Attention Visuelle dans les Applications Multimédias

Pr. Patrick Le Callet

www.irccyn.ec-nantes.fr\~lecallet

POLYTECH NANTES

IRCCyN

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

Attention Visuelle @ IRCCYN/IVC @ LS2N/IPI

(Image Perception & Interaction)

Accompagnement des évolutions technologiques multimédias par les sciences de la vision

Approche interdisciplinaire

Informatique Interaction + IHM

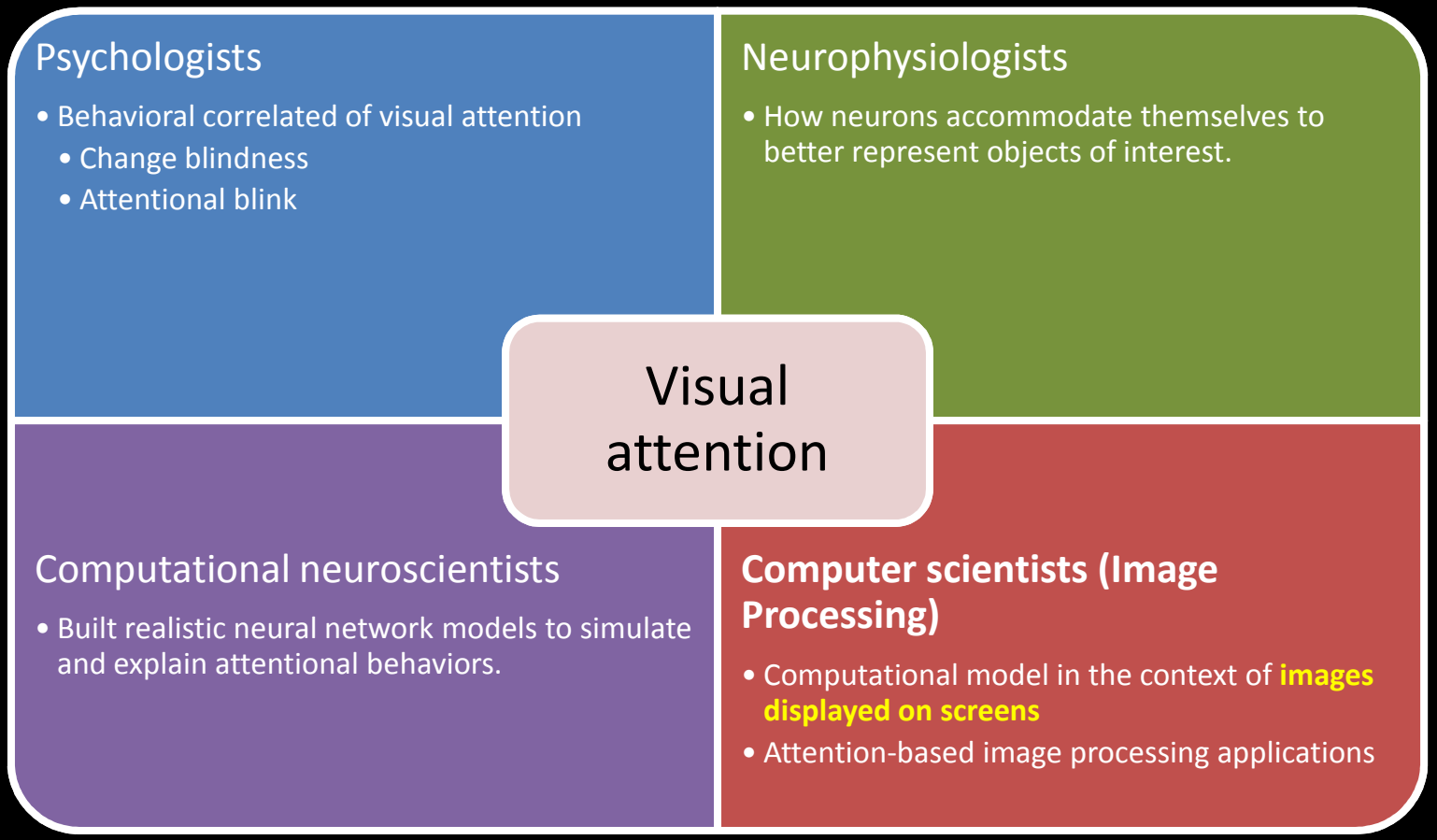
Traitement de signal + science de la vision

Ophtalmologie + Neuroscience

L'attention visuelle: un mécanisme **Ecologique**

….largement étudié

(d'abord en en science de la vision) 38621 occurrences dans pubmed

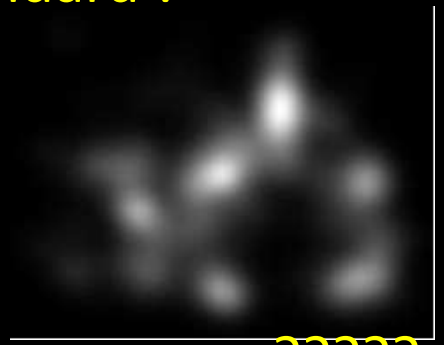


Occulométrie : mesure de localisation du regard

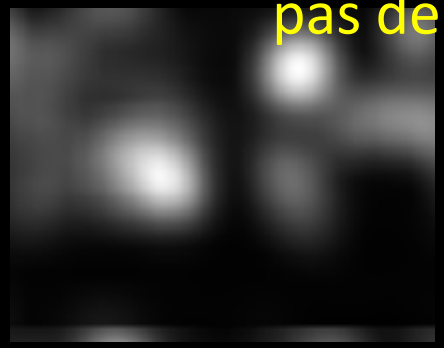
Des données occulométriques aux modèles computationnels

La vérité « terrain »

?????
pas de standard !

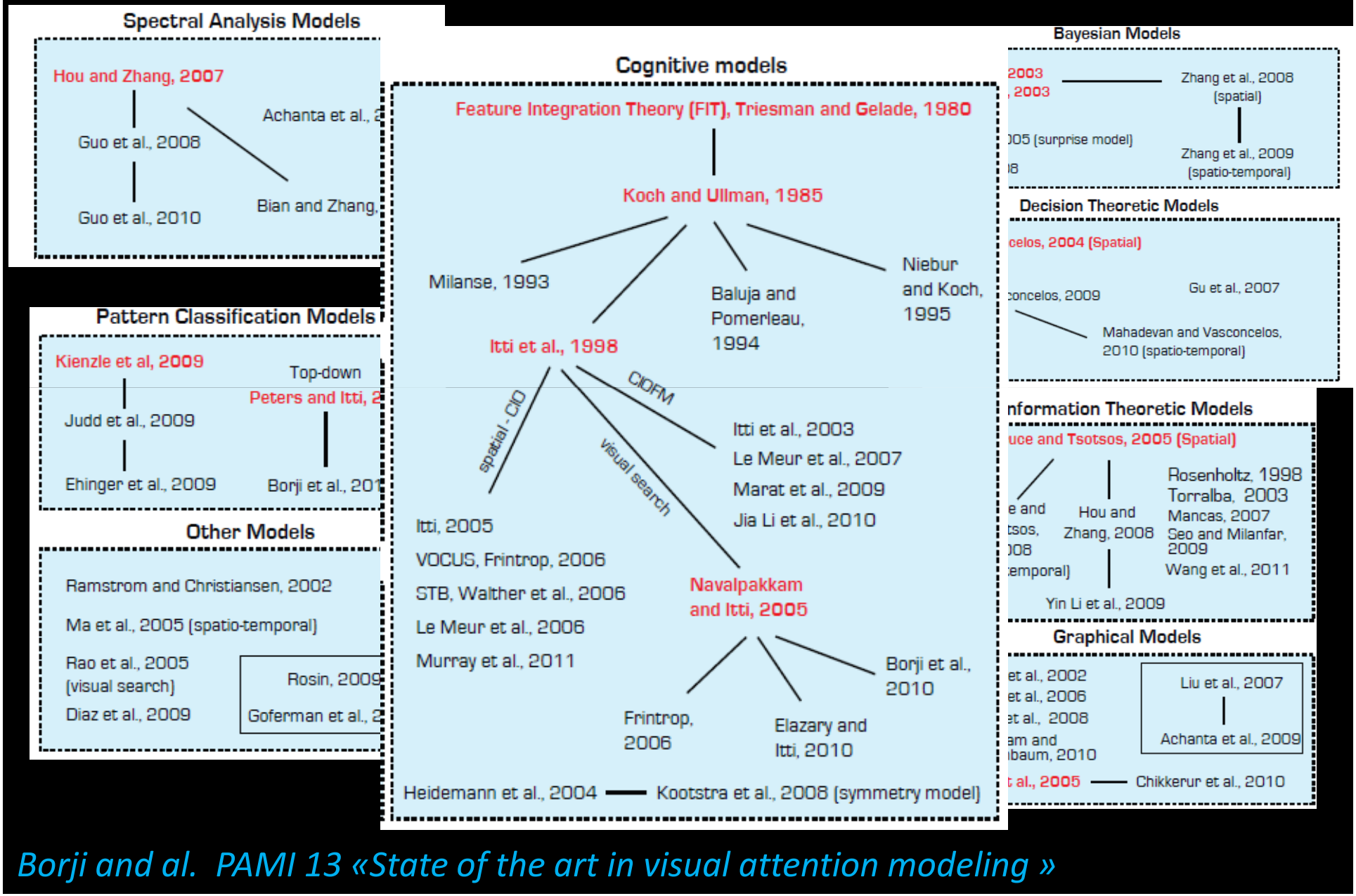


?????
pas de standard !



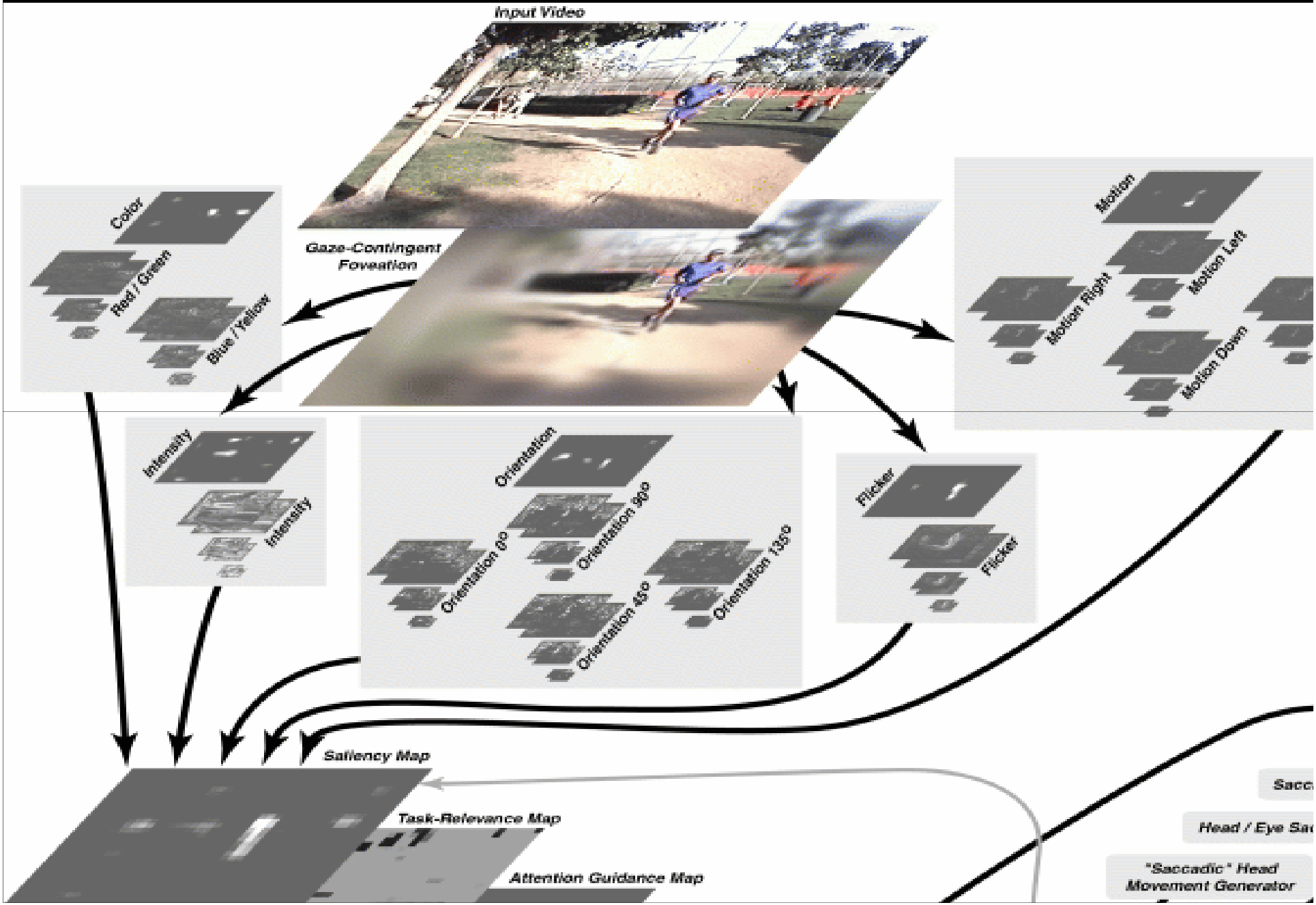
L'objectif (sortie du modèle):
la carte de saillance

Modèle d'attention visuelle: taxonomie



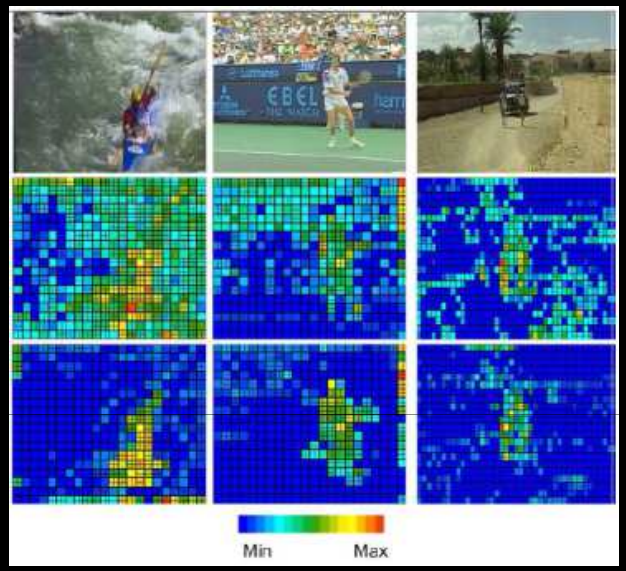
Borji and al. PAMI 13 «State of the art in visual attention modeling »

Saliency model: Feature Integration theory



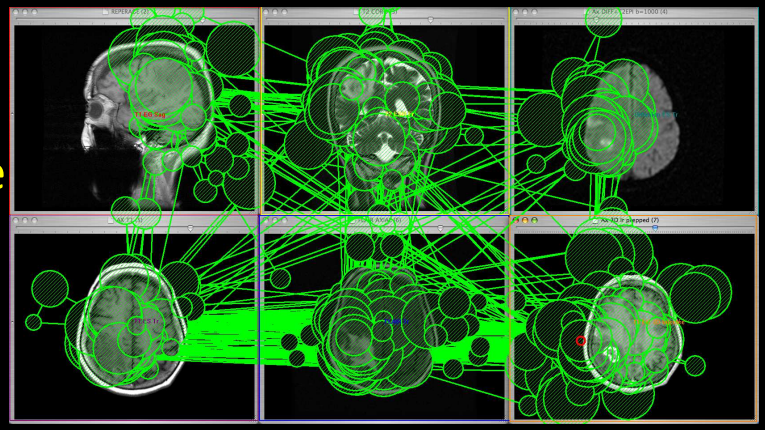
Quelques Applications TIC

Codage d'images



Reformatage de contenus

Ergonomie visuelle



Réalité virtuelle

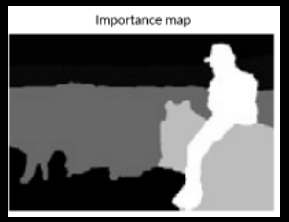
Focus intelligent et adaptatif

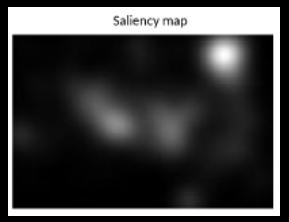
Gestion du conflit vergence/ accomodation

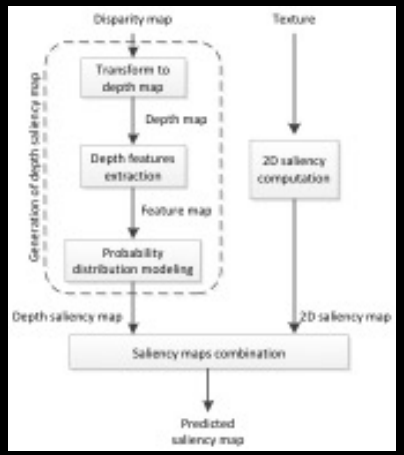
Le Callet and E. Niebur Proc. Of IEEE 2013 «Visual Attention and Applications in Multimedia Technologies»

Quelques Applications TIC (2)

Sous titrage « ergonomique »







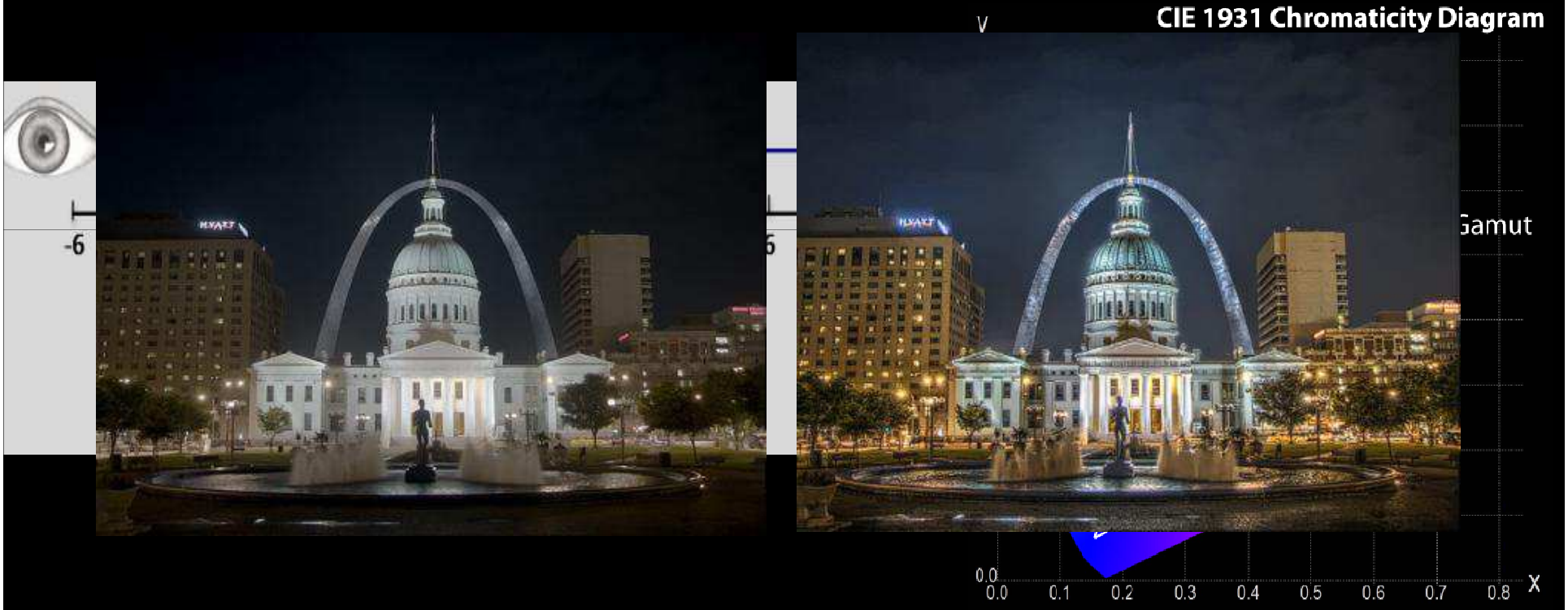
Visualisation 3D stéréo :
Modèle computationnel
JEMR12, IEEE TIP13, IEEE TIP14

=> 3D retargeting, 3D confortable

QoE & Artistic intention

Non transparent technology

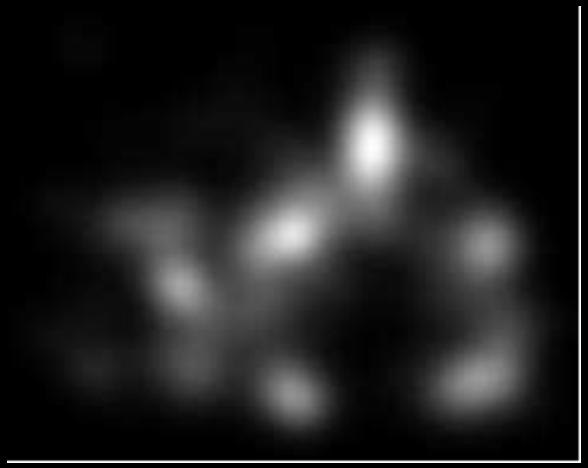
Visual contents can be seen in various conditions



…that can even change the original artistic intention

(emotions, image reading …)

Effects on Visual Attention deployement

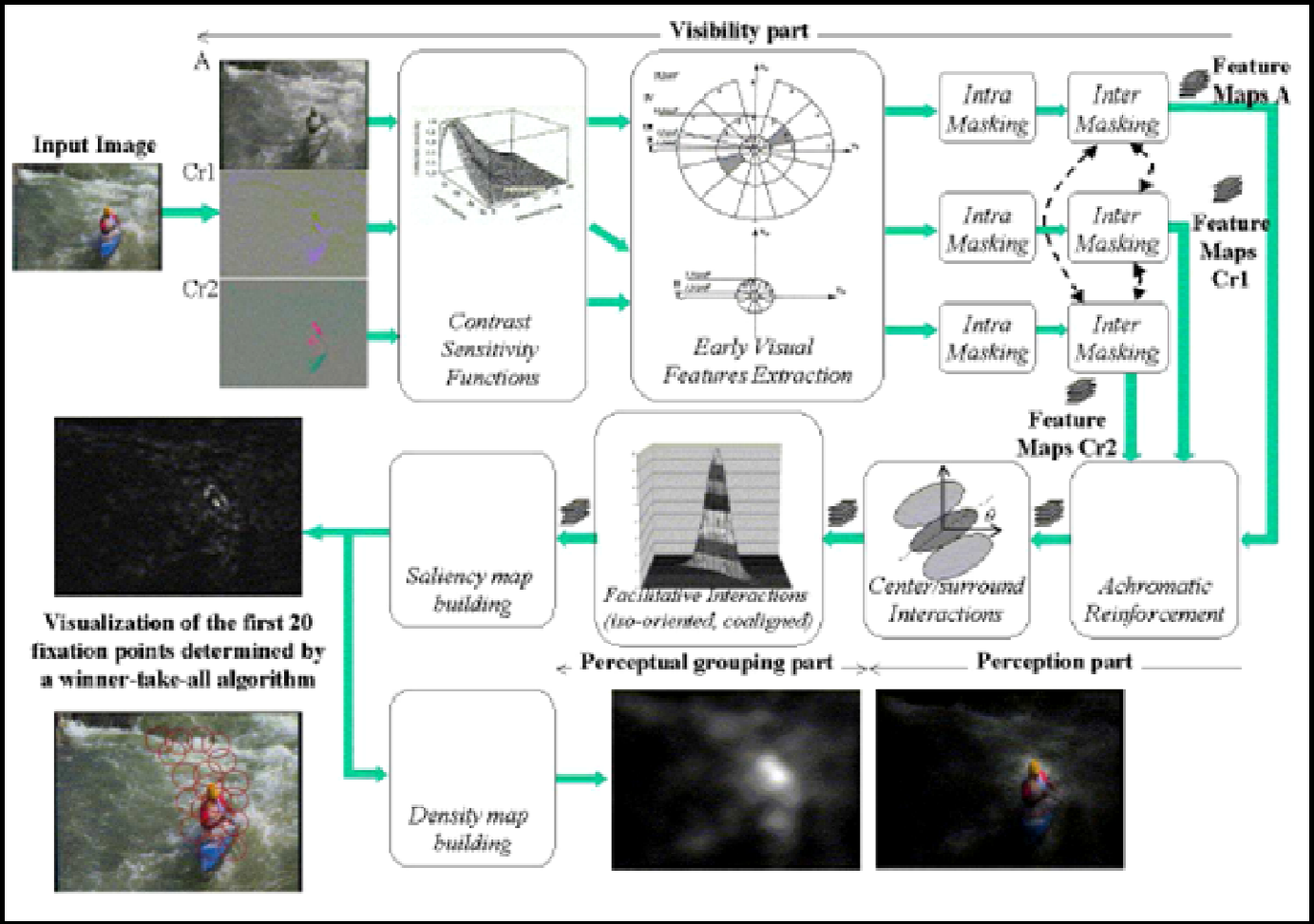


Effects of TMO on Visual Attention deployement

Saliency maps from 2 display conditions

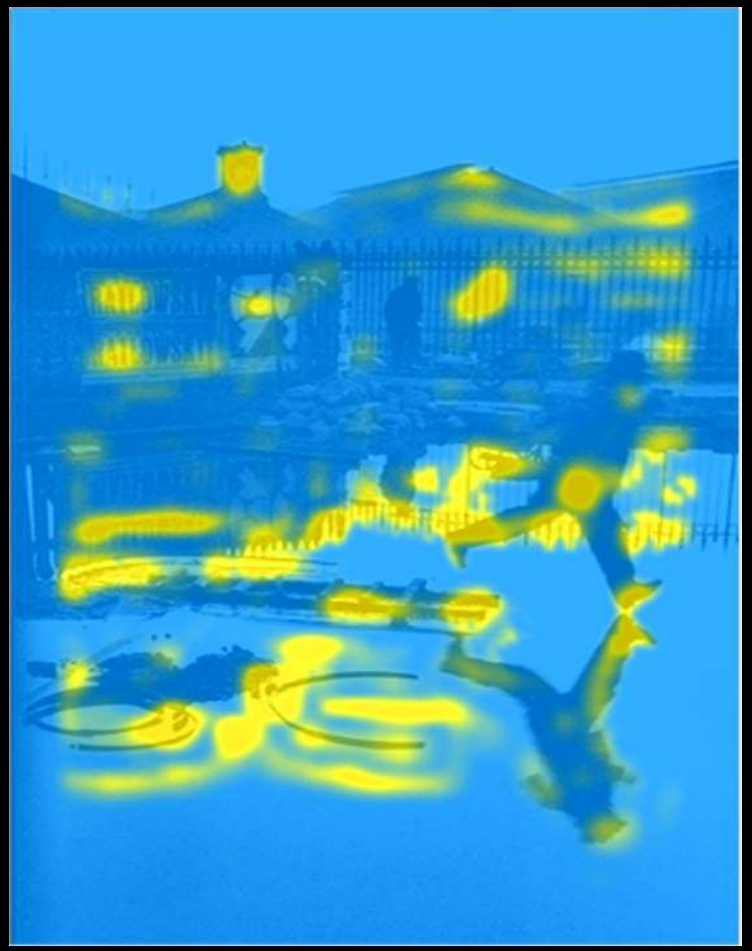
M. Narwaria, M. Silva, P. Callet and R. Pepion “Tone mapping Based High Dynamic Range Compression: Does it Affect Visual Experience?”, Signal Processing: Image Communication (Special Issue on Recent Advances in High Dynamic Range Video Research), 2013

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

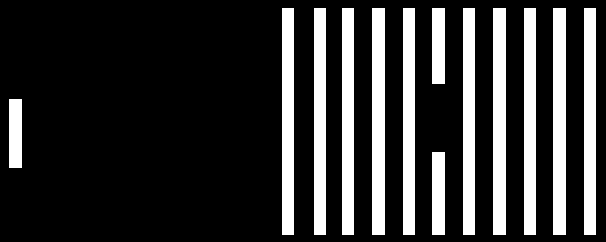
Saillance et Compression vidéo

Du tatouage d'images à une
représentation efficace de la saillance

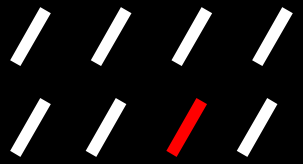
M. Ammar, M. Mitrea, I. Boujelben, et P. Le Callet, « HEVC saliency map computation », in Electronic Imaging, 2016

Saliency = f(Singularity)

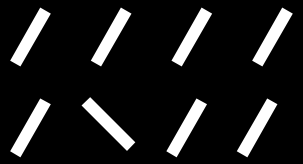
- Contrast



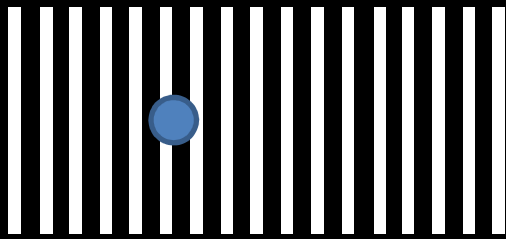
- Color



- Orientation



- Motion



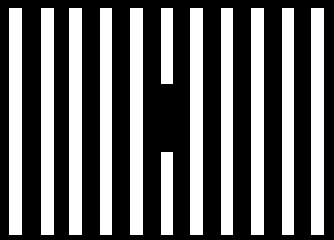
Saliency, singularity and image/video coding

Source coding: efficient representation
removing spatial/temporal redundancy

Core Idea:

Saliency = Singularity = Non-redundancy

$\Rightarrow$ **Hard to encode**



Seminal work: Saliency as Incremental Coding Length (ICL)

For certain lossy coding scheme

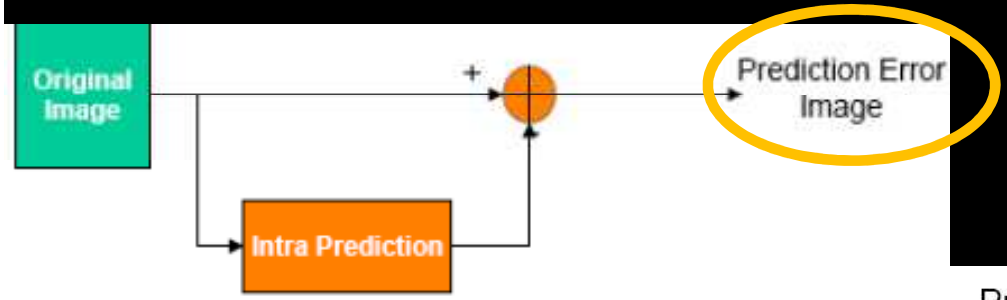
Li and al. ICIP09 «Incremental Sparse Saliency Detection»

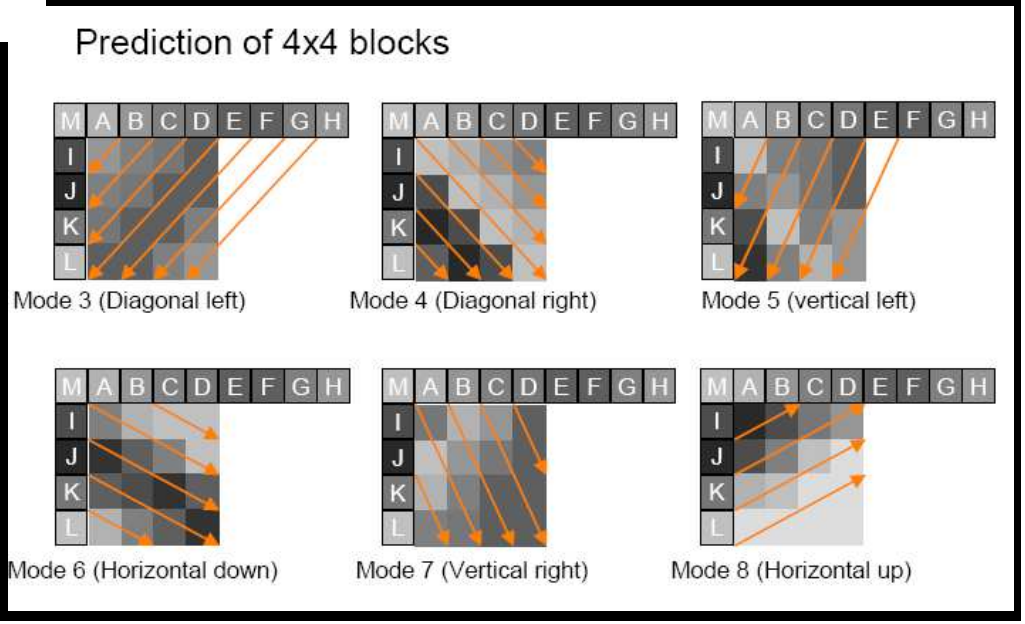
video coding & redundancy

Spatial and temporal redudancy removal:

⇒ Intra & inter prediction

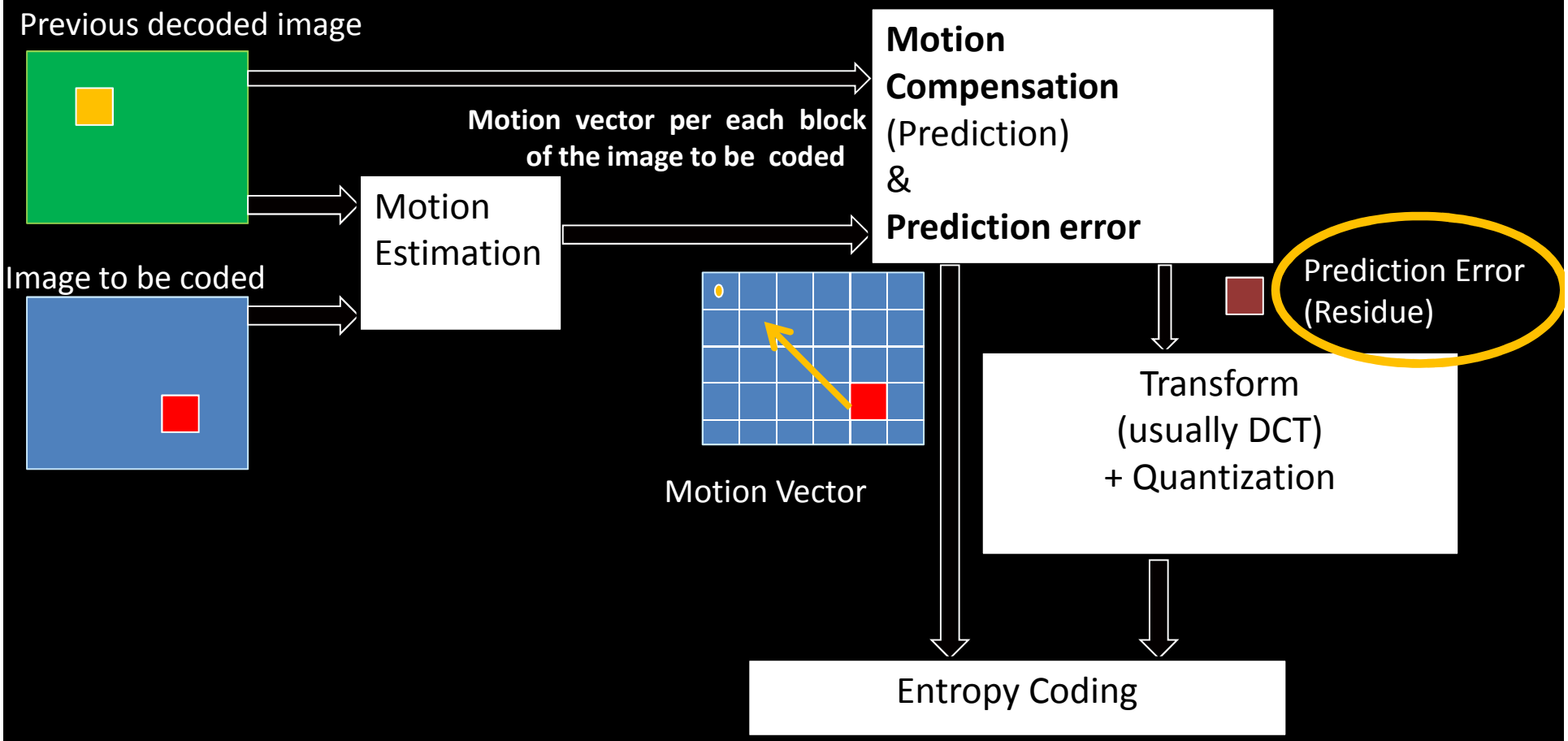
from previous coded/decoded information (e.g: a neighbor block, a bloc from a previous frame)





video coding & redundancy

spatial and temporal redudancy removal: inter prediction



Main idea: hybrid model for saliency prediction

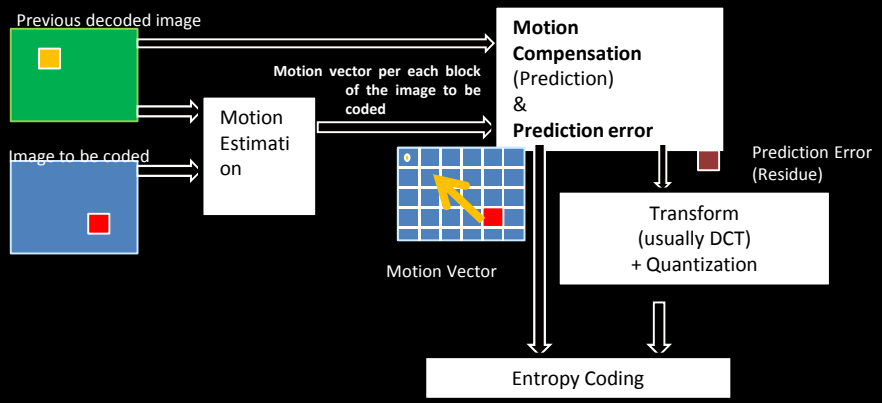
bitstream based

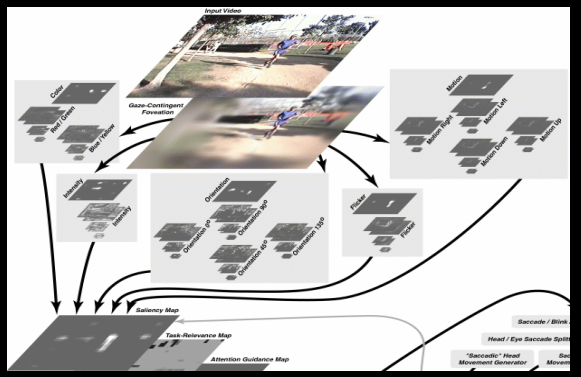
Combining

source coding representation (saliency/singularity is hard to encode)

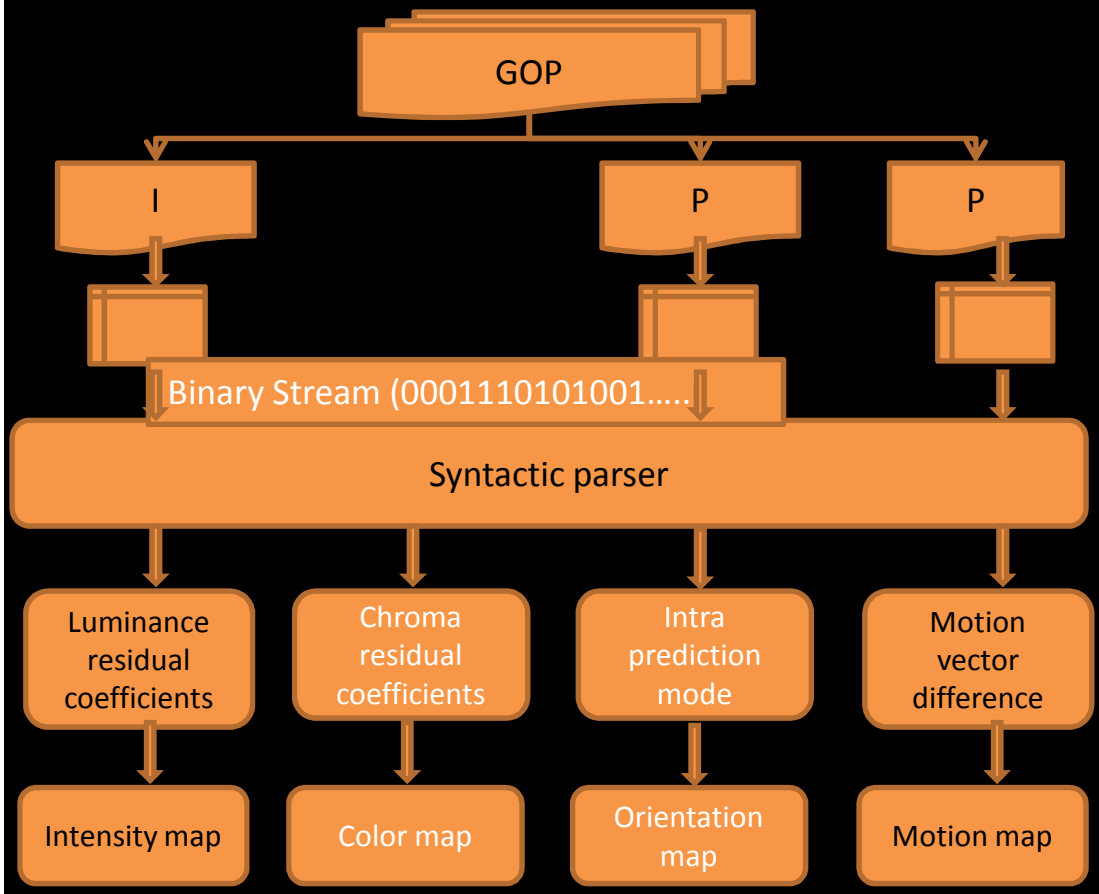
&

FIT scheme





Saliency in MPEG/AVC: a first POC



Competitive approach compared to SoA

MPEG-4 AVC, although not designed by exploiting saliency principles standard preserves the visual saliency at the stream syntax elements level

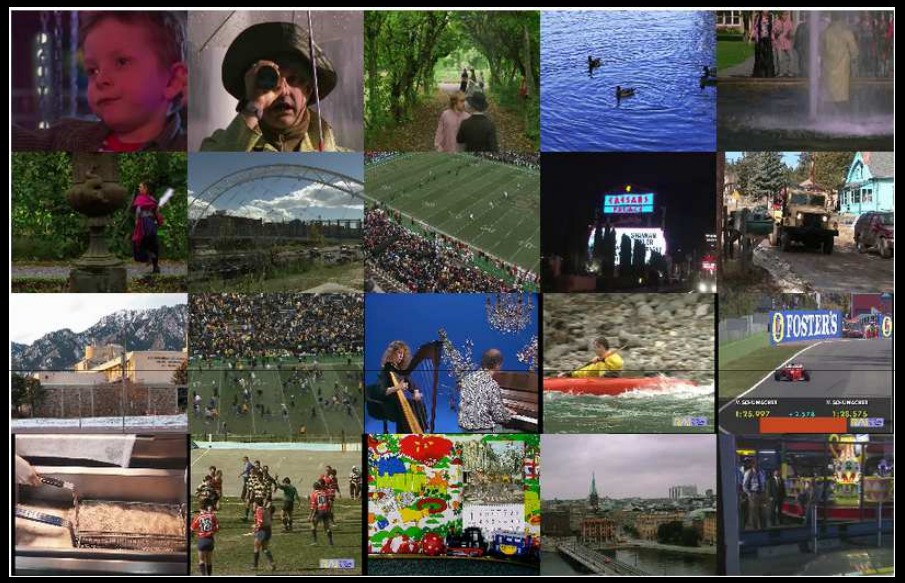
Ammar and al. HVEI15 «Saliency map for MPEG-4 AVC »

Test set up

- Eye tracking Dataset

  http://ivc.univ-nantes.fr/en/databases/Eyetracker_SD_2009_12/

  30 observers



Engelke and al. Qomex10 «Modelling saliency awareness for objective video quality assessment»

- Encoding: HEVC reference software (JCT-VC HEVC)
- Comparison with 3 soA uncompressed domain methods references

  + previous study devoted to the MPEG-4 AVC

Results

	Ming[7]	**Hae**[12]	**Gof**[13]
KLD gain	0.16	0.13	0.01
AUC gain	0.18	0.15	0.13

MPEG-4 AVC, HEVC standard preserves the visual saliency at the stream syntax elements level !

A perpetually friendly representation for saliency (much better than for quality purposes)

M. Ammar, M. Mitrea, I. Boujelben, et P. Le Callet, « HEVC saliency map computation », Electronic Imaging, 2016

M. Ammar, M. Mitrea, M. Hasnaoui, et P. Le Callet, « Visual saliency in MPEG-4 AVC video stream », Electronic Imaging, 2015

L'attention visuelle: un mécanisme **Ecologique**

….largement étudié

(d'abord en en science de la vision)

….mais pas forcément bien approprié/défini

(dans les communautés « applicative »: computer vision, traitement /communication images)

Le Callet and Nieburst « Visual Attention and Applications in Multimedia Technologies », Proceedings of IEEE 2013

La communauté « computer vision » s'en(m)mêle

Frequency Tuned salient region detection

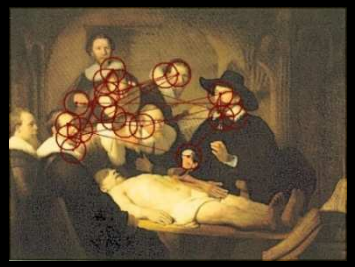
R. Achanta, S. Hemami, F. Estrada, S. Susstrunk

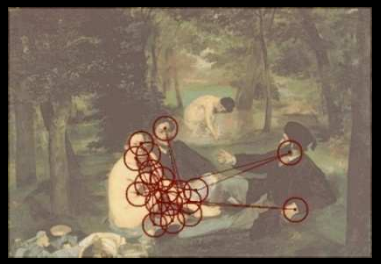
CVPR 2009

Types de modèles et de vérité terrain

Prédictions possibles :

- Chemin visuel





- Région d'intérêt perçu

- Carte de saillance

Pas les mêmes applications !

Warning : différences conceptuelles

"Concepts such as saliency, region-of-interest, visual attention, and gaze patterns are frequently used interchangeably in image processing and computer vision research communities ***but they should not…"***

Over vs covert

Top down vs Bottom up

RoI vs Saliency

Perceived Interest vs Overt Visual Attention

Engelke and al. HVEI 15 « Perceived Interest Versus Overt Visual Attention in Image Quality Assessment »

Le Callet and al., PIEEE« Visual Attention and Applications in Multimedia Technologies »

Warning (2) : différences méthodologiques

Design and validation of the models need “ground truth” data (=> experimental collection of the data)

quelle vérité terrain pour quel modèle (et pou quelle application)?

eye gaze tracking vs ROI selections

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”

Top down vs Bottom up

Un autre regard sur la vérité terrain occulométique

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

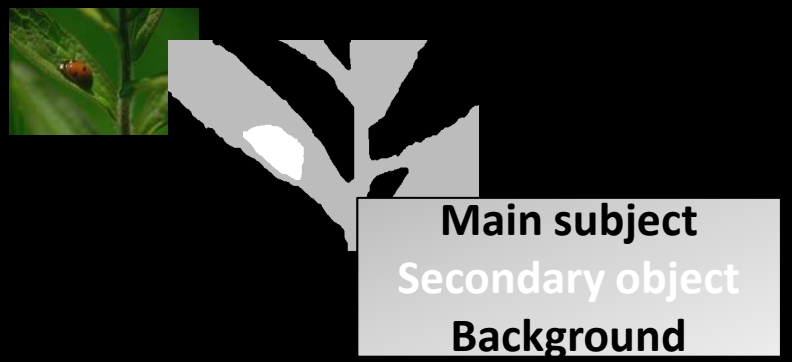
- Two types of ground truths for visual attention
  - Fixation density map (visual saliency, bottom-up)
  - Region of interest (visual importance, top-down)

- 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 »

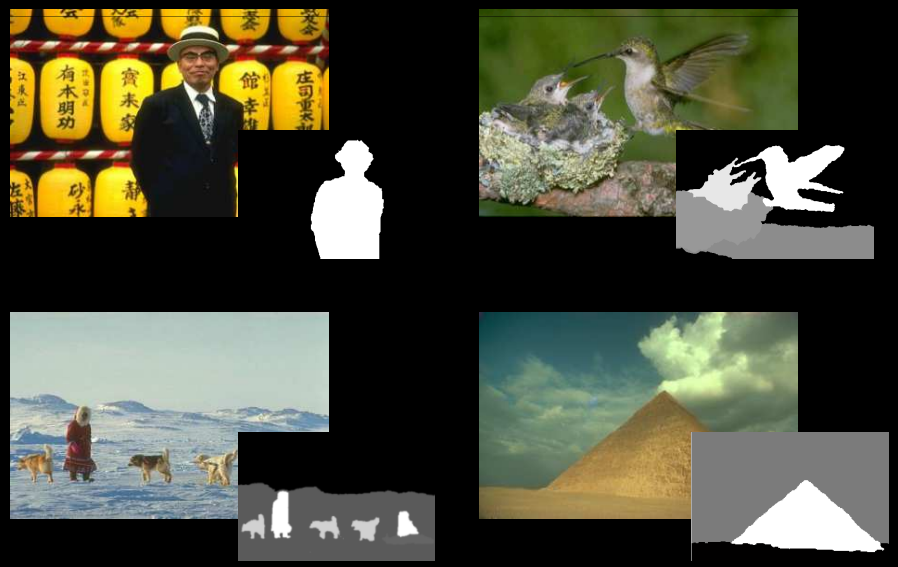
Experiment I: a scoring experiment

- 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

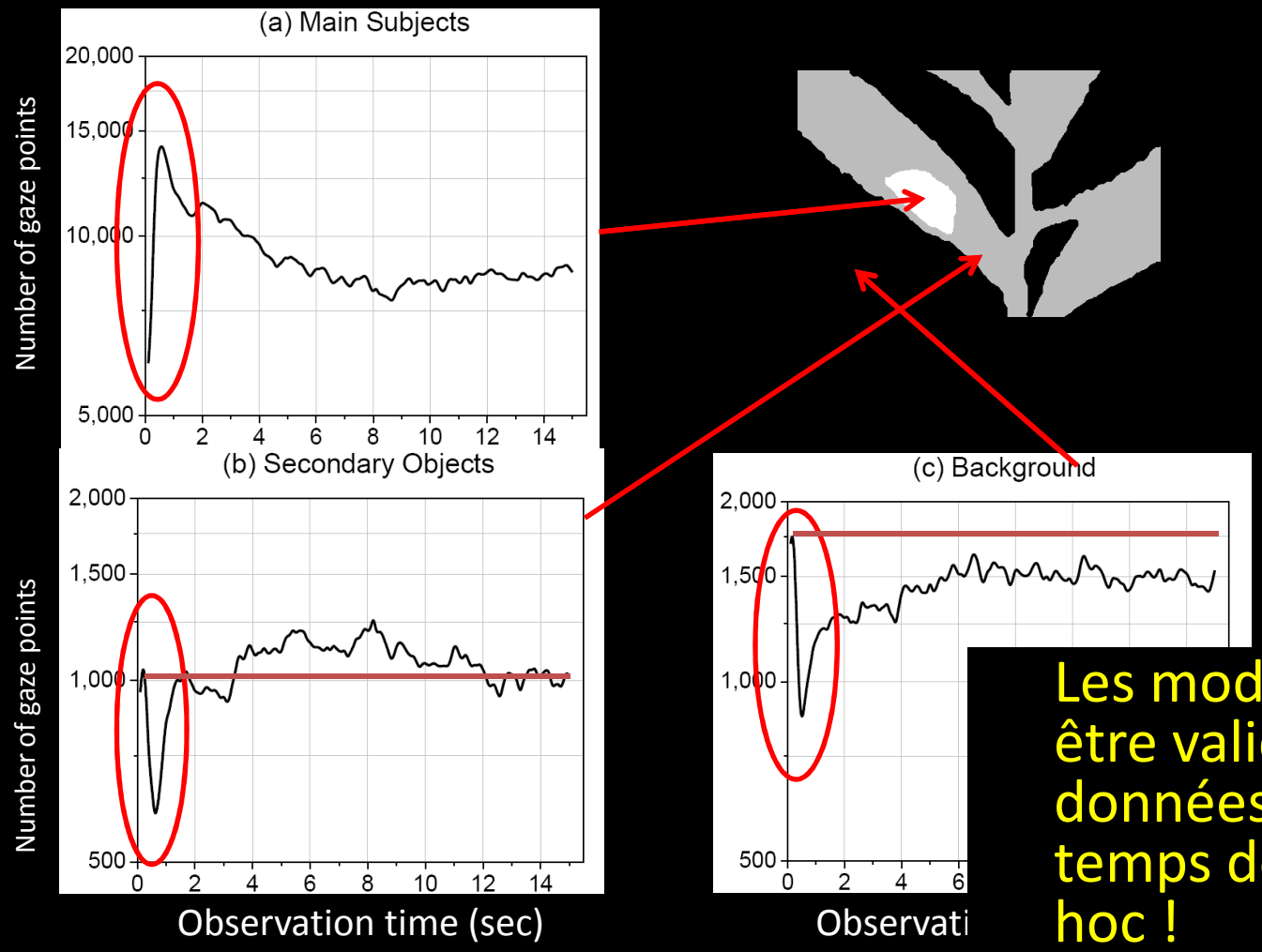

- Task
  - Free-viewing

- 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

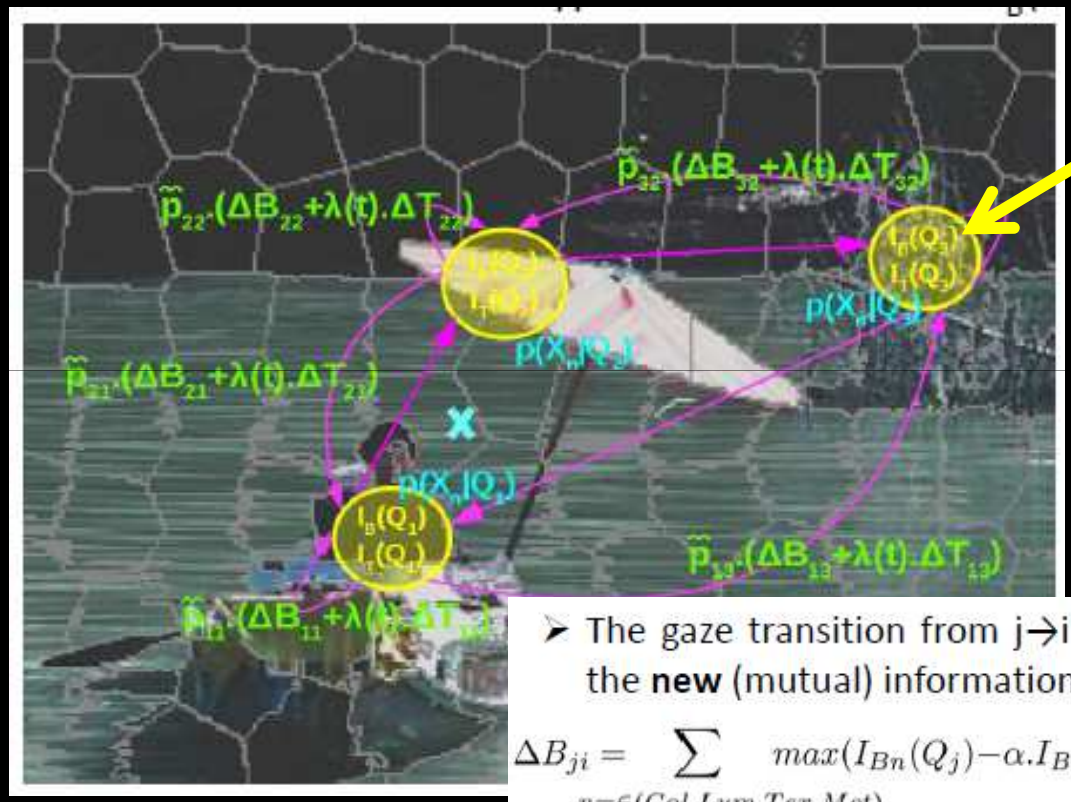


Les modèles devraient être validés avec les données correspondant au temps de visualisation ad hoc !

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 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 $\tilde{p}_{ji}$

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

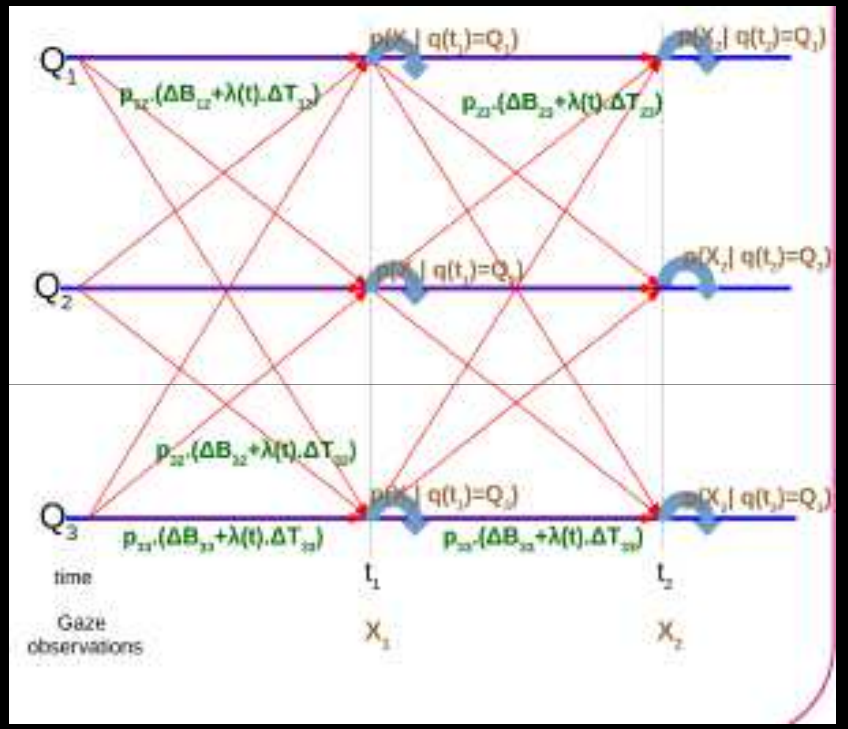
$$M_{ji} = \tilde{p}_{ji}(\Delta B_{ji} + \lambda(t)\Delta T_{ji})$$

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

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

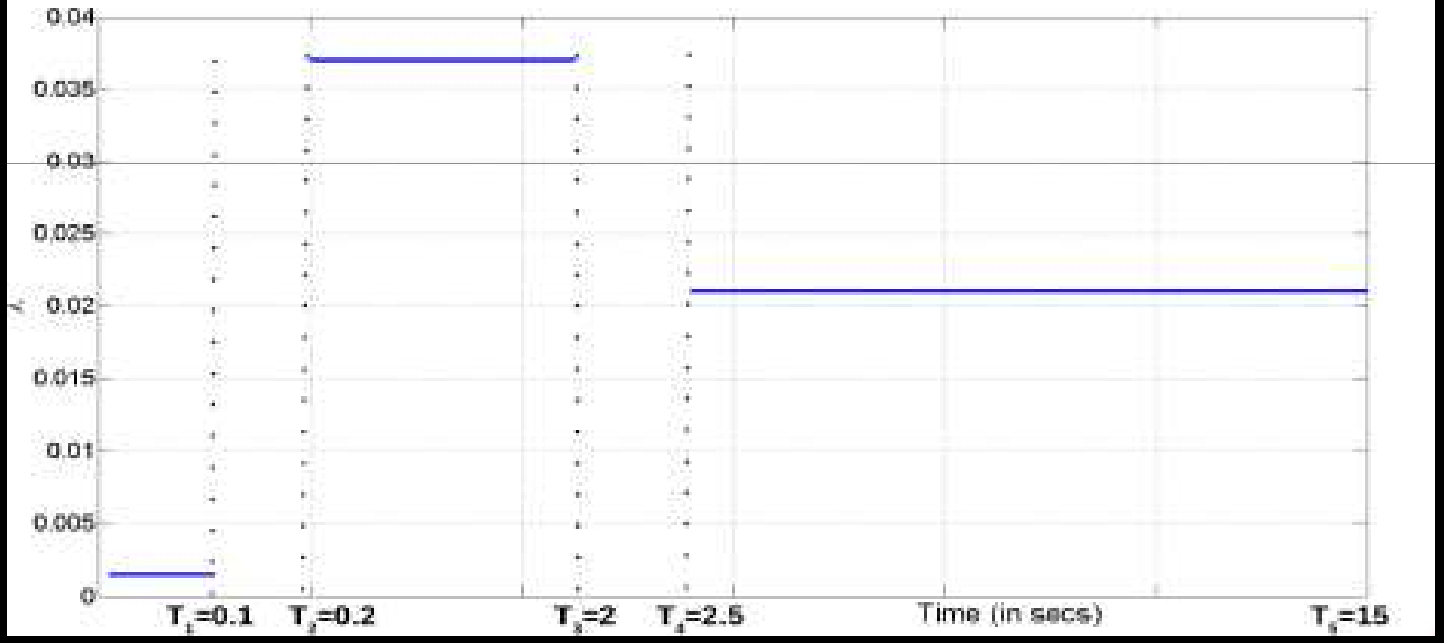
$$(\alpha, \lambda_{t_m - t_n})_{opt} = arg \max_{\alpha,\lambda} p(X_{t_m:t_n} | \mathcal{M}(\alpha, \lambda))$$



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.

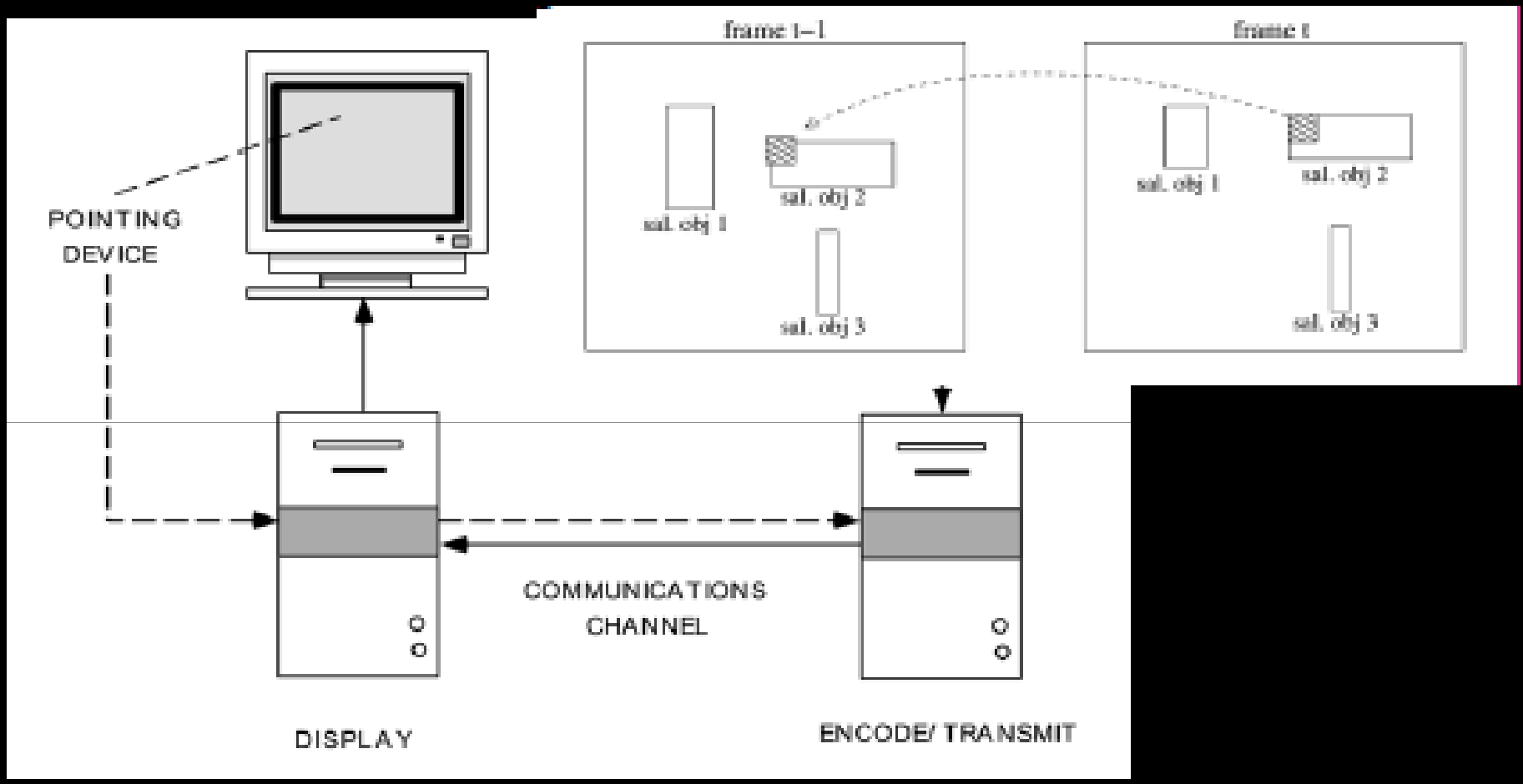


Y. Rai , P. Le Callet and G. Cheung« Quantifying the relation between perceived interest and visual salience during free viewing using Trellis based Optimization » IEEE IVMSP16

Streaming Interactif

Vers le développement de modèles saccadiques?

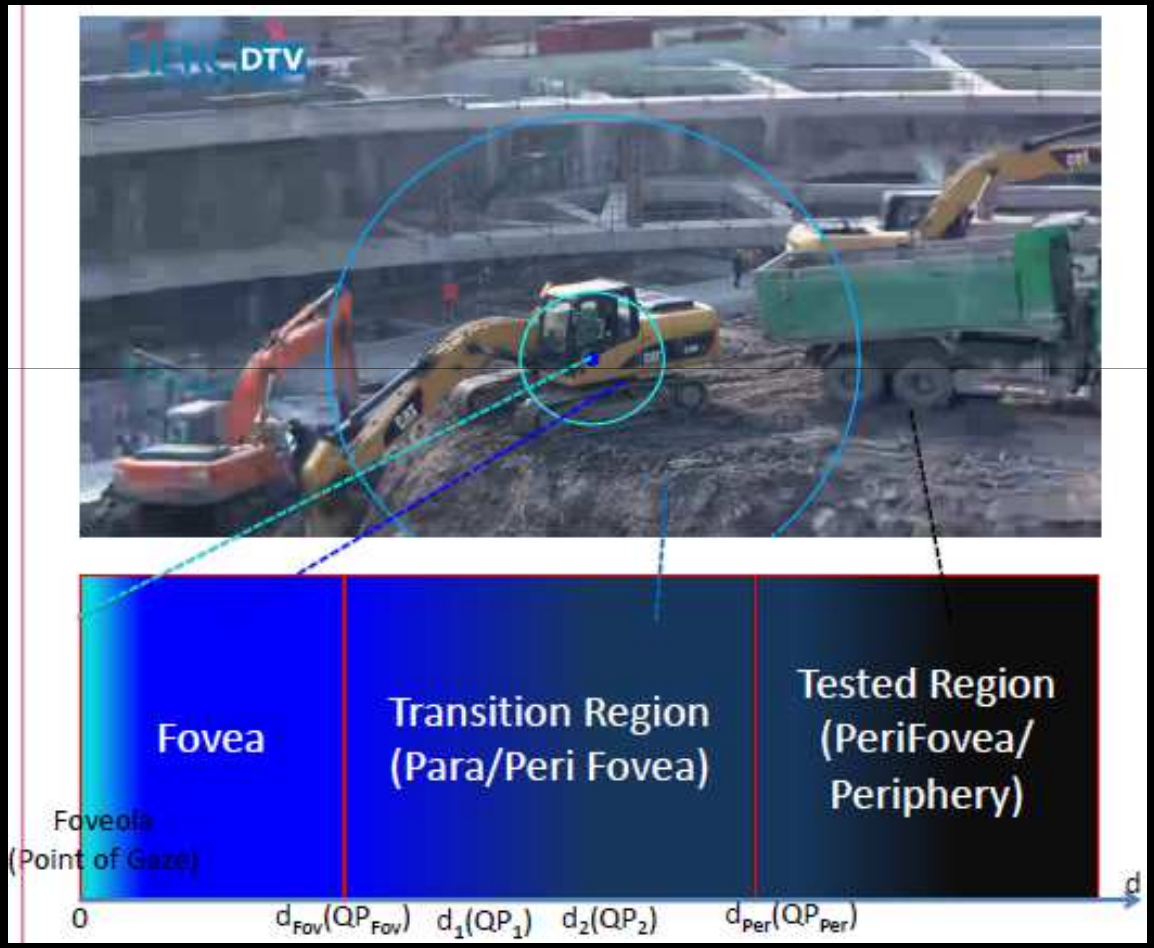
streaming interactif ?



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.

Expérience

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



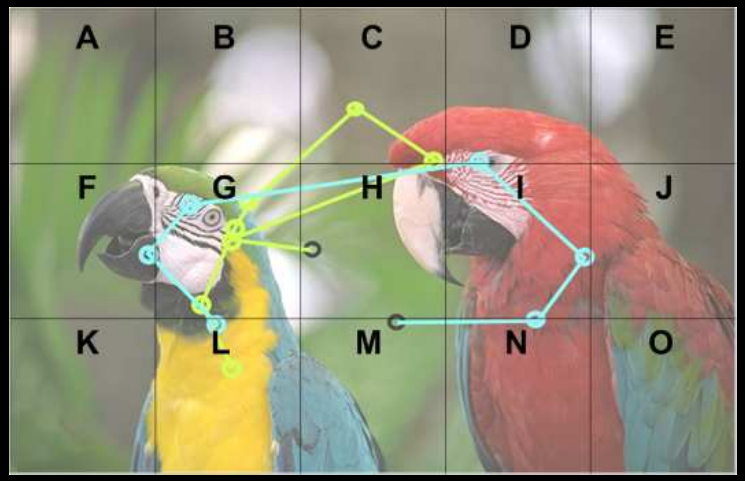
Y. Rai , M. Barkowsky & P. Le Callet « Role of peripheral Spatio-Temporal distortions indisrupting natural attention deploymentzation » HVEI'16 (best student paper)

Comment comparer 2 scan paths?

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

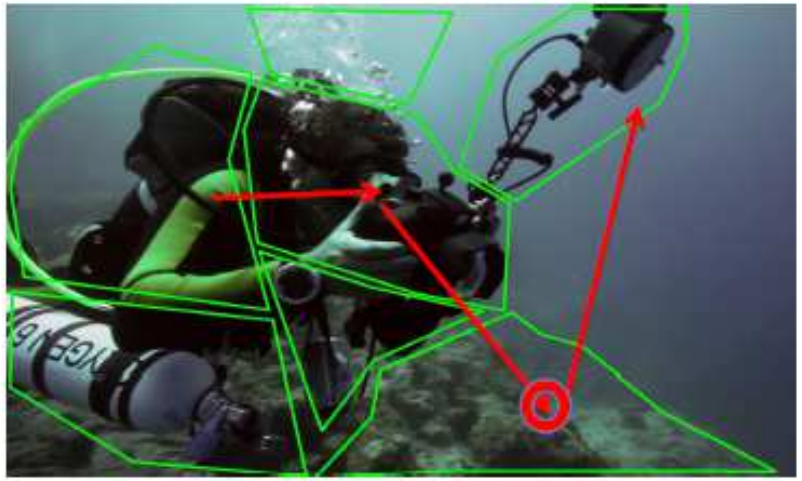
Nombre d'opération minimum pour transformer un chaine de carectère en une autre

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.

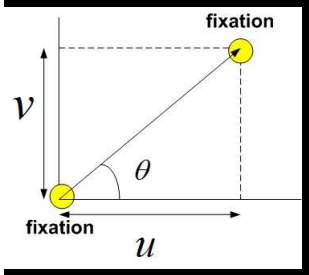
Ref \ Test	ROI	BG
ROI	26	10
BG	11	1

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?

Méthode basées vecteur



ScanPath
A &B

For each ScanPath: Simplification (Merging process)

- Merging small consecutive saccadic vectors
- Merging consecutive vectors having similar directions

Temporal Alignment

- Similarity Matrix M
- Adjacency Matrix M
- Goal: find the shortest path

Scanpath comparison

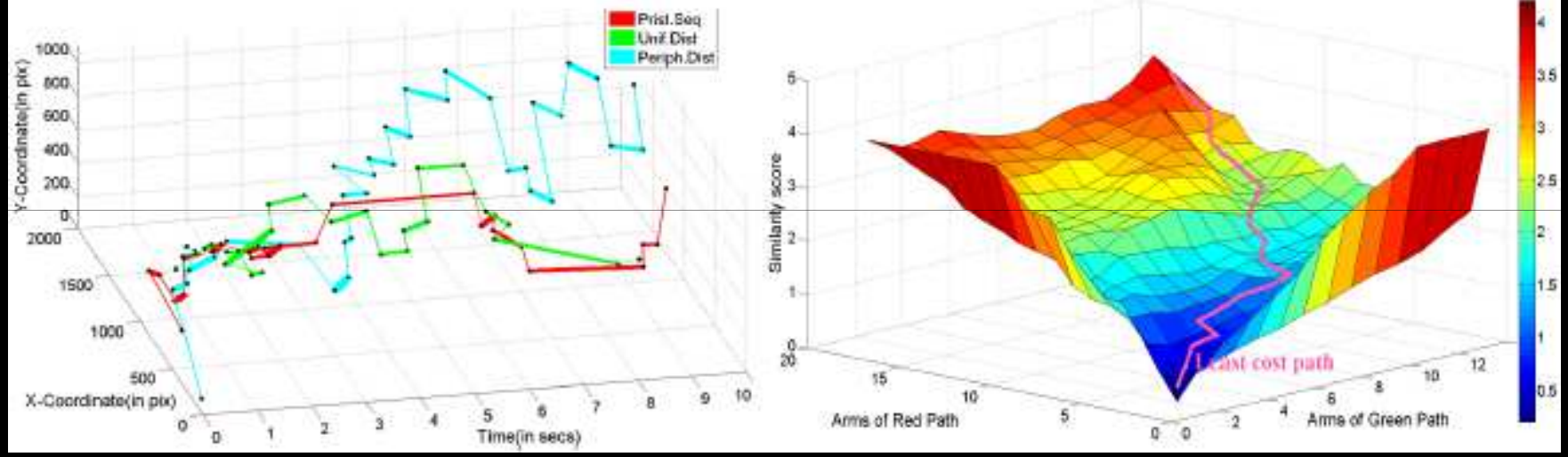
- Difference in shape (vector difference)
- Difference of saccade amplitude
- Spatial localization difference
- Direction difference
- Duration difference

preserves:
shape of the scanpath;
length of the scanpath (almost);
direction of the scanpath saccades;
position of fixations;
duration of fixations.

Advantages:
+ No pre-defined AOIs
+ Alignment of scanpaths (based on their shapes or on other dimensions)
Drawbacks
Eye movements such as smooth pursuit are not handled
It compares only two scanpaths

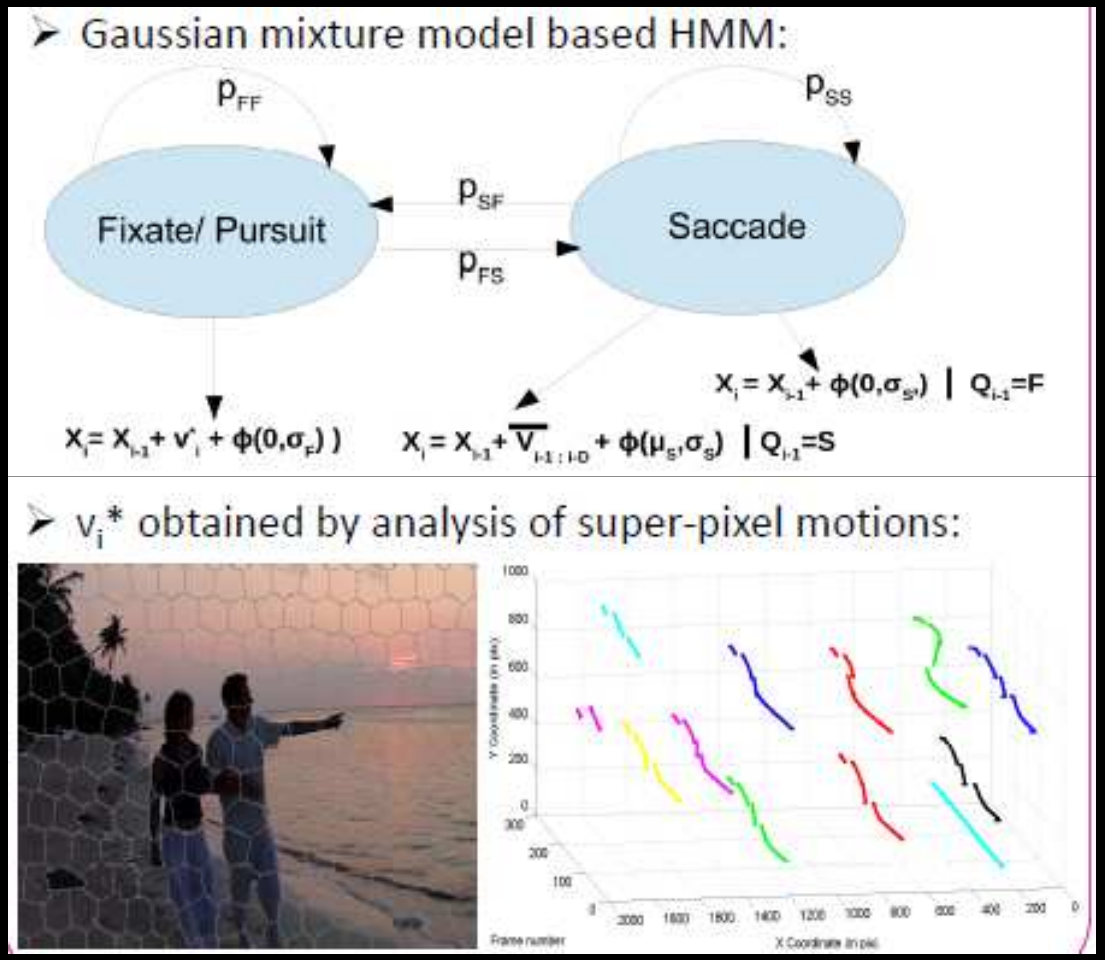
Comment comparer 2 scan paths?

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



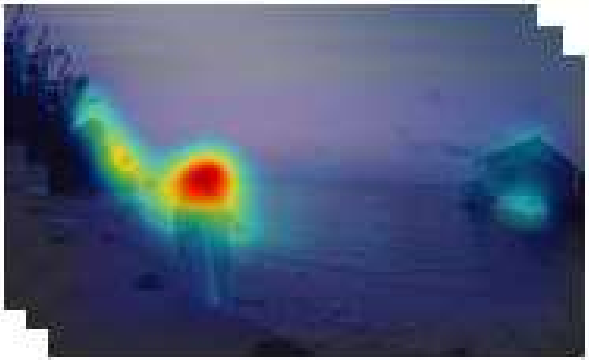
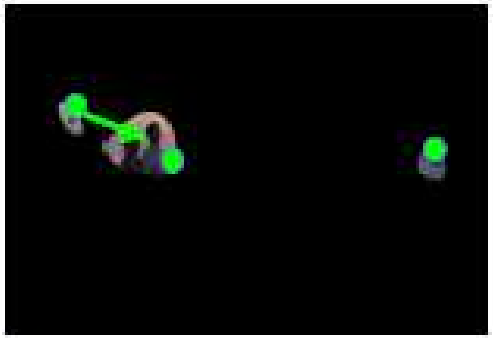
Y. Rai , M. Barkowsky & P. Le Callet « Role of peripheral Spatio-Temporal distortions indisrupting natural attention deploymentzation » HVEI'16 (best student paper)

Comment comparer 2 scan paths?



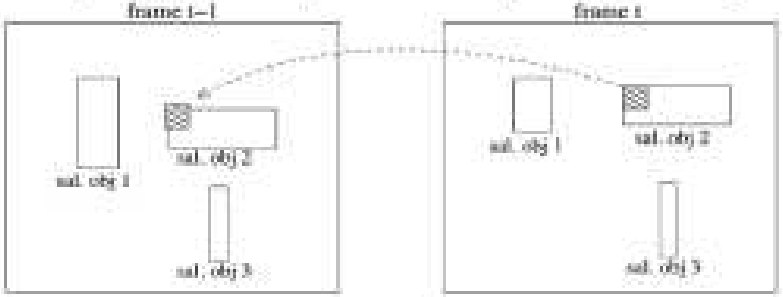
Y. Rai , G. Cheung & P. Le Callet « Role of HEVC Coding artifacts on gaze prediction in interactive video streaming systems» ICIP 16

Un espoir pour le streaming interactif

- Determine the state and transition probabilities for all saliency algorithms considered [Bruce et al, Li et al, Cheng et al, Riche et al, Judd et al, Harel et al]

$$p_{t,1} + p_{t,2} + p_{t,3} \dots + p_{t,n} + p_{t,s} = 1$$
$$p_{t-1,1} + p_{t-1,2} + p_{t-1,3} \dots + p_{t-1,n} + p_{t-1,s} = 1$$
$$p_{FF} + p_{FS} = 1$$
$$p_{SS} + p_{SF} = 1$$
$$p_{t,1} = p_{t-1,1} \cdot p_{FF} + p_{t-1,s} \cdot p_{SF}$$
$$p_{t,S} = \sum_{i=1}^{n} p_{t-1,i} \cdot p_{FS} + p_{t-1,s} \cdot p_{ss}$$



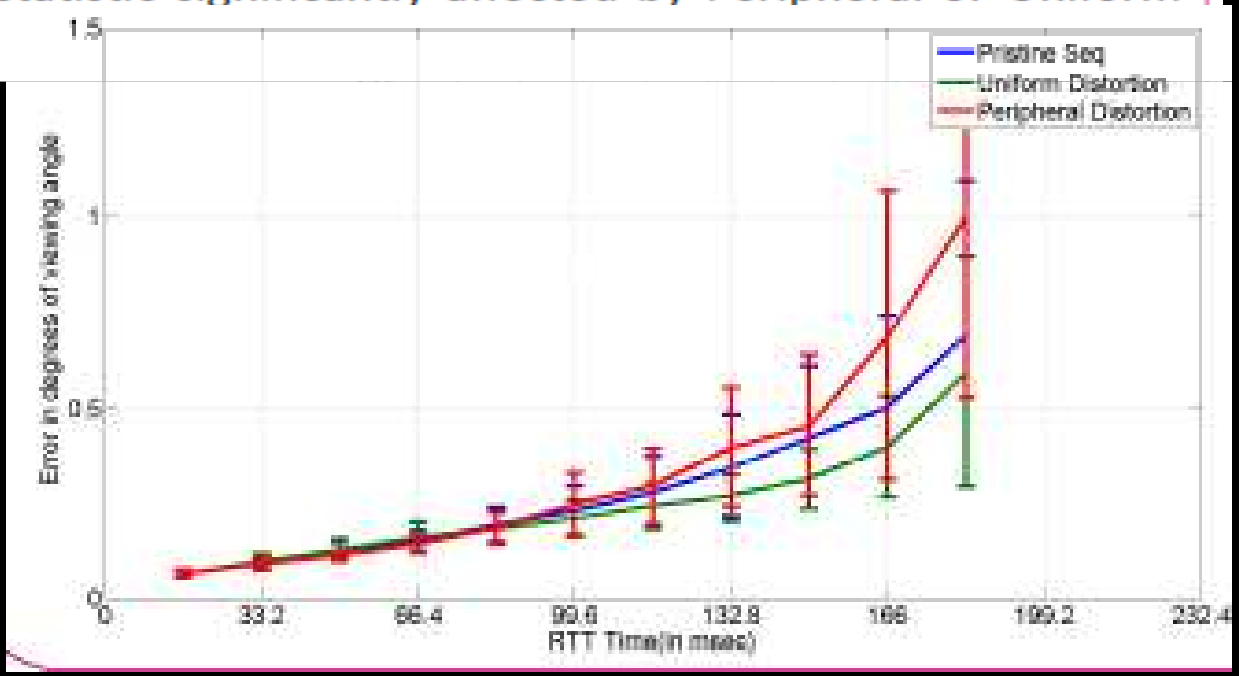
- Considering the Transition & Steady state probabilities, we choose the saliency algorithm producing results closest to subjective data (Harel et al : GBVS)

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

Un espoir pour le streaming interactif

Can the models predict gaze accurately?

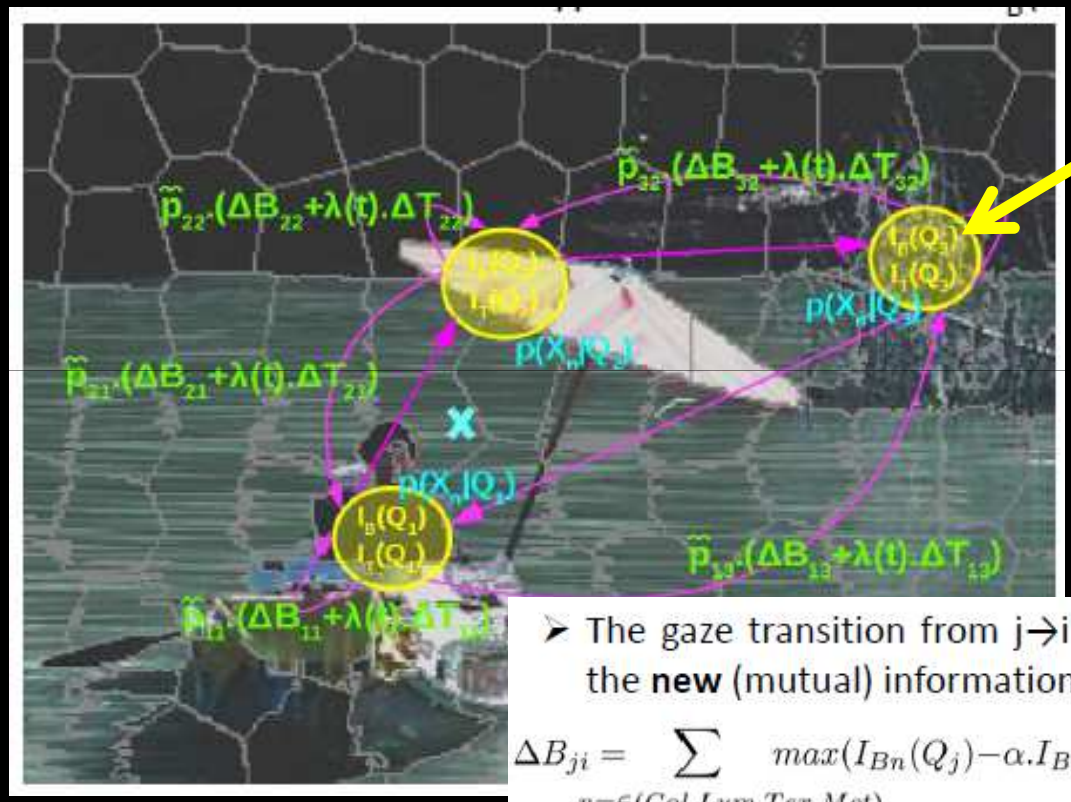
- Prediction error increases with time in case of peripheral distortions: more so if the sequence contains a lot of temporal activity.
- Prediction error can be restricted to a maximum of about 1.5 degrees of viewing angle within 200ms.
- Gaze prediction is not statistic significantly affected by Peripheral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

Vers un modèle de mesure de la « disruptiveness » 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

Attention overt vs covert

Nouvelle génération de modèles d’attention visuelle ?

(pour de nouvelles applications)

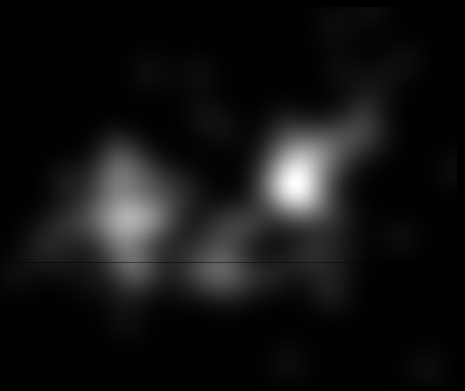
Genèse du projet : un manifeste scientifique

25 ans de recherche en modélisation d'attention visuelle

Input Image

Model

Saliency map

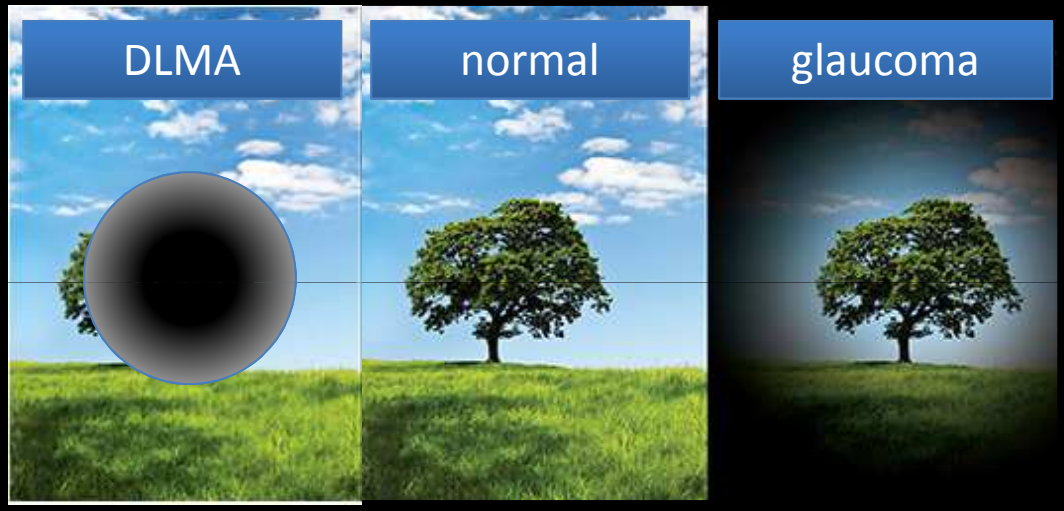


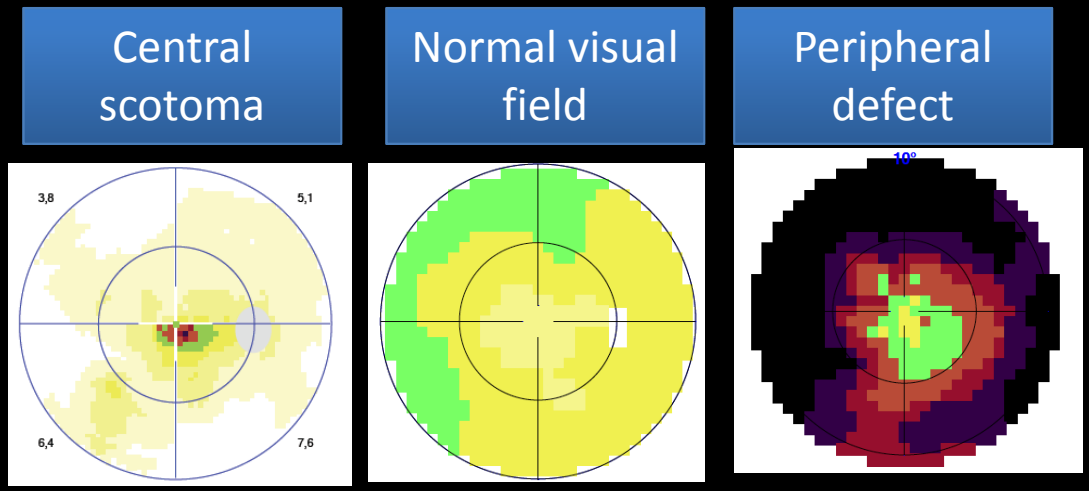
Les modèles actuels sont des modèles fovéaux

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

Notre approche (Projet VAM2020)

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





Attention Visuelle dans les Applications Multimédias

Pr. Patrick Le Callet

www.irccyn.ec-nantes.fr\~lecallet

POLYTECH NANTES

IRCCyN

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES