THOMSON

images & beyond

IEEE International Conference on Image Processing
***ICIP 2007*, 16-19 September 2007**

Session:MP-L6: Image Scanning, Display, Printing, Color and Multispectral Processing I
Topic:Image Scanning, Display and Printing: Image Quality Assessment

Does where you gaze on an image affect you perception of quality? Applying visual attention to image quality metric

A. Ninassi [1,2], O. Le Meur [1], P. Le Callet [2], D. Barba [2]

[1] THOMSON R&D, VCL, Cesson-Sevigné, France

[2] IRCCyN UMR 6597 CNRS, Nantes, France

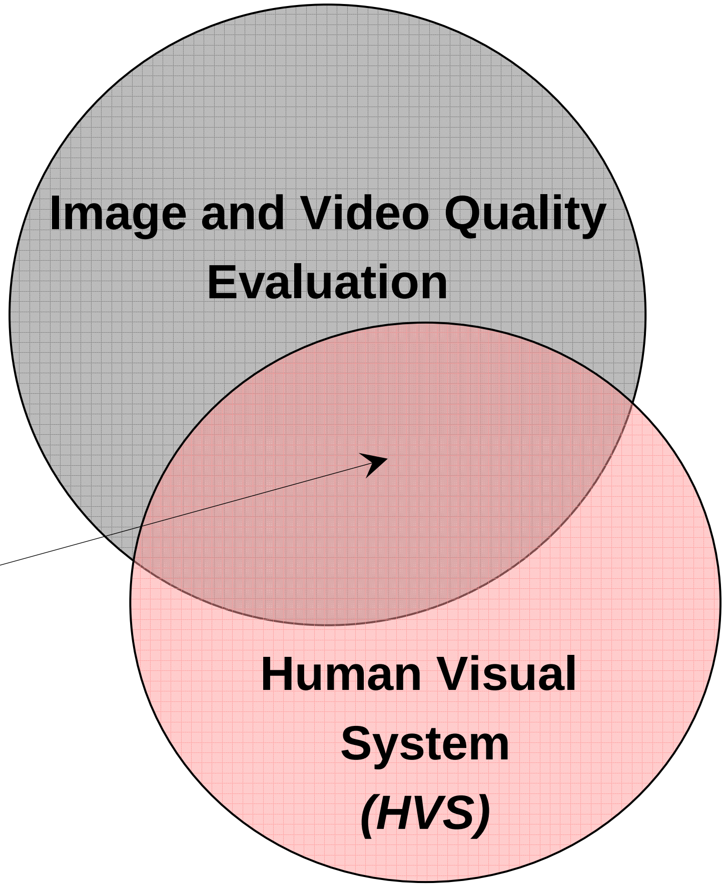
Introduction

“Most of the efficient quality metrics are based on the Human Visual System ”

An important component of the HVS

=

Visual attention



Outline

1. Image quality assessment and visual attention
2. The ground truth
3. Saliency-based quality metrics
4. Results
5. Conclusion

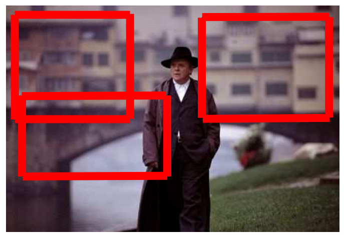
Quality assessment and visual attention 1/3

Visual attention ?

“ Visual attention allows us to select the relevant information in our environment, in connection or not with a particular task ”

<u>Two mechanisms are involved in visual attention control</u>

Bottom-up	**Top-down**
(involuntary attention)	(voluntary attention)
Salient elements of the visual field catch the attention	Our attention is guided by the task to accomplish
⇔	⇔
Mechanism based on signal	*Mechanism based on task*

THOMSON images & beyond

Quality assessment and visual attention 2/3

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 :

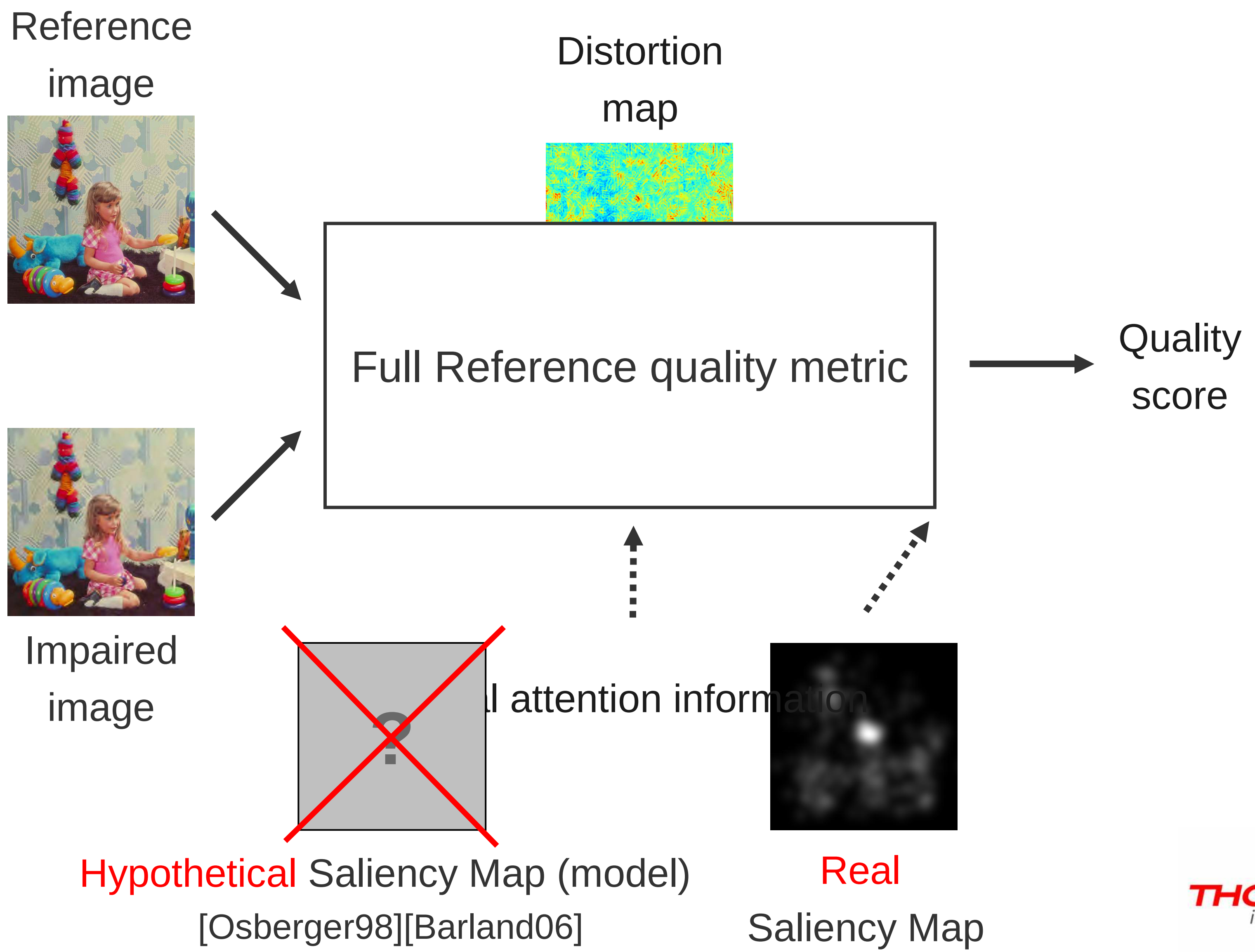
Give more weight to the distortions appearing on the saliency areas

Distortion weight

Distortion weight

THOMSON images & beyond

Quality assessment and visual attention 3/3



Outline

1. **Image quality assessment and visual attention**
2. The ground truth
3. Saliency-based quality metrics
4. Results
5. Conclusion

The ground truth *1/4*

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)*

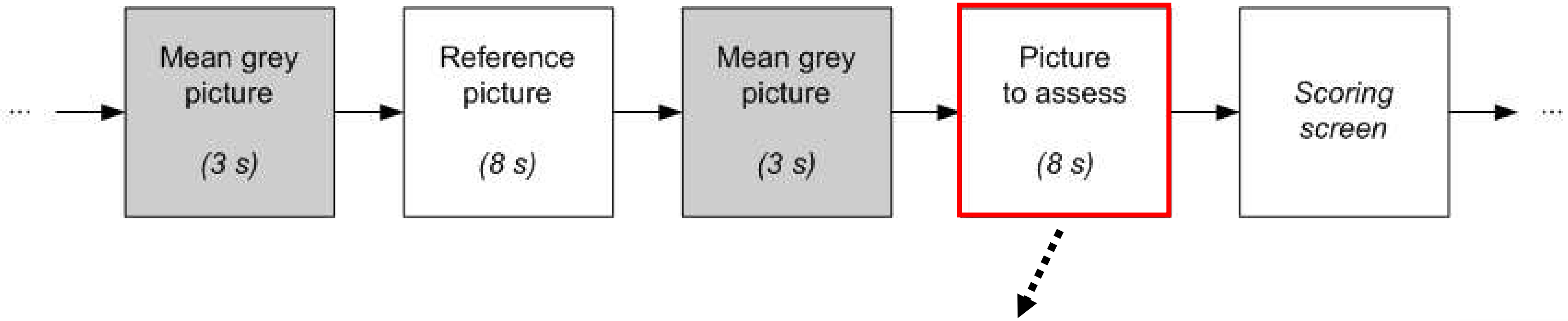
An ***eye-tracker*** is a device elaborated to track and to record the position where a human observer is looking at

THOMSON images & beyond

The ground truth : *Mean Observer Score* *2/4*

Picture quality assessment campaign :

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

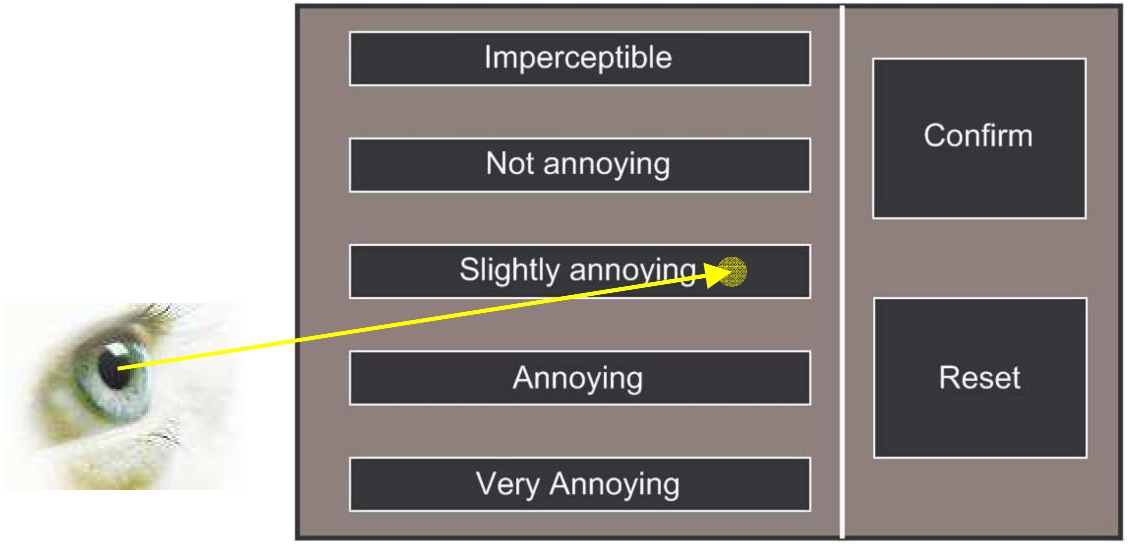


Real Saliency Map

The ground truth : *Mean Observer Score* 3/4

Degradation category rating (5 scores) :

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

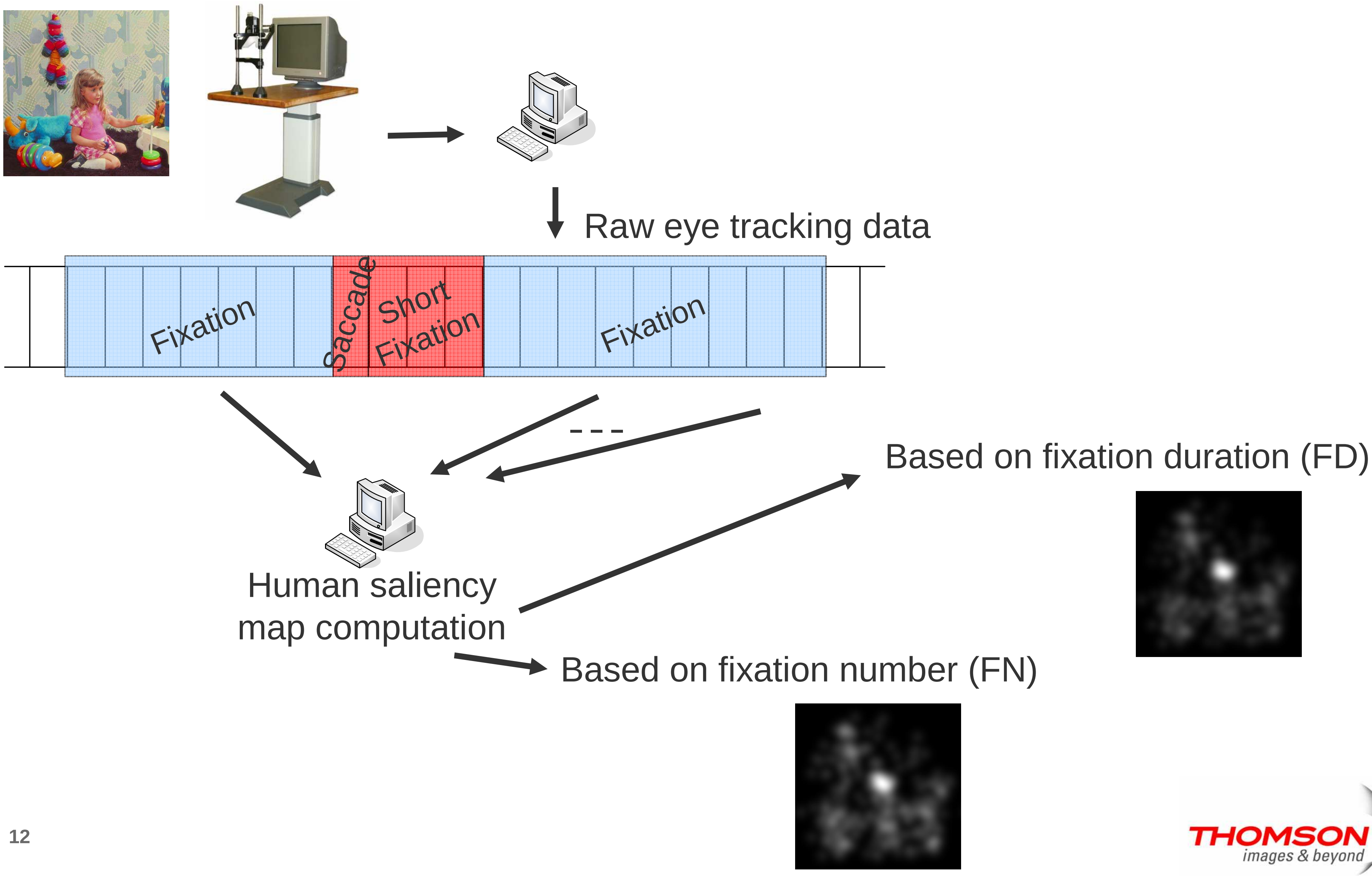


How to rate ?

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

Score is *Not annoying*

The ground truth : *Saliency Information* 4/4

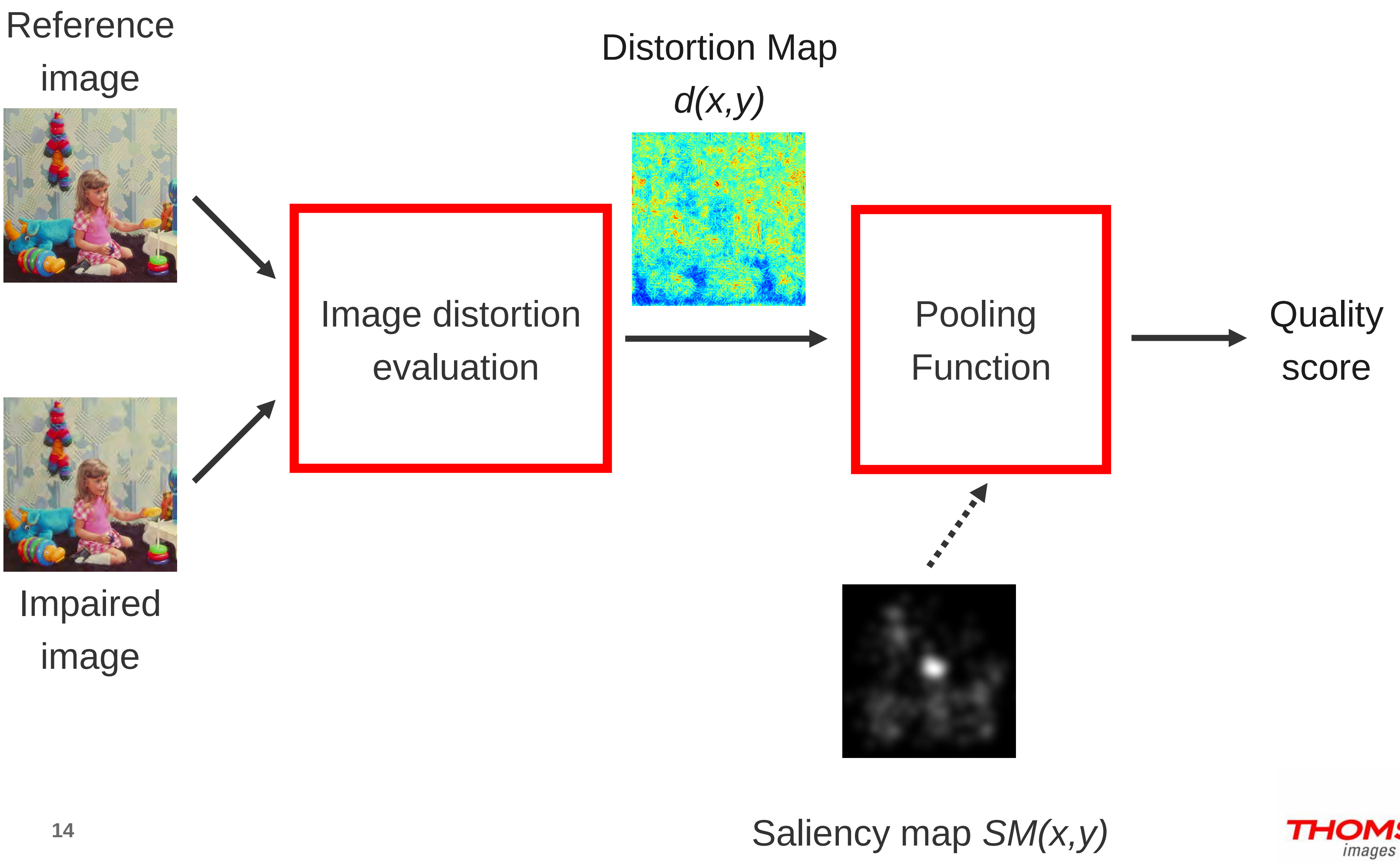


THOMSON images & beyond

Outline

1. Image quality assessment and visual attention
2. **The ground truth**
3. Saliency-based quality metrics
4. Results
5. Conclusion

Saliency-based quality metrics *1/2*



THOMSON images & beyond

Saliency-based quality metrics : *Pooling function* 2/2

Saliency-based spatial pooling :

(according to previous literature works [Osberger98][Barland06])

$$Q = \frac{\sum_{x=1}^{W} \sum_{y=1}^{H} w_i(x,y) \cdot d(x,y)}{\sum_{x=1}^{W} \sum_{y=1}^{H} w_i(x,y)}$$

Where Q *is the quality score,*
$d(x,y)$ *is the distortion map,*
$w_i(x,y)$ *is the weight function,*
W *and* H *are the width and the height of the picture*

4 simple weight functions :

$$w_1(x,y) = SM_n(x,y)$$
$$w_2(x,y) = 1 + SM_n(x,y)$$
$$w_3(x,y) = SM(x,y)$$
$$w_4(x,y) = 1 + SM(x,y)$$

Where $SM(x,y)$ *is the saliency map and* $SM_n(x,y)$ *is the normalized saliency map*

Outline

1. Image quality assessment and visual attention
2. The ground truth
3. **Saliency-based quality metrics**
4. Results
5. Conclusion

Results : *Predicted MOS vs. MOS* 1/5

Mean observer behavior
Comparison between MOS and MOSp (predicted MOS) :

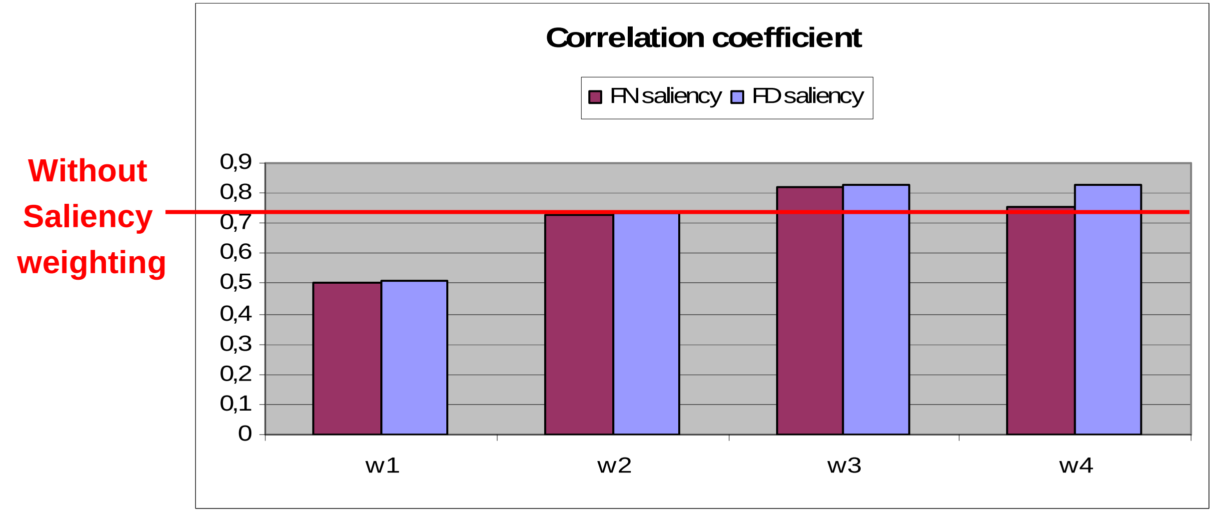
- **Of simple quality metrics (with and without saliency weighting)**
  - 2 distortion maps
    - Absolute difference
    - Structural similarity (SSIM [Wang04])
  - 4 spatial pooling functions (w_i)
  - 2 "**mean observer**" saliency maps
    - 1 based on fixation number (FN)
    - 1 based on fixation duration (FD)

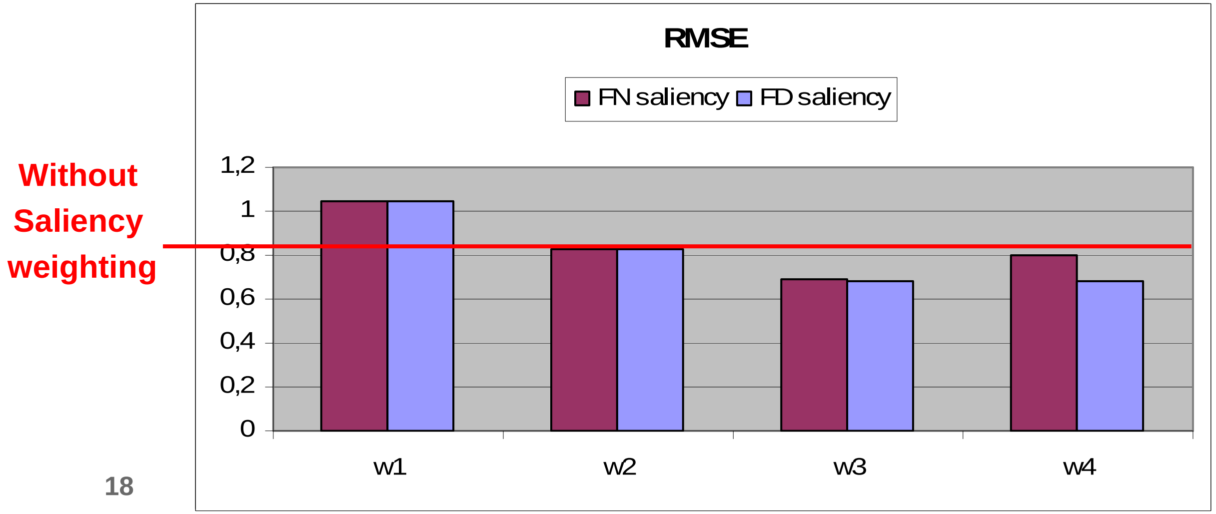
- **On the whole database**
- **In terms of :**
  - Correlation coefficient (CC)
  - Root mean square error (RMSE)

Results : *Predicted MOS vs. MOS*

Mean observer behavior

Distortion maps based on the **absolute difference**





→ Some positive impact of the saliency information

→ The prediction improvement depends on :

- weight function

- saliency map without normalization is better

- saliency based on Fixation Duration (FD) is better

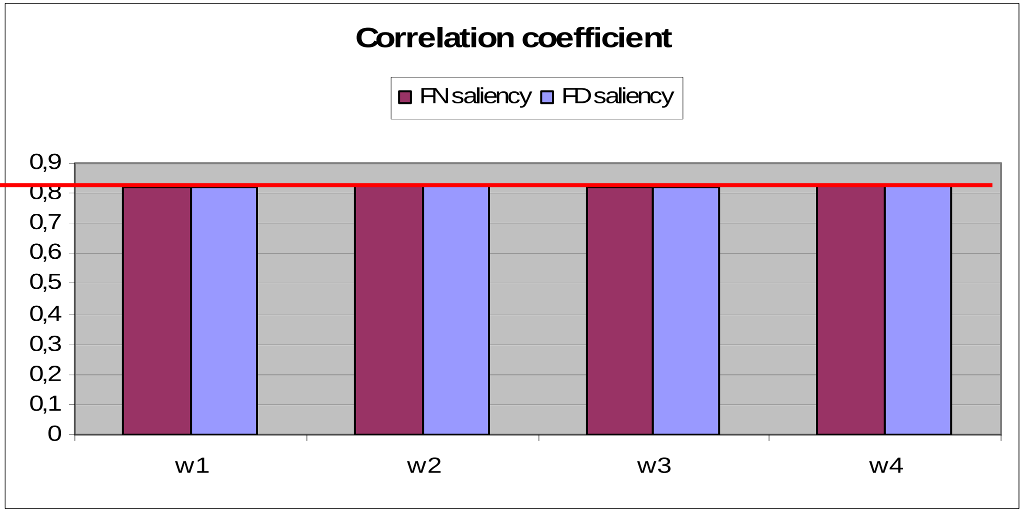
THOMSON *images & beyond*

Results : *Predicted MOS vs. MOS* *3/5*

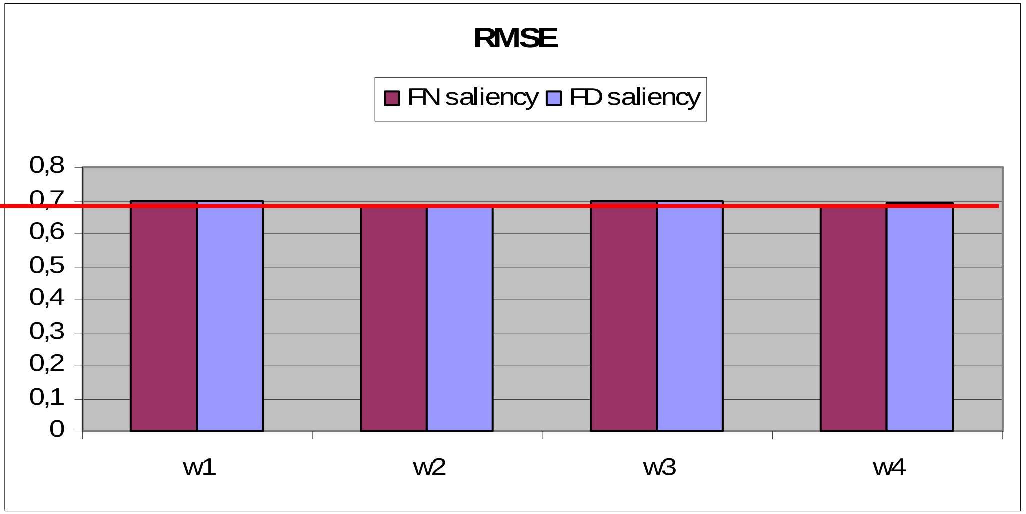
Mean observer behavior

Distortion maps based on the SSIM

Without Saliency weighting



Without Saliency weighting



No meaningful prediction improvement

→ *No positive impact of the saliency information*

THOMSON images & beyond

Results : *Predicted MOS vs. Observer Score* *4/5*

Particular observer behavior
Comparison between observer score and MOSp (predicted MOS) :

- **Of simple quality metrics (with and without saliency weighting)**
  - 2 distortion maps
    - Absolute difference
    - Structural similarity (SSIM [Wang04])
  - 4 spatial pooling functions (w_i)
  - 2 "**particular observer"** saliency maps
    - 1 based on fixation number (FN)
    - 1 based on fixation duration (FD)

- **On the whole database**
- **In terms of :**
  - Correlation coefficient (CC)
  - Root mean square error (RMSE)

Results : *Predicted MOS vs. Observer Score* *5/5*

Particular observer behavior

- Distortion maps based on the absolute difference

Mean of the Δ_{CC} are **-0.08** and **-0.07** for the FN and FD saliency maps respectively

No meaningful prediction improvement

- Distortion maps based on the SSIM

Mean of the Δ_{CC} are **-0.01** and **-0.02** for the FN and FD saliency maps respectively

No meaningful prediction improvement

→ The general non-improvement of the prediction are not due to the average saliency map building

Outline

1. Image quality assessment and visual attention
2. The ground truth
3. Saliency-based quality metrics
4. **Results**
5. Conclusion

Conclusion

→ Prediction improvement is not clearly established

Simple weight functions using the real saliency information, do not clearly improve the prediction capability of simple metrics

→ Prediction improvement on some particular cases is promising (but not acceptable)

→ The way to take into account the visual attention in a quality metric cannot be limited to a simple spatial pooling

Future work :

More studies are required to understand well the visual attention mechanisms in an image quality assessment context,

It seems that the saliency information and the distortion intensity have to be jointly considered in the pooling function

Thank you for your attention !

Appendix

References

[Wang04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, IEEE Trans. on Image Processing, vol. 13, pp. 600–612, 2004.

[Osberger98] W. Osberger, N. Bergmann, and A. Maeder, “An automatic image quality assessment technique incorporating higher level perceptual factors”, IEEE ICIP, pp. 414–418, 1998.

[Barland06] R. Barland and A. Saadane, “Blind quality metric using a perceptual importance map for JPEG-2000 compressed images”, IEEE ICIP, pp. 2941–2944, 2006.

Human saliency map computation

Based on fixation number (FN) :

$$SM'^{(k)}_{FD}(x,y) = \sum_{j=1}^{N_{FP}} \Delta(x - x_i, y - y_i)$$

Based on fixation duration (FD) :

$$SM'^{(k)}_{FD}(x,y) = \sum_{j=1}^{N_{FP}} \Delta(x - x_i, y - y_i) \cdot d(x_j, y_j)$$

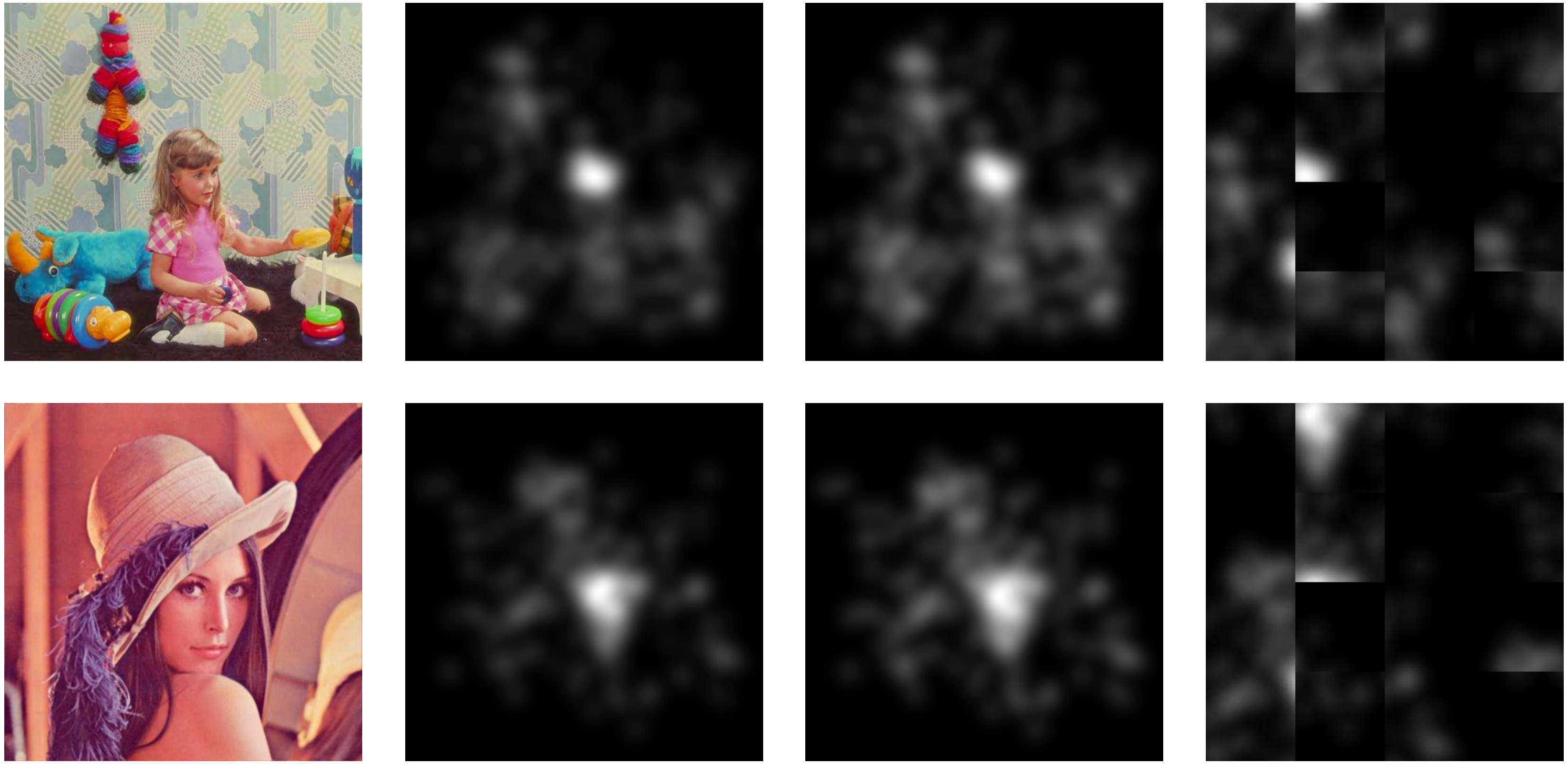
Saliency map (average) :

$$SM'(x,y) = \frac{1}{K} \sum_{k=1}^{K} SM'^{(k)}(x,y)$$

Density saliency map :

$$SM(x,y) = SM'(x,y) * g_{\sigma}(x,y)$$

Examples



Experiments : *The ground truth*