UNIVERSITÉ DE NANTES

Polytech'Nantes
École d'ingénieurs de l'université de Nantes

IRCCYN
UMR CNRS 6597

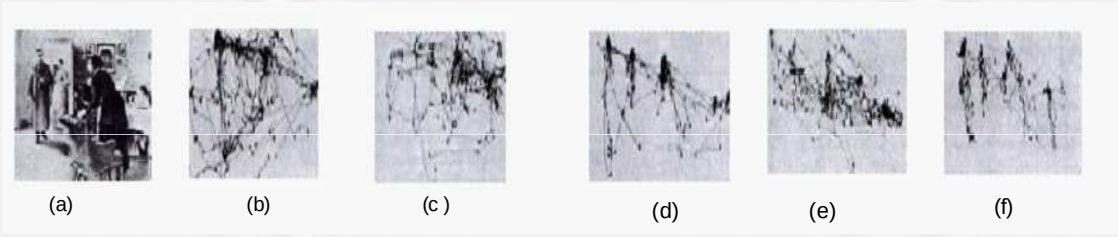
Visual attention modeling and applications

Patrick LE CALLET

Visual attention

Top-down
=> Task driven

Bottom-up
=> Signal driven



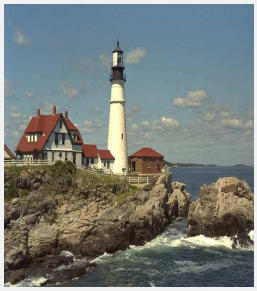
(a) Painting « *Unexpected Return* », I. Repin
(b) No task
(c) Social level of people
(d) People age
(e) People activity
(f) Memorise clothes

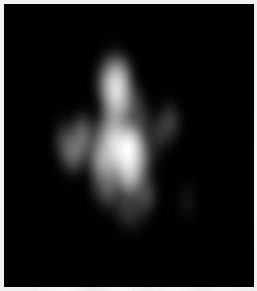
Experiment from A. Yarbus [Yarbus 67]

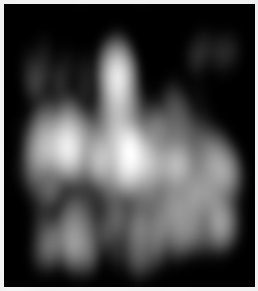
issue : modeling bottom-up process of visual attention

Detect areas of image or video that are salient

visual attention: psychophysics experiments







databases

http://www.irccyn.ec-nantes.fr/ => plateformes

Still images databases :

- Free Task :
- IRCCyN/IVC Eyetracker 2006 05

An eyetracker on images database. This database is composed by eyetracker data associated to 27 original still images.

- IRCCyN/IVC Eyetracker Berkeley Database

A 481x321 pixels images database. This database is composed by eyetracker data associated to 80 original still images.

- Quality Task :

Videos databases :

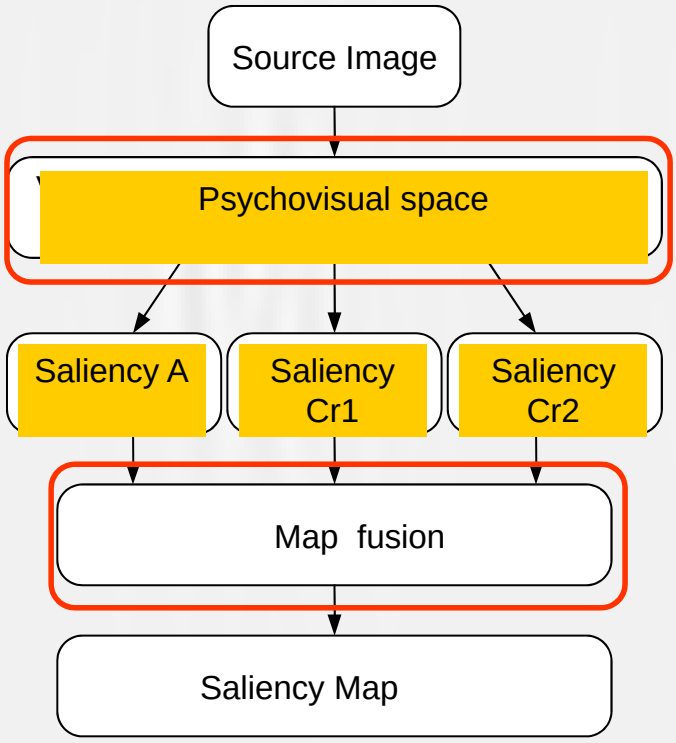
- Free Task :
- IRCCyN/IVC Eyetracker SD 2008 11 Database

A SD videos database. This database is composed by 51 distorted videos by h264 coding. The eyetracker data are available for this database.

- Quality Task :
- IRCCyN/IVC Eyetracker SD 2009 12 Database

A SD videos database. This database is composed by 20 original sequences and 80 distorted videos by h264 coding with transmission errors. The h264 streams and the eyetracker data are available for this database.

Design and architecture of usual bottom-up models



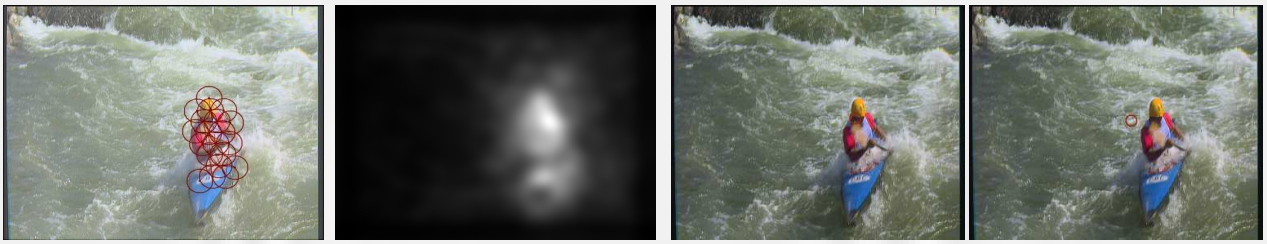
Heterogeneous extraction

Provide a coherent approach for extraction

Non obvious mapping of different features type

Homogenous data

Model performances

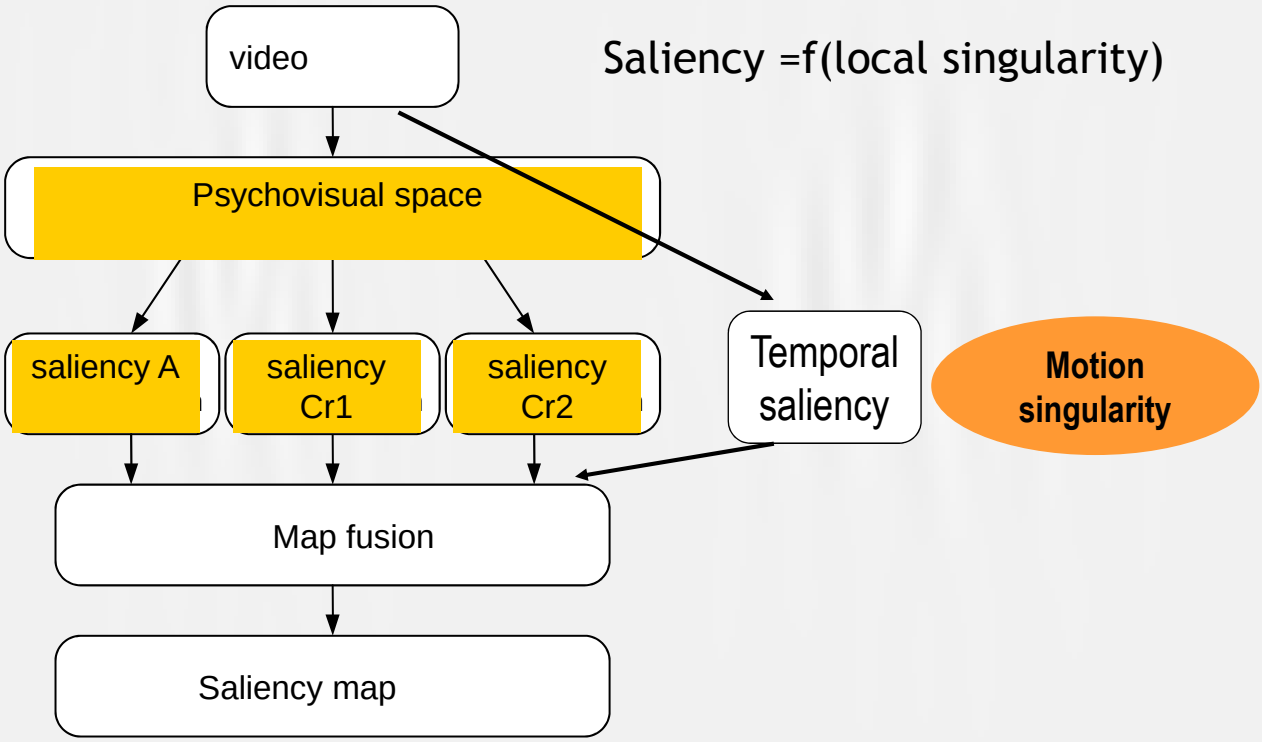


Model dispersion < natural inter-observers dipersion

Improved SoA (Itti's model)

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

Extension to temporal dimension



Model performances

Outperform State of the Art
Used in Thomson products
-encoders
- content repurposing and reframing

O. Le Meur, P. Le Callet et D. Barba
A spatio-temporal model of bottom-up visual selective attention: description and assessment
Vision Research, Vol 47, issue 19, pp 2483-2498 DOI info: 10.1016 /j.visres.2007.06.015

Applications: coding

direct selective H.264 encoding :

Principle

Selection prior:

saliency map, RoI

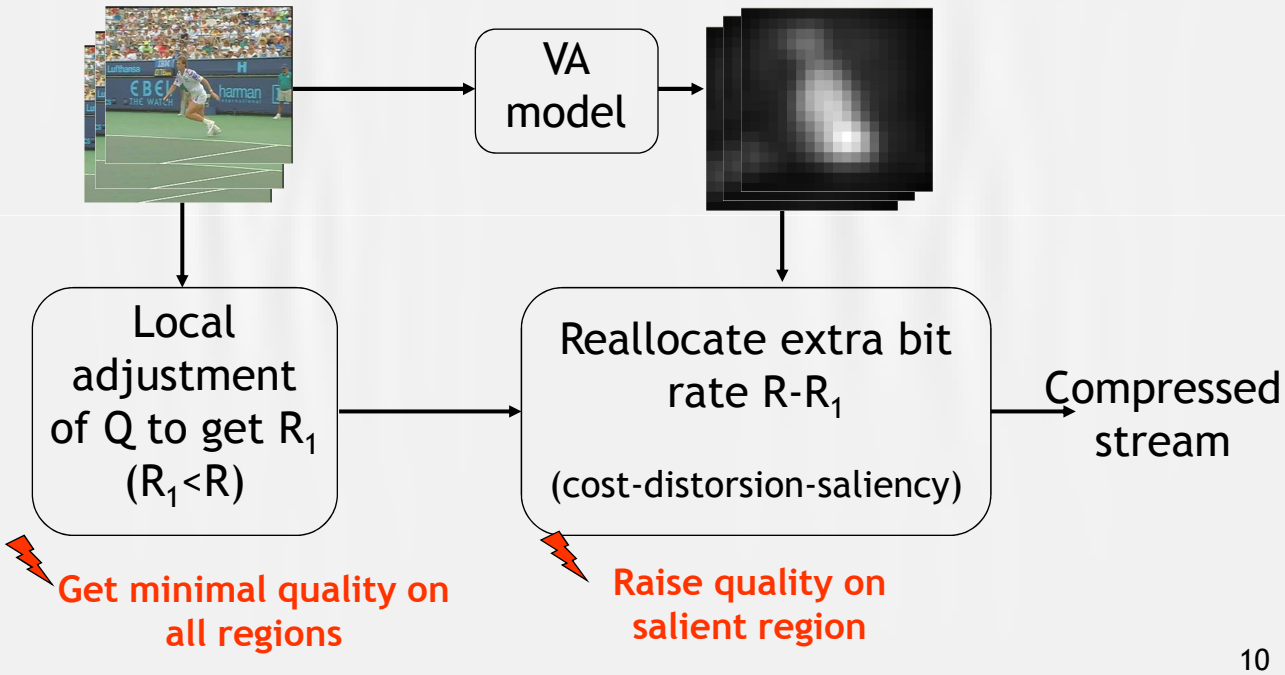
Compression prior:

modification of quantization to favorize salient region

⇔ take rate in non salient regions and reallocate it in salient regions

Applications: coding

Method: 2 pass coding (target bit rate= R)



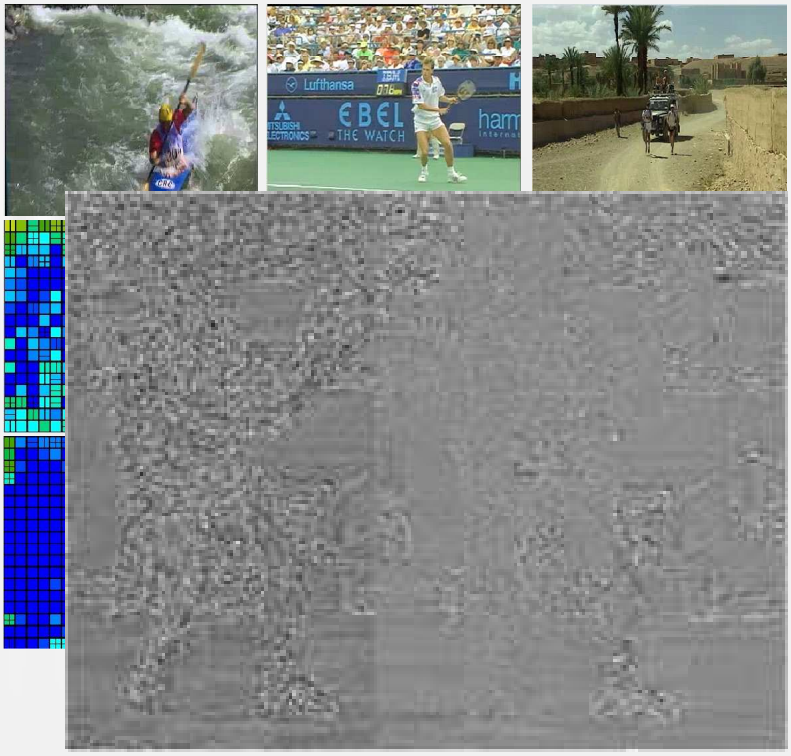
Applications: coding

Results: macroblocks bit rate allocation

Source

Usual Coding

Proposed coding

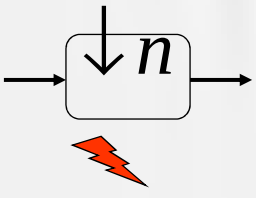


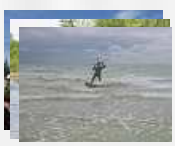
Applications: reframing

Adaptation to terminal resolution

Thumbnails mode

Usual approach:

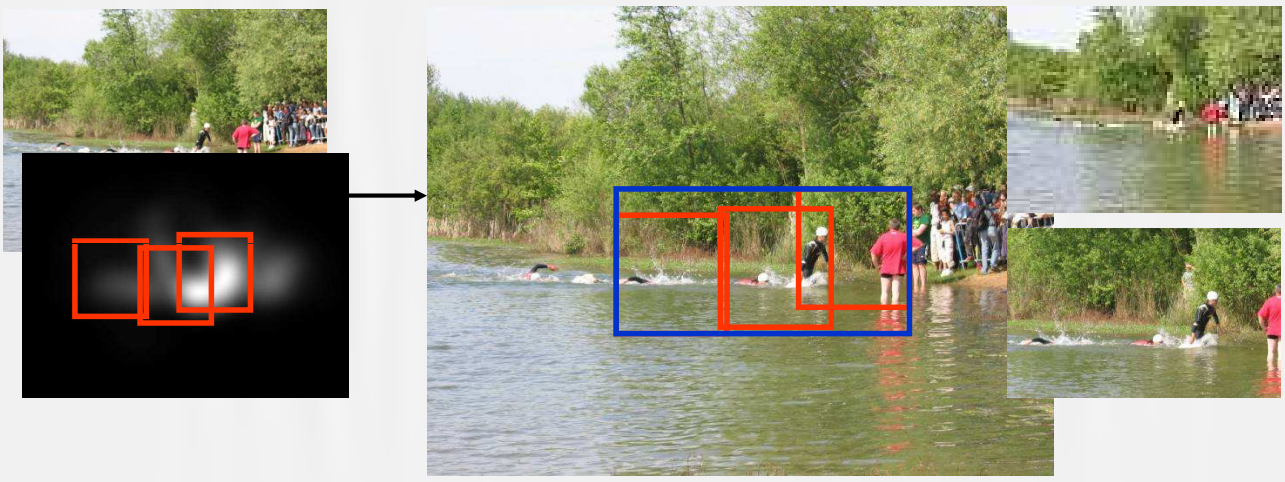




Poor usability
(lack of details)
aliasing

Applications: reframing

Proposed approach :



- A coherent computational Approach to model the bottom-up visual attention. , O. Le Meur, P. Le Callet, D. Barba et D. Thoreau. IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI) 2006,
- Efficient Saliency-Based Repurposing Method, O. Le Meur, P. Le Callet, D. Barba et X. Castellan, IEEE ICIP 06,

Applications: Quality assessment and visual attention

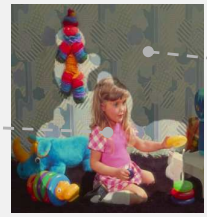
Visual attention in image quality assessment ?

Basic idea :

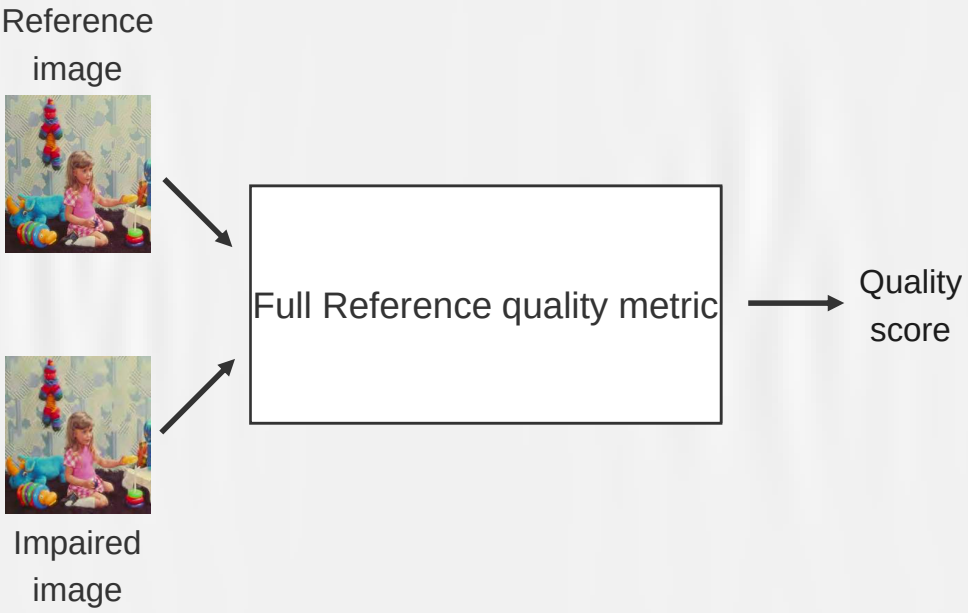
"A distortion that appears on region of interest is more annoying than a distortion appearing on an inconspicuous area"

Simple application to image quality metric :

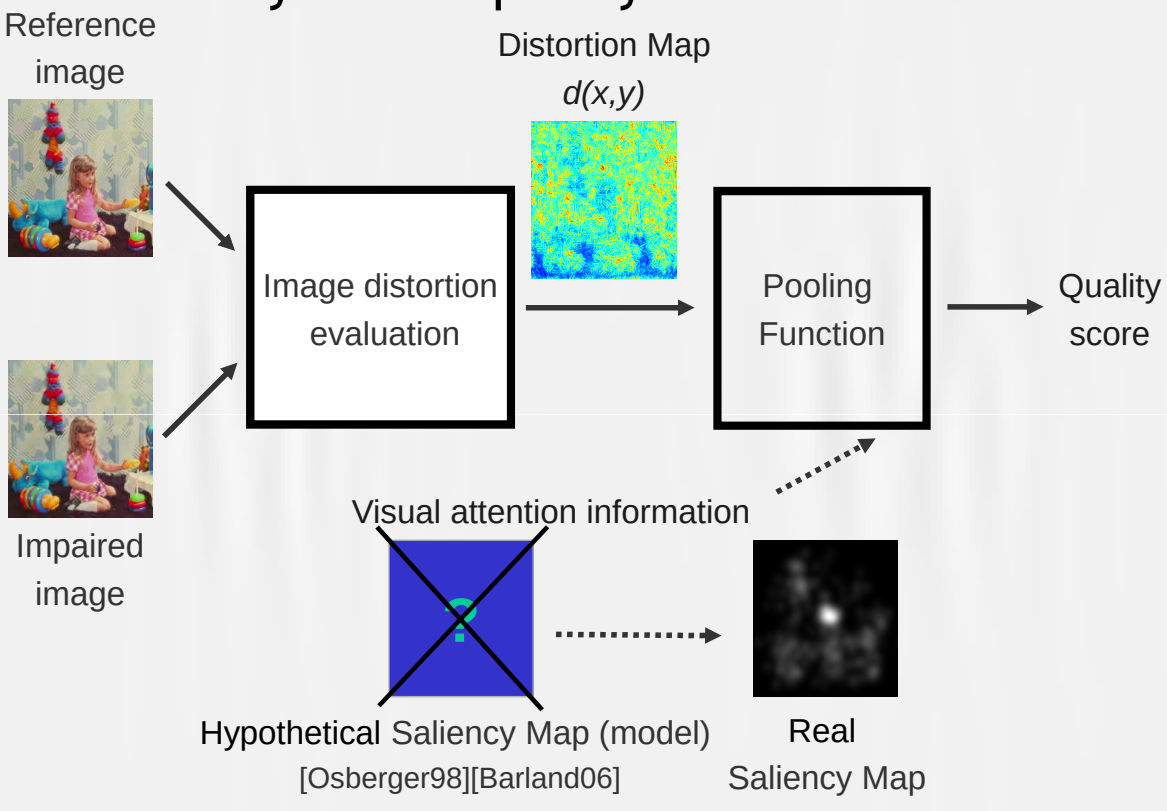
Give more weight to the distortions appearing on the saliency areas



Quality assessment and visual attention



Saliency-based quality metrics



The ground truth

Eye gaze tracking experiments during a quality assessment campaign

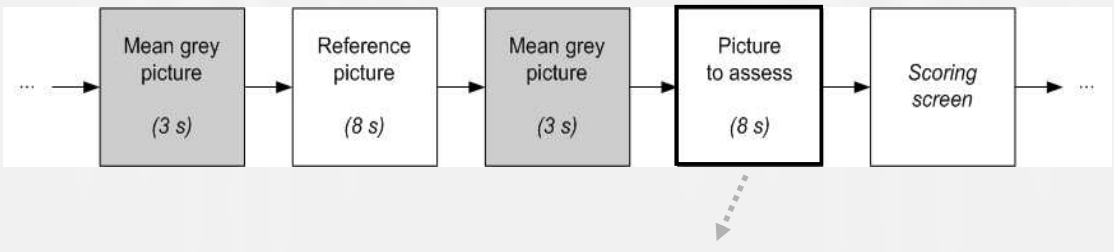
Collected data :

- *Mean observer score (MOS, from observers' scores)*
- *Real saliency information (from eye gaze position)*

The ground truth : *Mean Observer Score*

Picture quality assessment campaign :

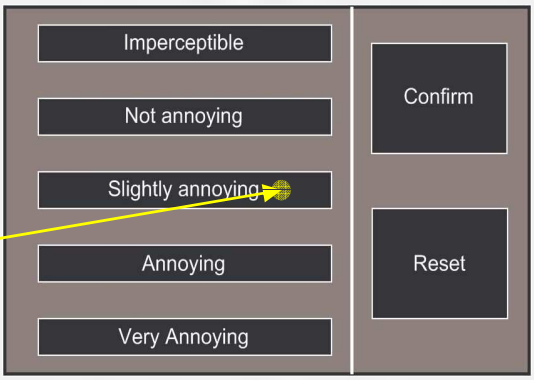
- 20 observers
- 130 pictures *(10 unimpaired references,120 impaired versions)*
- Visualization distance 4H
- DSIS protocol *(Double Stimulus Impairment Scale)*



The ground truth : *Mean Observer Score*

Degradation category rating (5 scores) :

- *Imperceptible*
- *Not annoying*
- *Slightly annoying*
- *Annoying*
- *Very annoying*

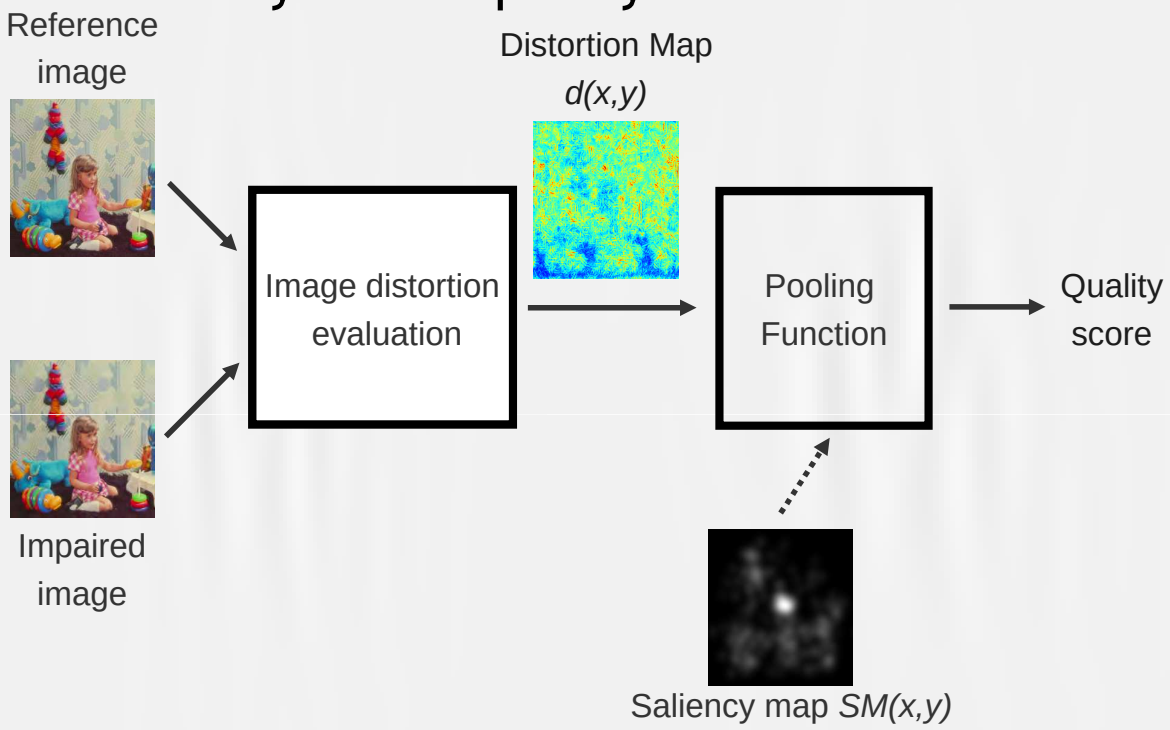


How to rate ?

- *Scoring screen*
- *selection and confirmation based on eye gaze position*

Score is *Not annoying*

Saliency-based quality metrics



Saliency-based quality metrics

Pondération spatiale (image par image) des distorsions : $Distorsions(t,x,y)$

$$D_t^S = \left(\frac{\sum_{k=1}^{K} \sum_{l=1}^{L} w_i(t,x,y) \cdot \left(Distorsions(t,x,y) \right)^{\beta_s}}{\sum_{k=1}^{K} \sum_{l=1}^{L} w_i(t,x,y)} \right)^{\frac{1}{\beta_s}} \quad (2)$$

	Images	
	PSNR	SSIM
Sans saillance	0.742	0.827
Saillance réelle	0.825	0.825

Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric, A. Ninassi, O. Le Meur, P. Le Callet et D. Barba, IEEE ICIP 07

Overt visual attention for free-viewing and quality assessment tasks. Impact of the regions of interest on a video quality metric O. Le Meur, A. Ninassi, P. Le Callet and D. Barba, Elsevier, Signal Processing: Image Communication. 2010

Do video coding impairments disturb the visual attention deployment? O. Le Meur, A. Ninassi, P. Le Callet and D. Barba, Elsevier, Signal Processing: Image Communication 2010

Video over IP

- Considerable growth of video traffic over IP in recent years (*e.g.*, TVoIP, VoD, videoconferencing).
- Internet originally not designed for multimedia applications => packet loss, large delay, jitter…

➡ Fluctuation in quality

Erroneous macroblocks

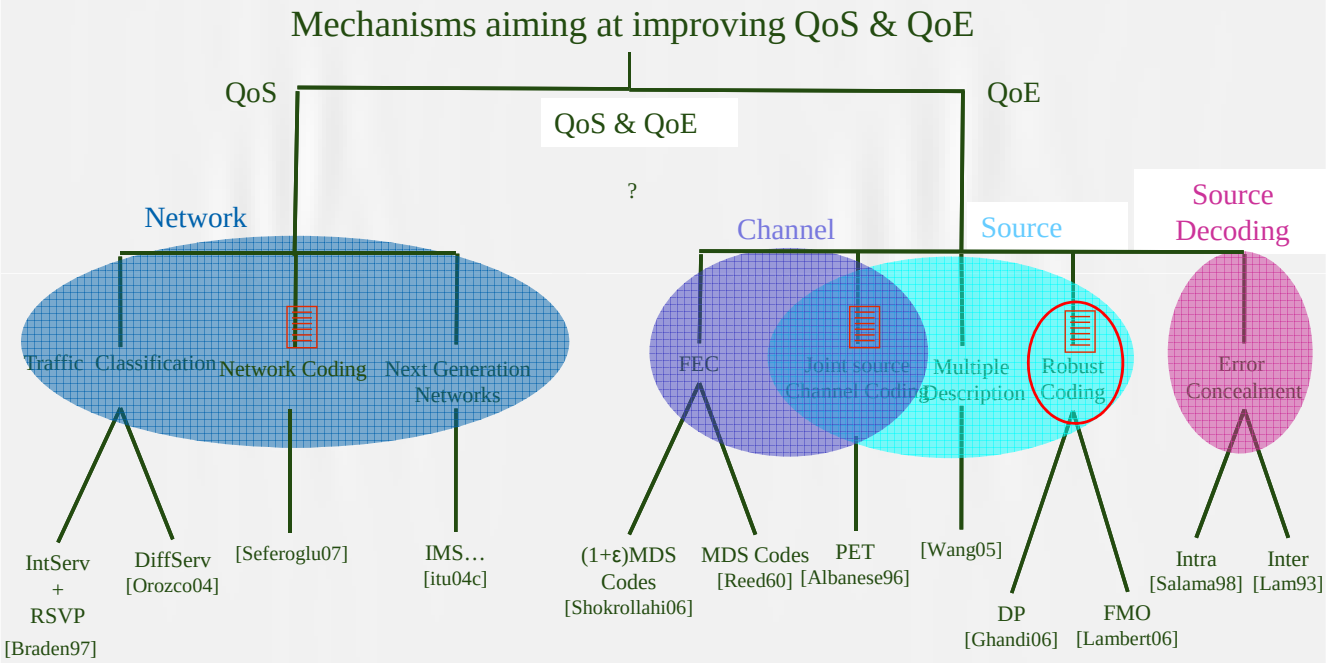
Jerkiness

Image freeze

QoE and QoS

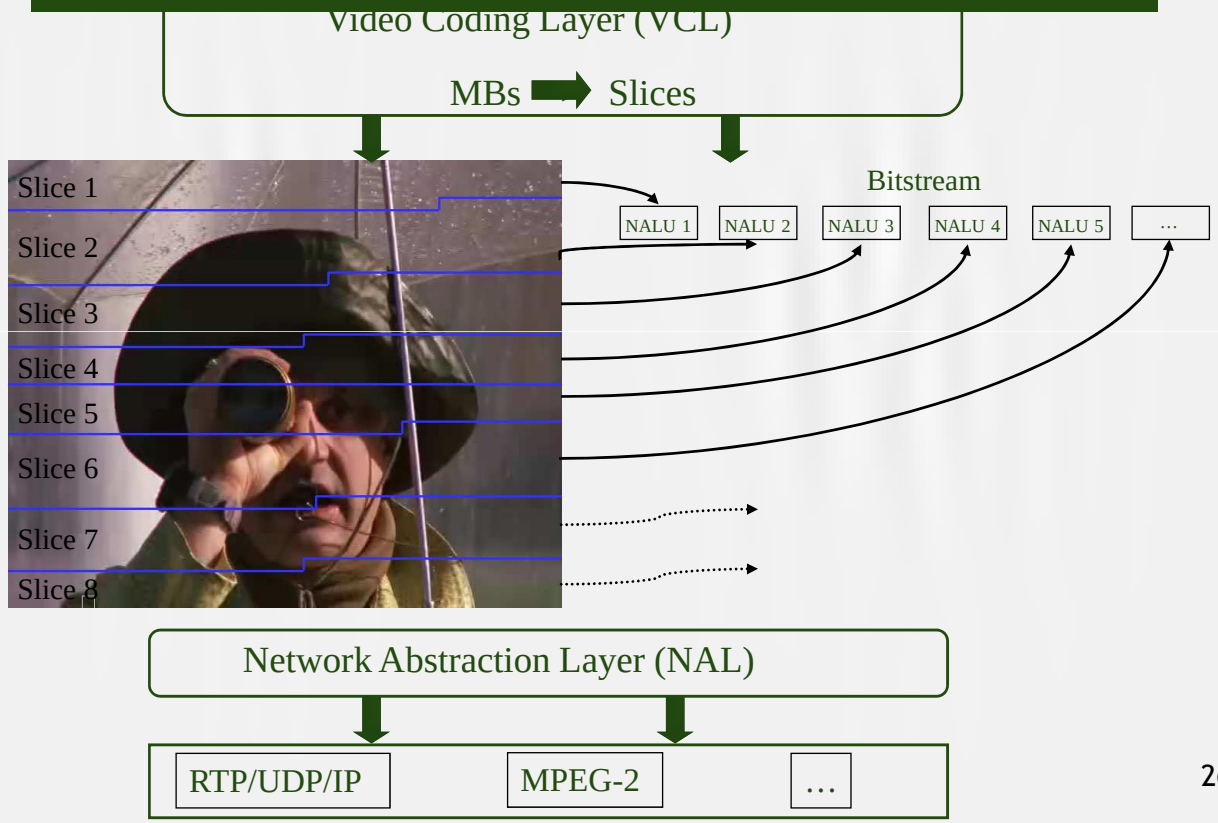
- Quality of Experience: measure of end-user satisfaction w.r.t. the multimedia service.
- Quality of Service: objective measure of the network's performance (loss, delay and jitter).

Sustaining an Acceptable Quality Level



Importance of Protection

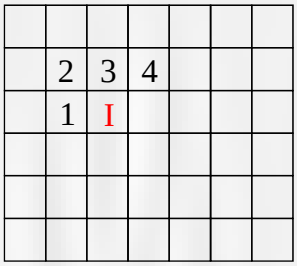
H.264/AVC Bitstream



H.264/AVC Coding

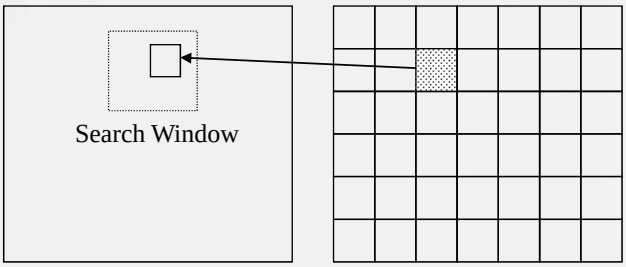
– Block-based video coding standard.

– Spatial prediction: I pictures.



Intra-Coded Picture

– Temporal prediction: B and P pictures.



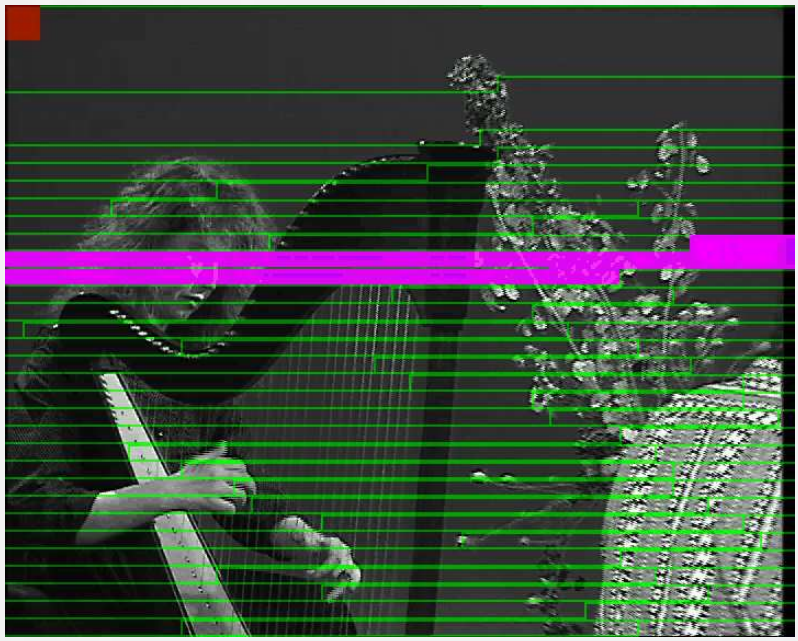
Reference Picture B or P Picture

Loss Simulation Example (1)

I packets lost

B packets lost

Loss Simulation Example (2)



Flexible Macroblock Ordering

– Error resilience tool available in H.264/AVC Extended Profile.

– Flexible mapping of macroblocks to slices.

Type 6

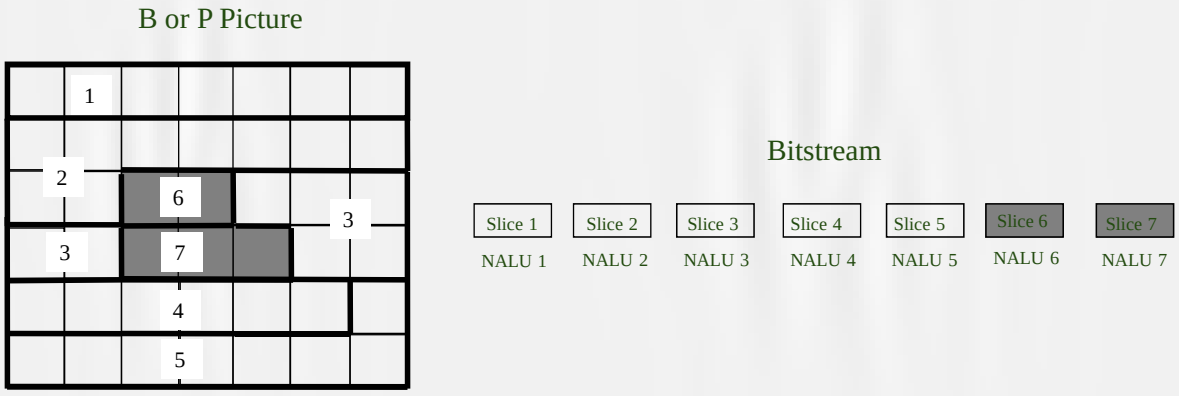
0	1	2	…			
						n

Specify mapping on a per-macroblock basis:

MB 0 --> slice 1
MB 1 --> slice 1
…
MB n --> slice x

FMO-based RoI Intra-Prediction

– Group macroblocks in slices according to RoI criterion.



– FMO gives flexibility for later processing (*e.g.*, channel coding).

– Expected overhead due to extensive intra-coding.

From Saliency Maps to Regions of Interest

– The encoder needs macroblock-level information.

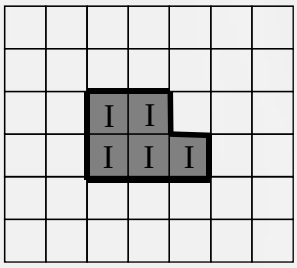
– A pixel belongs to the RoI if its saliency value is higher than a threshold.

– Threshold set empirically such as not to have a saliency map covering more than 20% of the image.

Constrained Intra-Prediction

– 2 methods can achieve secure intra-prediction:

1. **Constrained Intra-Prediction:** Force all intra-predictions to use residual data and decoded samples from I macroblock types only.
2. **FMO slicing:** Forbid intra-prediction across slices.



B or P Picture

Region-of-Interest Intra Prediction for H.264/AVC Error Resilience, F. Boulos, W. Chen, B. Parrein,P. Le Callet, *IEEE ICIP* 2009

A New H.264/AVC Error Resilience Model Based on Regions of Interest
F. Boulos, W. Chen, B. Parrein, P. Le Callet, *Packet Video 2009*

Performance Without Loss

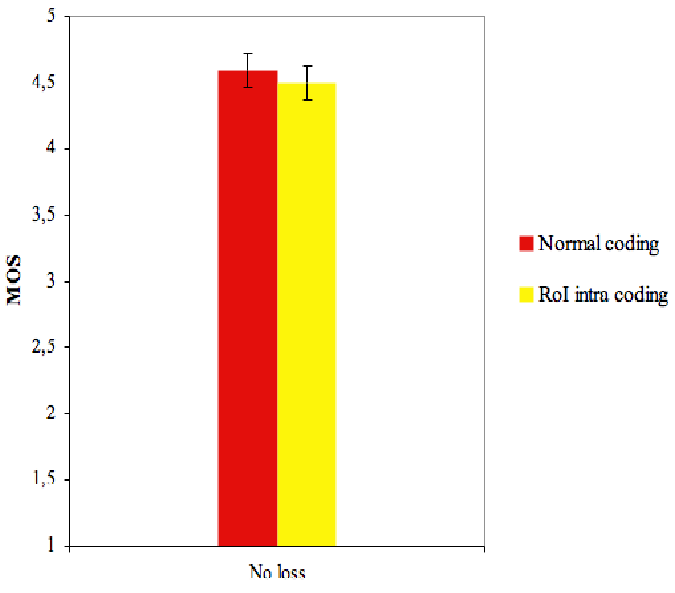
Excellent

Good

Fair

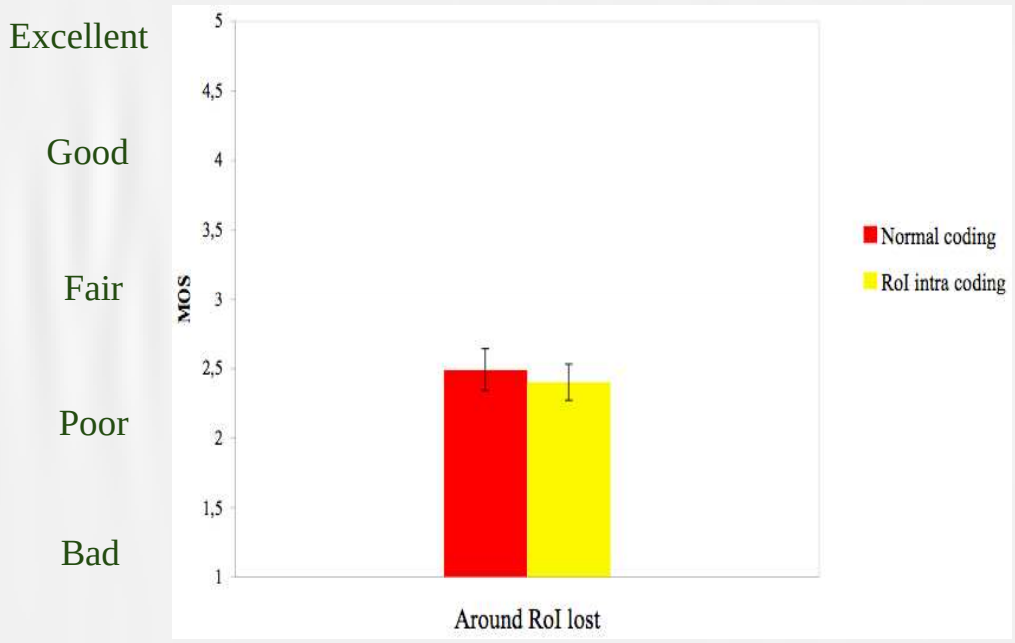
Poor

Bad



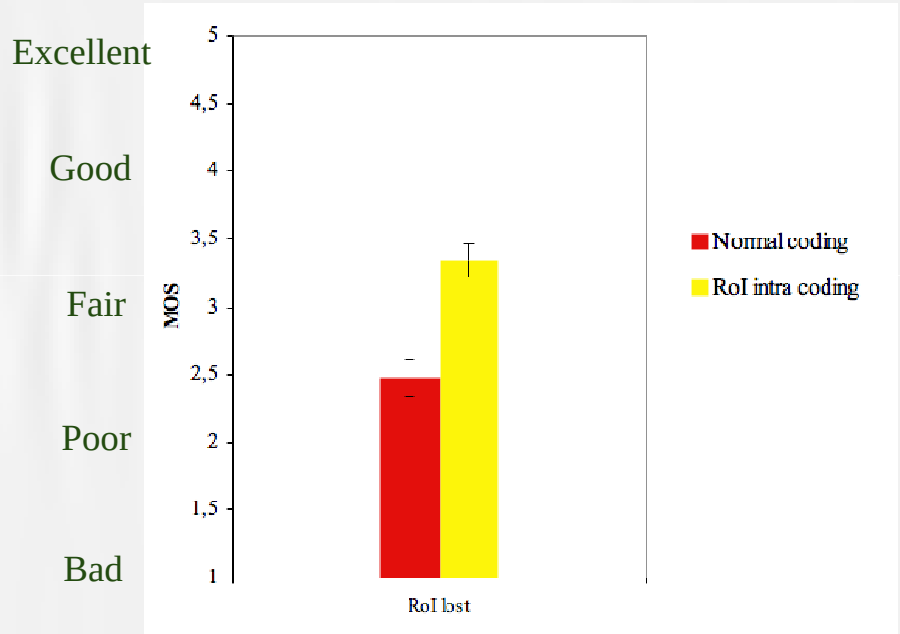
Decrease in coding efficiency insignificant perceptually at used bitrates (~ 4Mbit/s).

Performance in Presence of non-RoI Loss



Same performance of proposed error resilient coding and classical coding.

Performance in Presence of RoI Loss



1 category difference when proposed error resilient coding is used compared to classical coding.