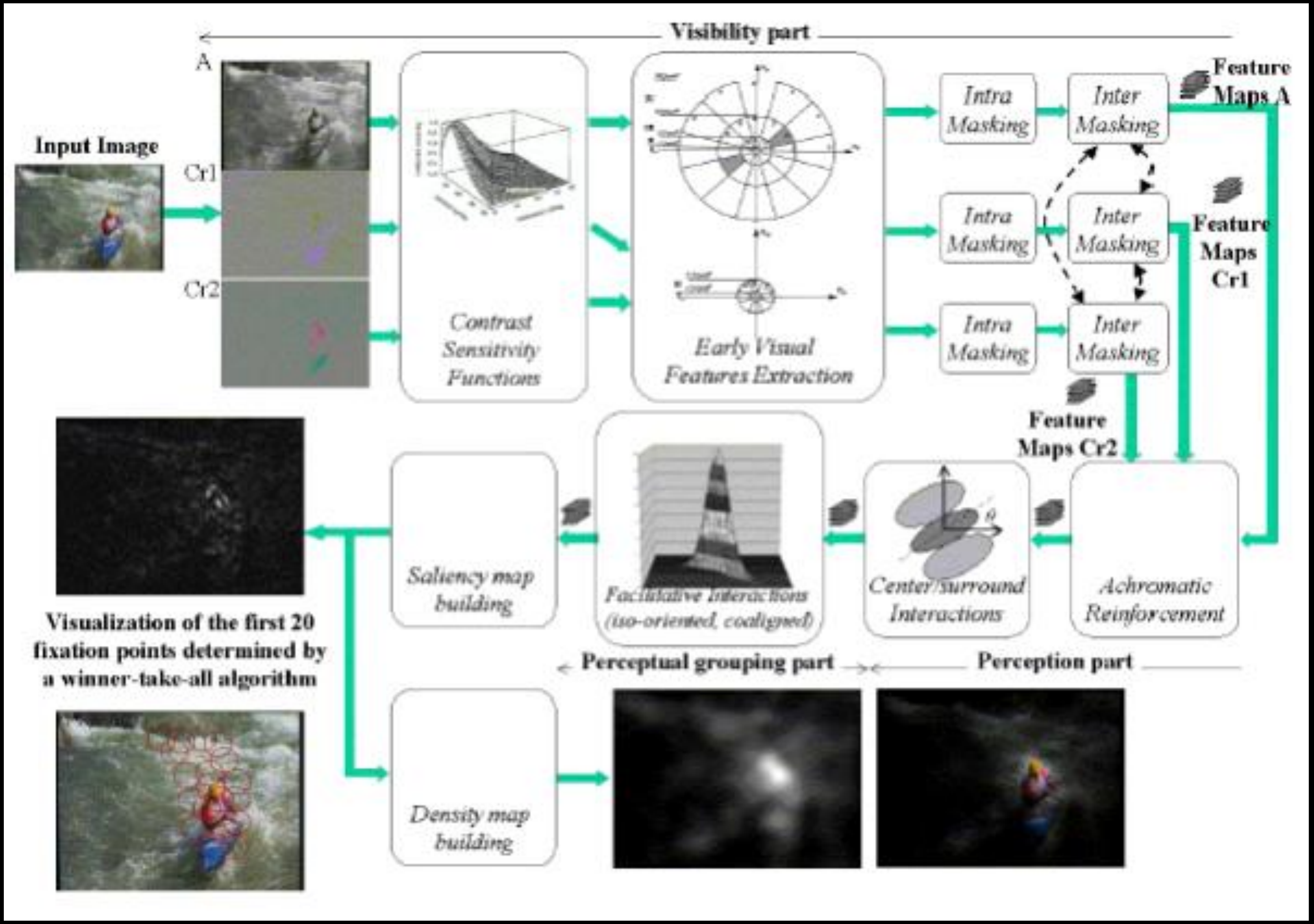
Effects of TMO on Visual Attention deployement



Saliency maps from 2 display conditions

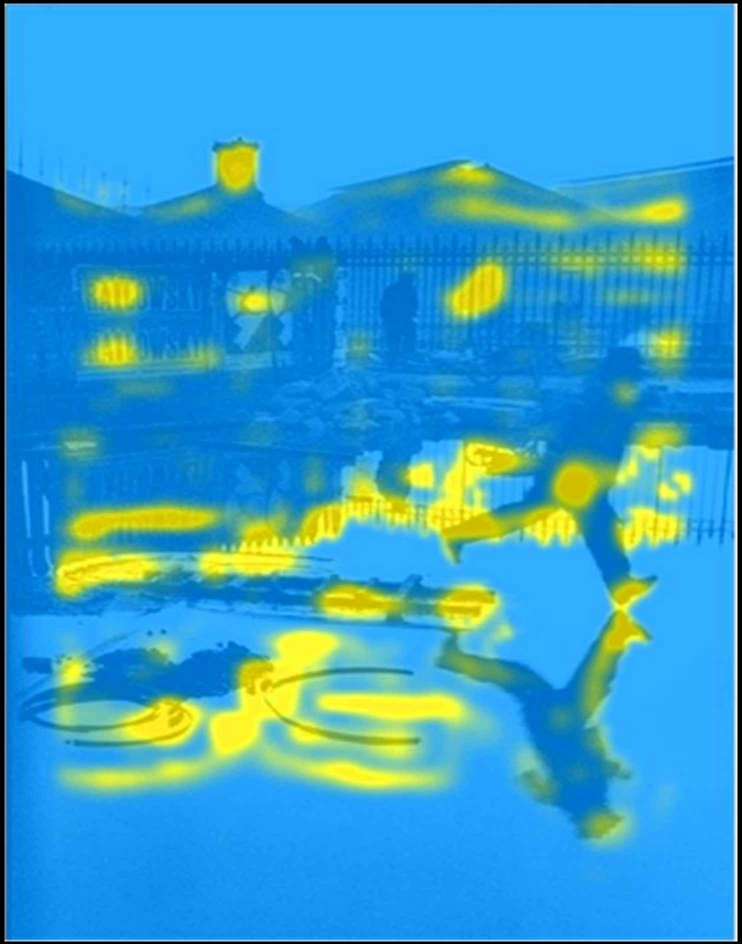
M. Narwaria, M. Silva, P. Callet and R. Pepion "Tone mapping Based High Dynamic Range Compression: Does it Affect Visual Experience?", Signal Processing: Image Communication (Special Issue on Recent Advances in High Dynamic Range Video Research), 2013

Artist intention: can visual attention models be helpful?



A coherent computational Approach to model the bottom-up visual attention
O. Le Meur, P. Le Callet and D. Barba, IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 28, Issue 5, Pages:802-817 , May 2006

Artist intention: visual attention models can help!



On the role of artistic intent of image quality, Scott J. Daly, Electronic Imaging 2008