Visual attention
Inter-Observer Visual Congruency (IOVC) and memorability prediction

O. Le Meur

Univ. of Rennes 1
http://people.irisa.fr/Olivier.Le_Meur/
GdR ISIS 2013

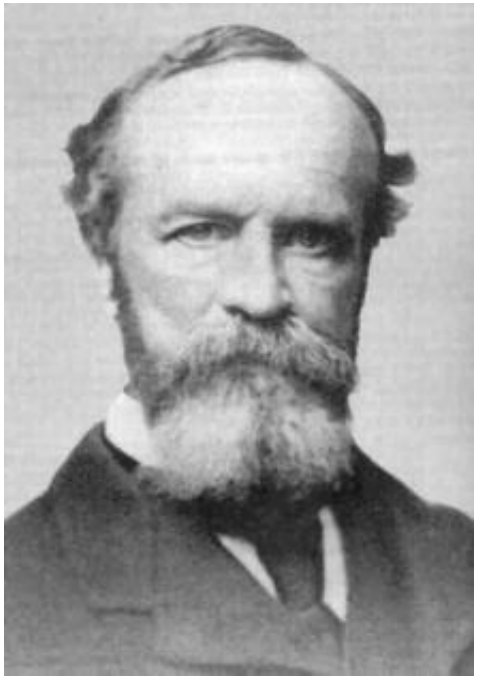
January 16, 2013

Outline

1. What is the visual attention?
2. Inter-Observer Visual Congruency
3. Memorability
4. Conclusion

William James, an early definition of attention...

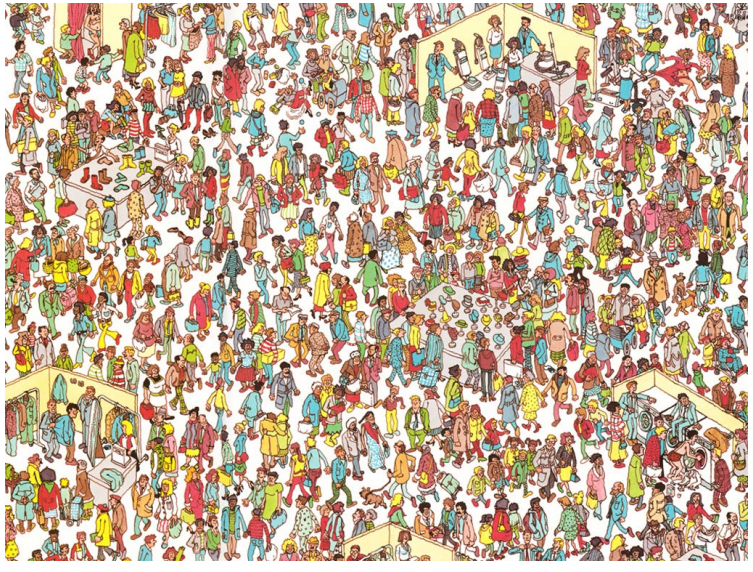
Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought... It implies withdrawal from some things in order to deal effectively with others. William James, 1890[James, 1890]



William James (1842-1910)
A pioneering American psychologist and philosopher.

Human Visual System

Natural visual scenes are cluttered and contain many different objects that cannot all be processed simultaneously.



Amount of information coming down the optic nerve $10^8 - 10^9$ bits per second

Far exceeds what the brain is capable of processing...

Human Visual System

- Example of change blindness.....

YouTube link: `www.youtube.com/watch?v=ubNF9QNEQLA`

Fundamental questions

WE DO NOT SEE EVERYTHING AROUND US!!!

Two majors conclusions come from the change blindness experiments:

- Observers never form a complete, detailed representation of their surroundings;
- Attention is required to perceive change.

It raises fundamental questions:

How do we select information from the scene?

Can we control where and what we attend to?

Do people look always at the same areas?

Outline

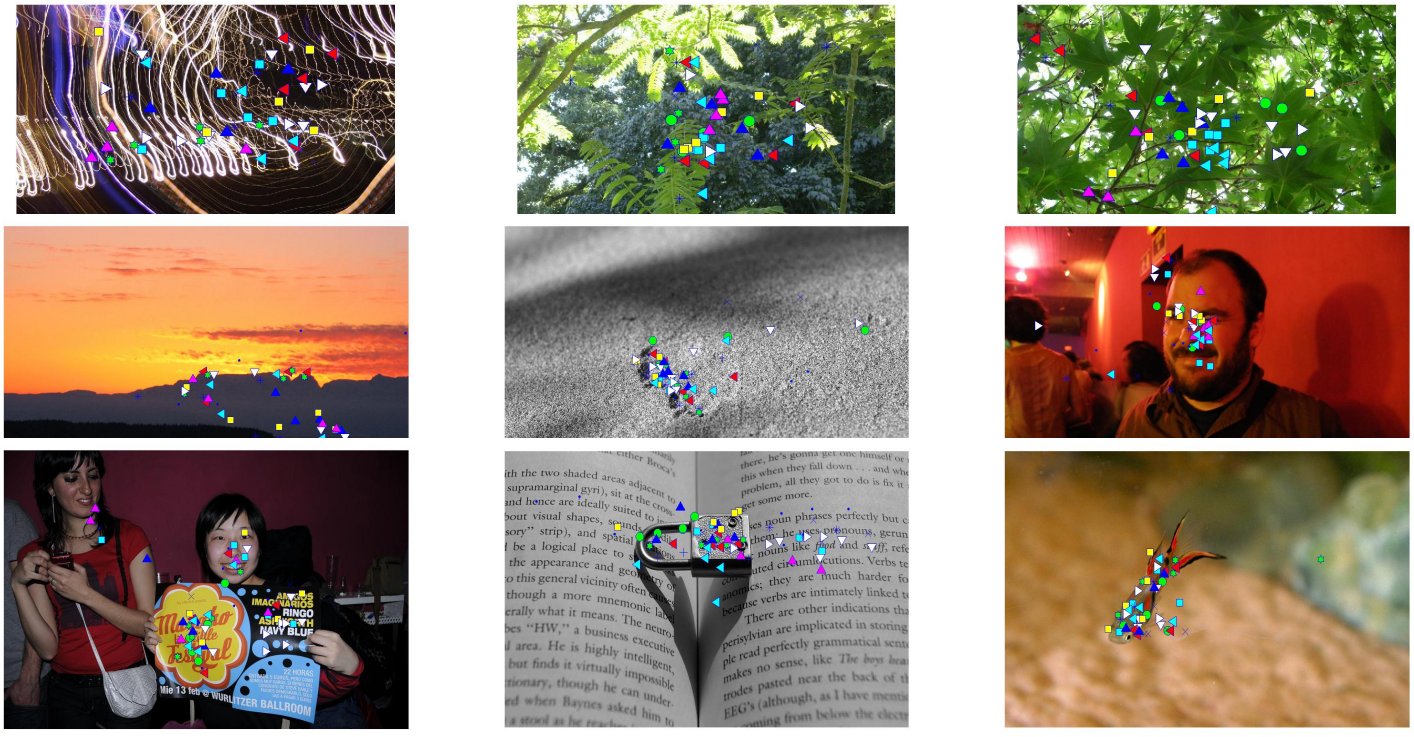
1. What is the visual attention?
2. Inter-Observer Visual Congruency
3. Memorability
4. Conclusion

Problematic and context

Definition (Inter-observer visual congruency)

Inter-observer visual congruency (IOVC) reflects the visual dispersion between observers or the consistency of overt attention (eye movement) while observers watch the same visual scene.

Do observers look at the scene similarly?

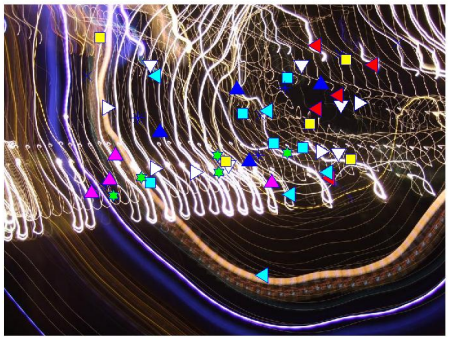


Problematic and context

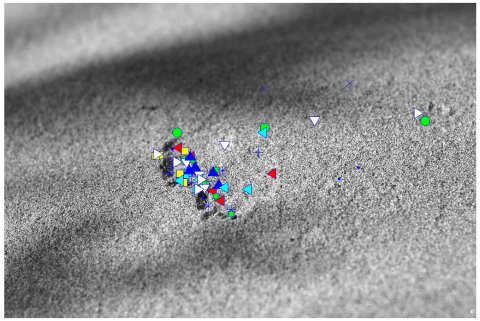
Two issues:

1. Measuring the IOVC by using eye data

LOW



HIGH



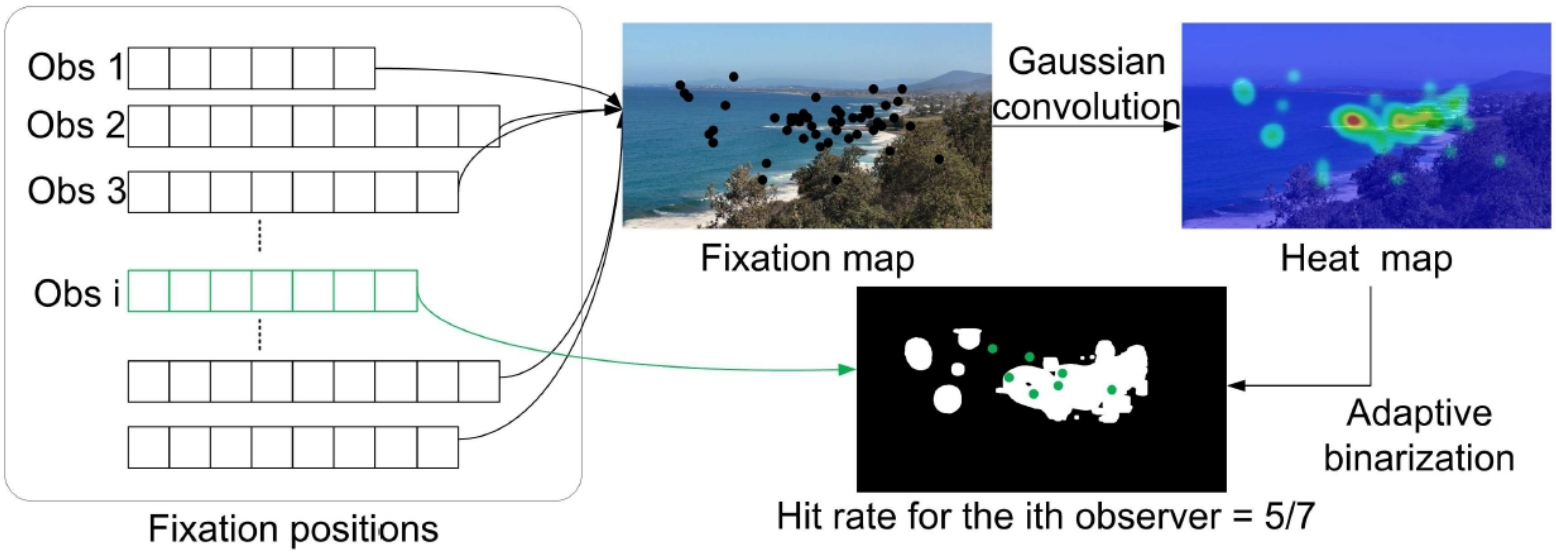
2. Predicting this value with a computational model

O. Le Meur et al., Prediction of the Inter-Observer Visual Congruency (IOVC) and application to image ranking, ACM Multimedia (long paper) 2011.

Measuring the IOVC

To measure the visual dispersion, we use the method proposed by [Torralba et al., 2006]:

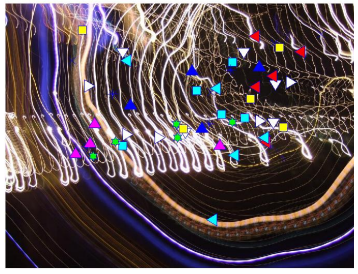
one-against-all approach



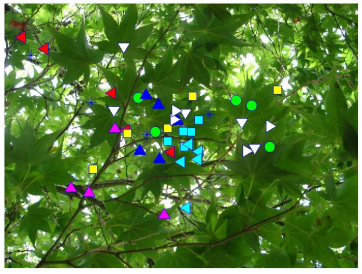
- $IOVC = 1$, strong congruency between observers.
- $IOVC = 0$, lowest congruency between observers.

Measuring the IOVC

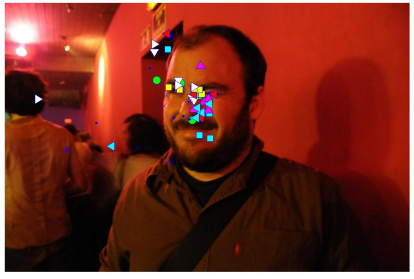
Examples based on MIT eye tracking database



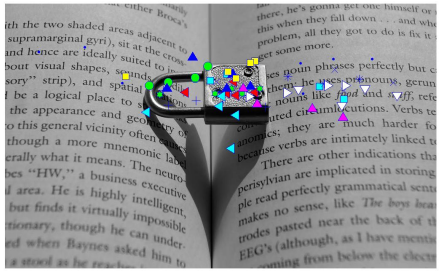
(a) 0.29



(b) 0.23

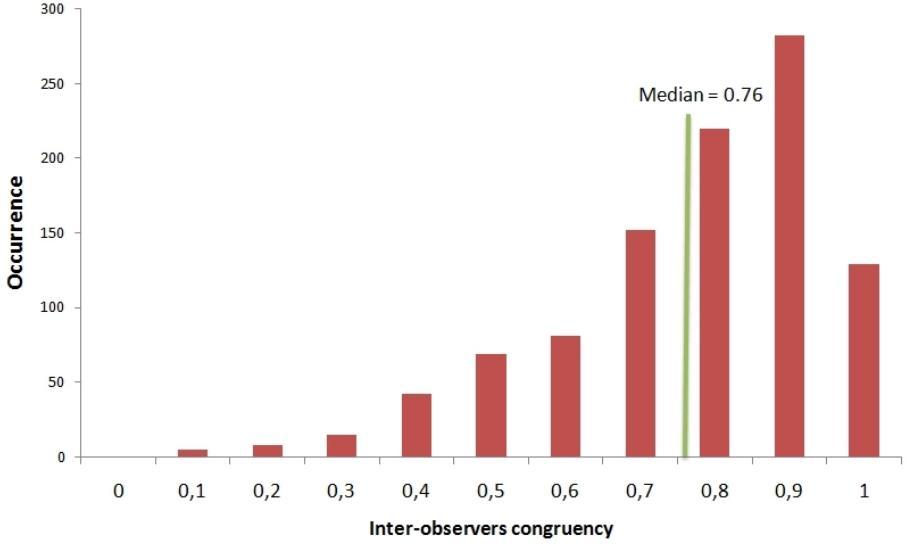


(c) 0.72



(d) 0.85

For the whole database:



Measuring the IOVC

Factors impacting the visual dispersion:

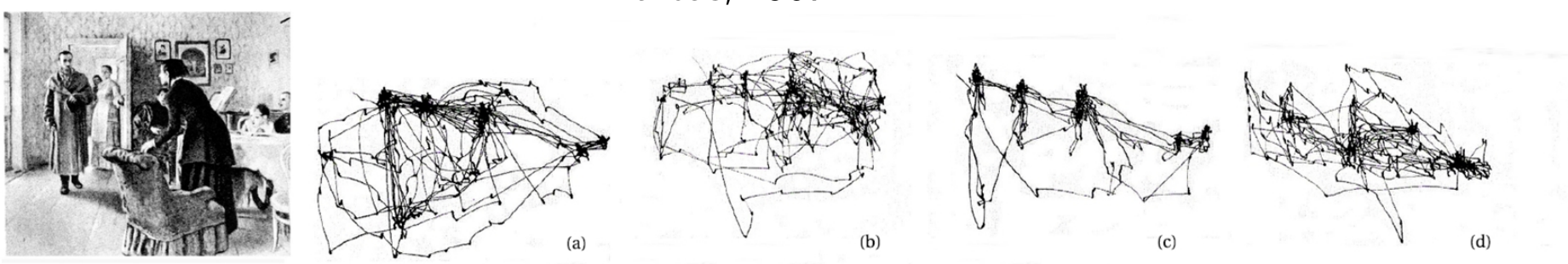
Bottom-up factors (stimulus-dependent)

- Scene complexity
- Scene consistency
- Human face
- Color
- ...

Top-down factors (observer-dependent)

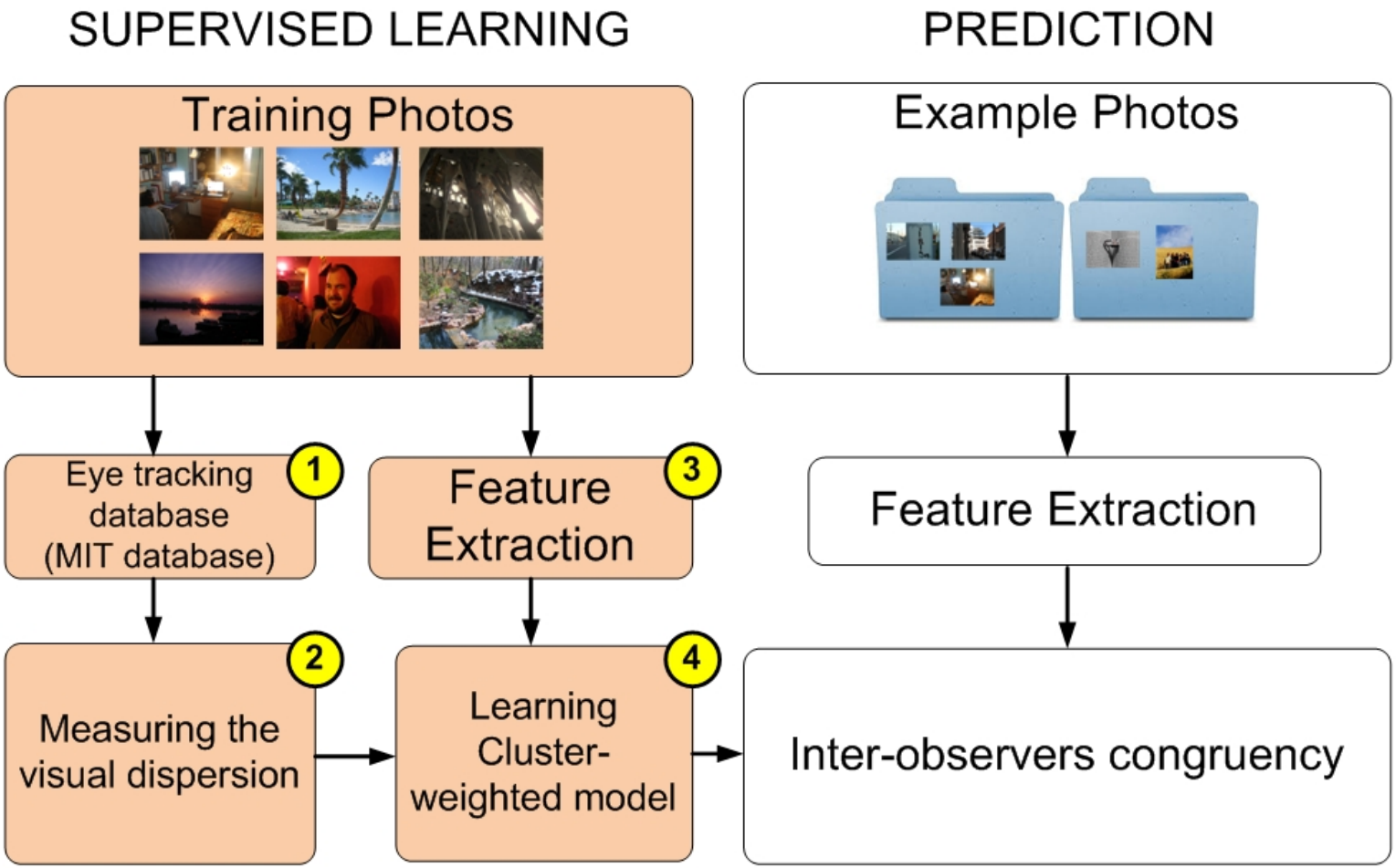
- Task at hand
- Cultural difference
- Prior knowledge
- Prior experience
- ...

Yarbus, 1967



IOVC computational model

Global architecture of the proposed approach:



Feature extraction

6 visual features are extracted from the visual scene:

Mixture between low and high-level visual features

- Color harmony [Cohen et al., 2006]
- Scene complexity:
  - → Entropy E [Rosenholtz et al., 2007]
  - → Number of regions (Color mean shift)
  - → Amount of contours (Sobel detector)
- Faces [Lienhart & Maydt, 2002] (*Open CV*)
- **Depth of Field**

$E = 14.67 dit/pel$, $NbReg = 103$

$E = 14.72 dit/pel$, $NbReg = 72$

Feature extraction

Depth of Field computation inspired by [Luo et al., 2008]:

The horizontal/vertical derivatives histogram is modified after a blurring operation

I the input picture and f_k the blurring kernel of size $k \times k$ ($k = \{3, 5, 7\}$).

1. The vertical and horizontal derivatives are given by:

$$\begin{aligned} p_{xk} &\alpha hist(I * f_k * d_x) \\ p_{yk} &\alpha hist(I * f_k * d_y) \end{aligned}$$

   where $d_x = [1 \ -1]$ and $d_y = [1 \ -1]^T$.

2. For a pixel (i, j) and for a kernel k,

$$D_k(i,j) = \sum_{(n,m) \in W_{ij}} KL(p_{xk}|p_{x1})(n,m) + KL(p_{yk}|p_{y1})(n,m)$$

3. The DoF value is finally computed as follows:

$$DoF = \sum_{(i,j) \in I} \sum_k D_k(i,j)$$

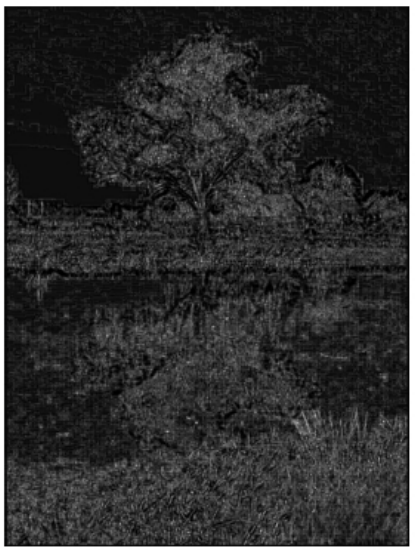
Feature extraction

Depth of Field computation inspired by [Luo et al., 2008]:

(a) Original

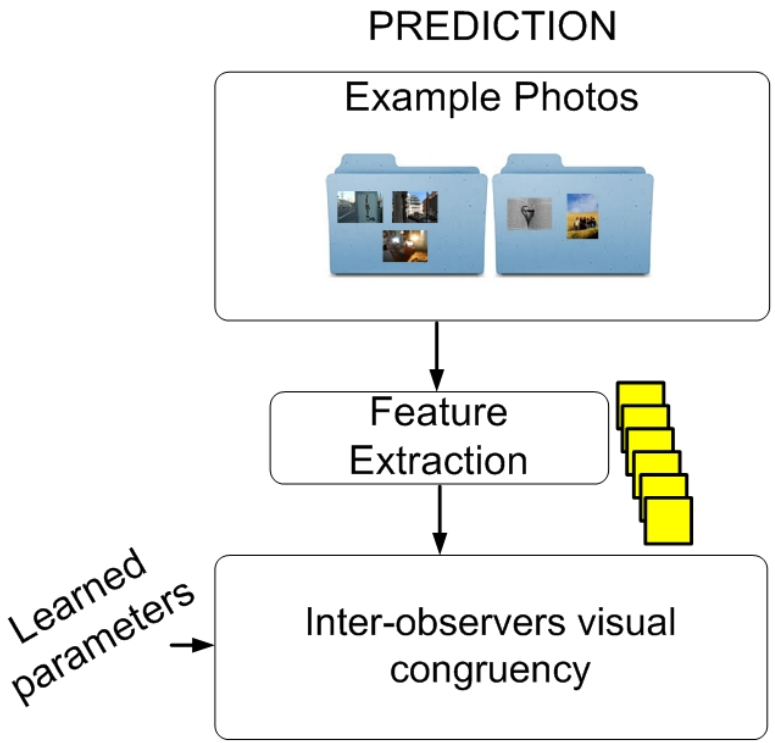
(b) DoF=0.12

(c) Original

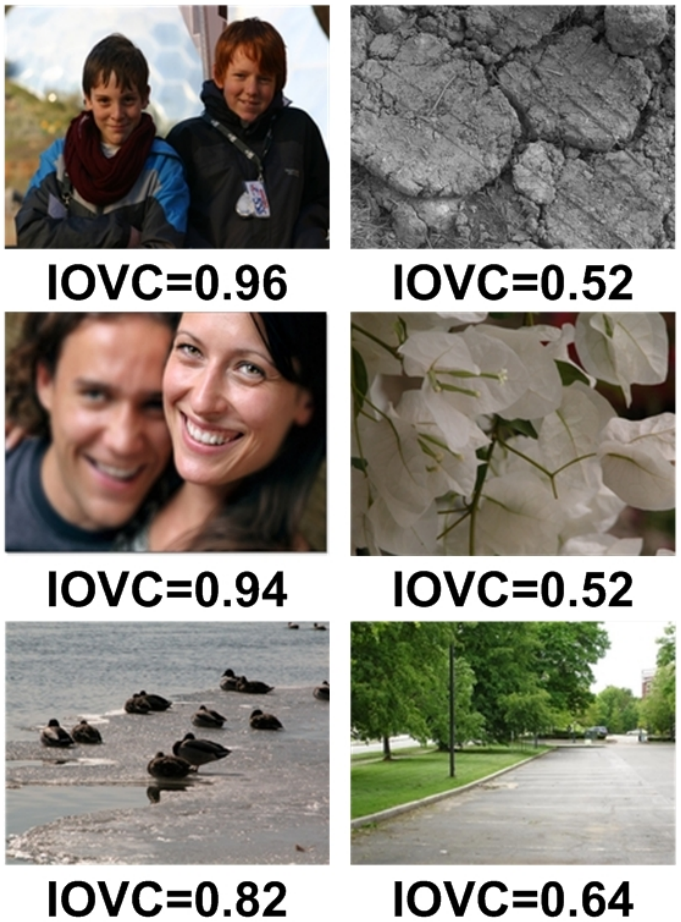

(d) DoF=0.75

Feature extraction

After the learning, we can now predict:



Predicted IOVC:



Performance

Two tracking databases are used to evaluate the performance:

- MIT's dababase (1003 pictures, 15 observers)
- Le Meur's database (27 pictures, up to 40 observers)

(a) 0.94,0.96

We compute the Pearson correlation coefficient between predicted IOVC and 'true' IOVC:

- MIT's dababase:
  $r(2004) = 0.34,\ p < 0.001$
- Le Meur's database:
  $r(54) = 0.24,\ p < 0.17$

(b) 0.89,0.94

(c) 0.91,0.91

Performance

Some cases for which the model fails....

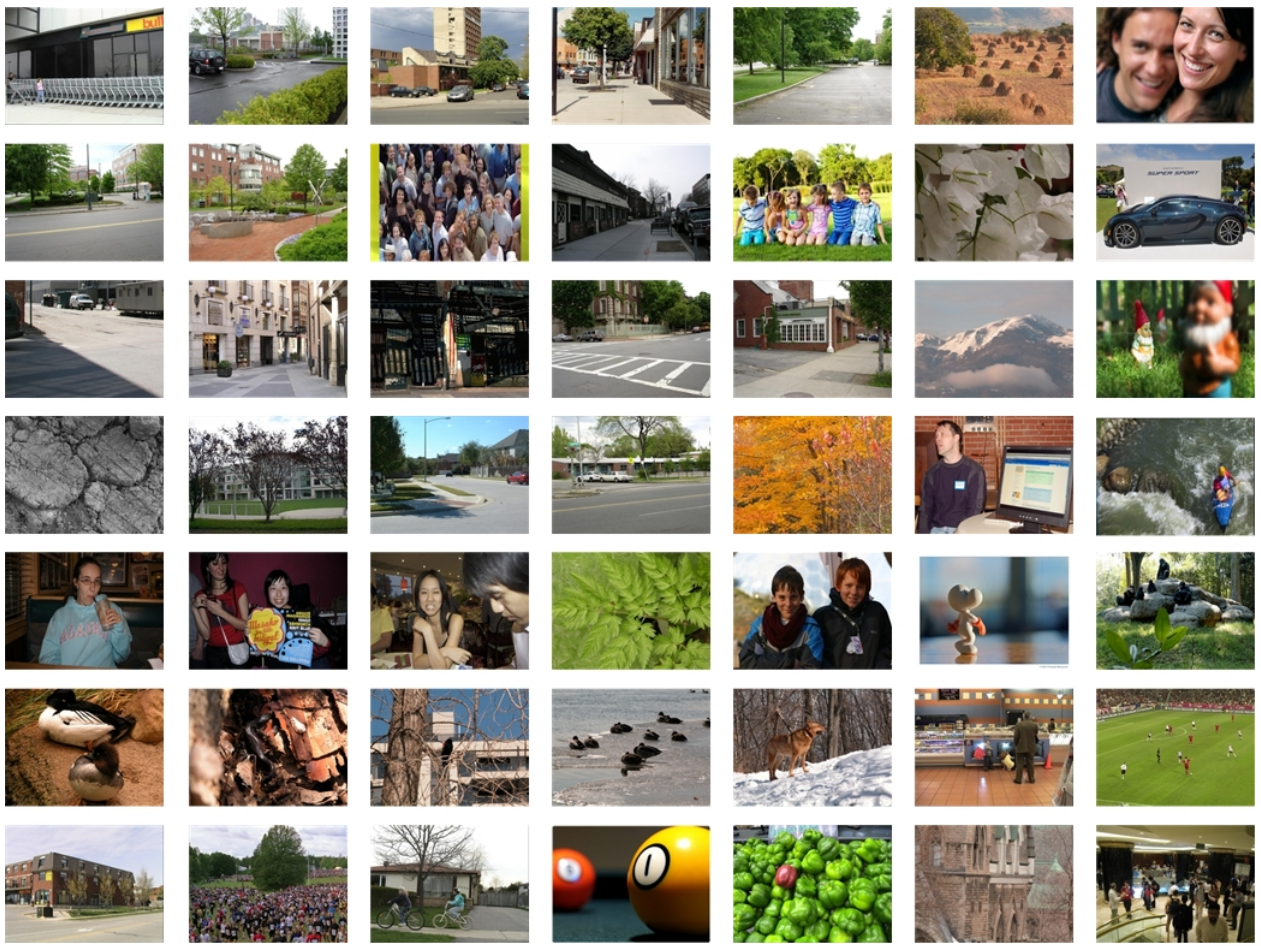
(a) (0.83,0.59)

(b) Heat map

Figure: (.,.), true IOVC and predicted IOVC

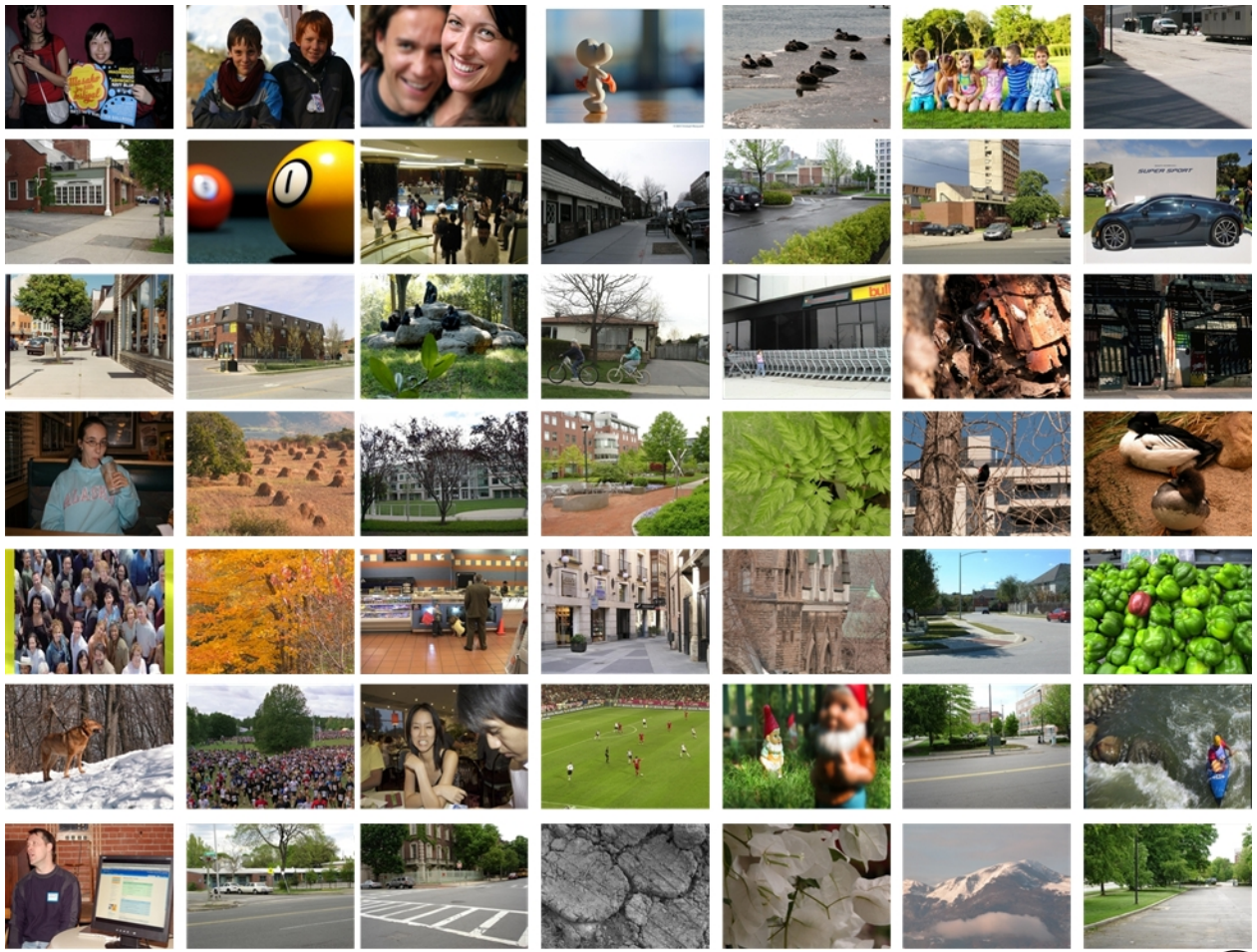
Application to image ranking

Goal: to sort out a collection of pictures in function of IOVC scores.



Application to image ranking

Goal: to sort out a collection of pictures in function of IOVC scores.



Application to image ranking

Goal: to sort out a collection of pictures in function of IOVC scores.

Top five pictures:

IOVC=0.98 IOVC=0.96 IOVC=0.94 IOVC=0.85 IOVC=0.82

Last five pictures:

IOVC=0.46 IOVC=0.52 IOVC=0.52 IOVC=0.64 IOVC=0.64

- can be combined with other factors to evaluate the interestingness or attractivness of visual scenes.

Conclusion

Contributions:

- Main contribution: A new approach to evaluate the visual dispersion between observers
  - → based on low-level visual features
  - → it outperforms the visual clutter approach [Rosenholtz et al., 2007]
- Side contributions:
  - → A new method to estimate the DoF
  - → IOVC-based image ranking

Rooms to improve applications such as:

- Video compression
- Computational model of visual attention
- Retargeting
- Estimating the attractiveness of picture...

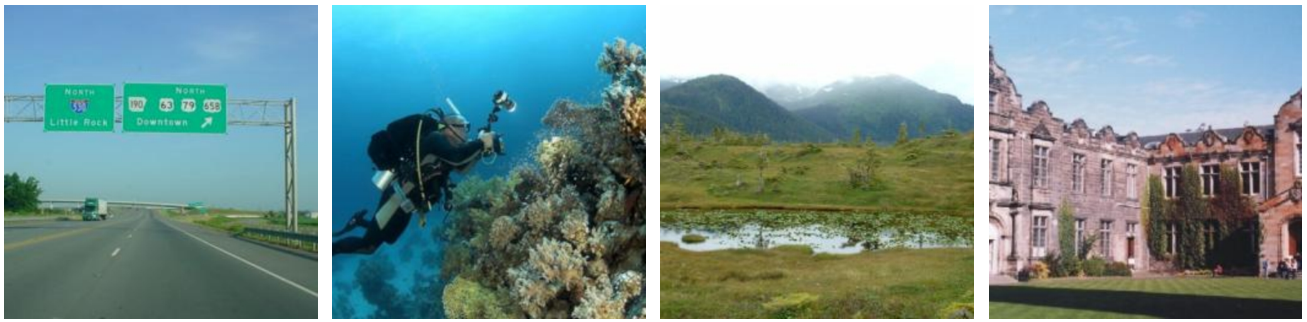
Outline

1. What is the visual attention?
2. Inter-Observer Visual Congruency
3. Memorability
4. Conclusion

Definition (Memorability)

Humans are extremely good at remembering thousands of pictures. However not all images are created equal. Some are much more memorable than others.

- A recent work done by Isola et al. [Isola et al., 2011] provides the first computational model predicting the memorability of a picture.



(a) M=0.85 (b) M=0.81 (c) M=0.42 (d) M=0.37

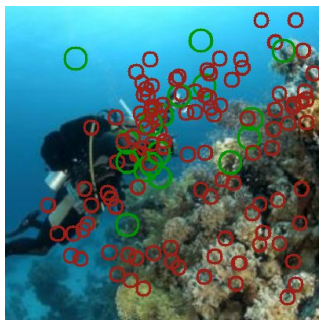
Extracted from [Isola et al., 2011]

- No link with visual attention....

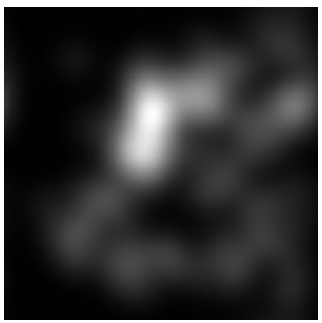
Submitted to ICIP'2013: **M. Mancas and O. Le Meur, Memorability of natural scenes: the role of attention.**

- Eye tracking on Isola's dataset:

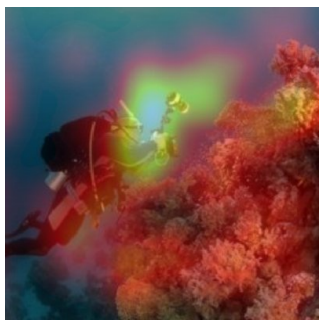
(a) Orig.



(b) Fix. Map.

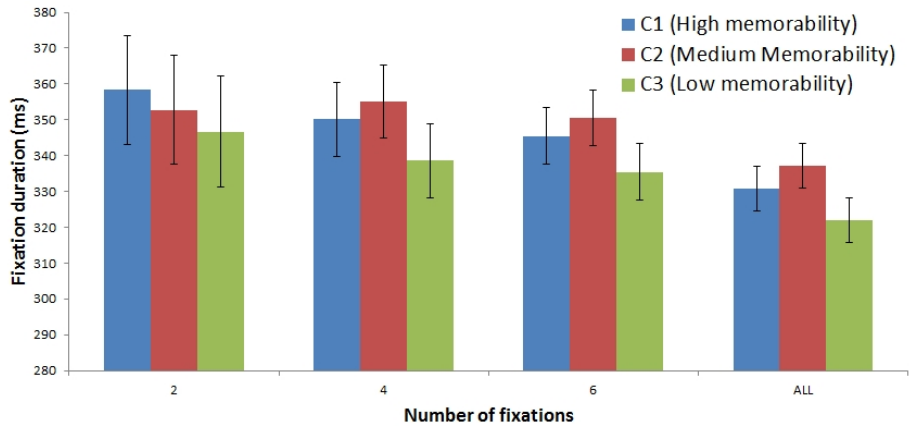


(c) Saliency

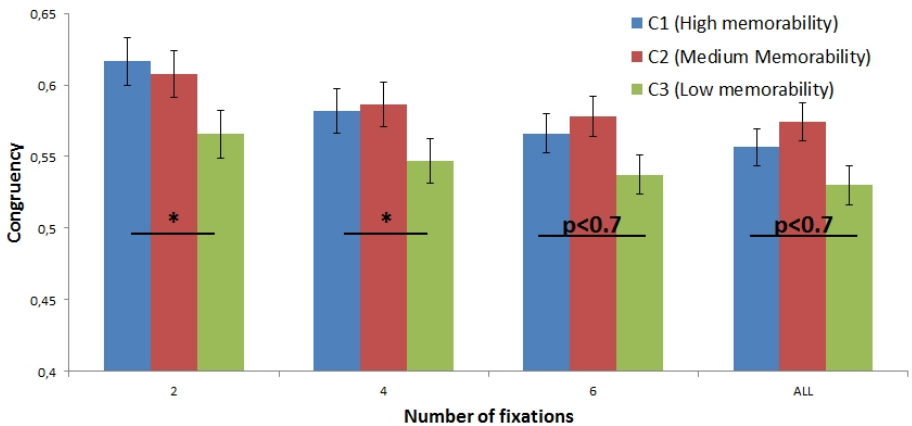


(d) Heat

From [Le Meur & Baccino, 2012]



(a) Fixation durations



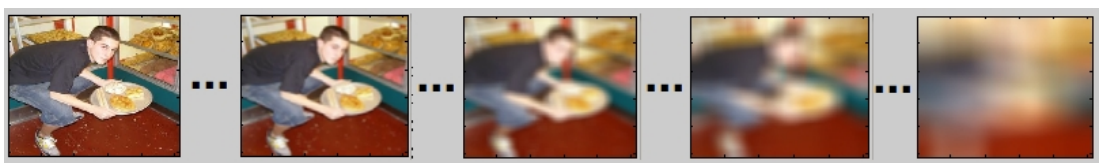
(b) Congruency

Isola's algorithm is based on:

- a support vector regression mapping global features to memorability scores
- global features:
  → GIST, SIFT, HOG2 $\times$ 2, SSIM (Matching Local Self-Similarities)
- the combination of global features performs at a rank correlation of 0.46.

We have used the same learning code and we add three features based on saliency but removed the 512 parameters of the gist:

- Coverage
- Two features related to the visibility of object:



We improve the correlation of 2 percents with less input features...

	Cov.	Vis.	Best (No GIST)	Best Isola
ρ	0.10	0.27	0.48	0.46

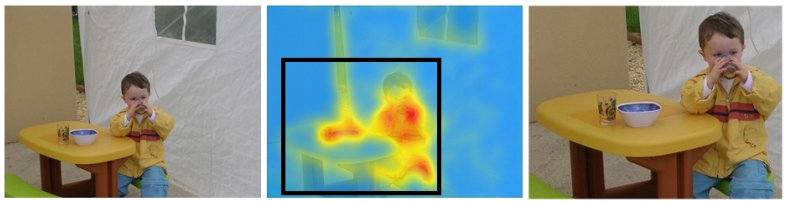
Outline

1. What is the visual attention?
2. Inter-Observer Visual Congruency
3. Memorability
4. Conclusion

Understanding how we perceive the world is fundamental for many applications.

- Visual dispersion
- Memorability
- Quality assessment
- Compression
- Indexation

- Retargeting:



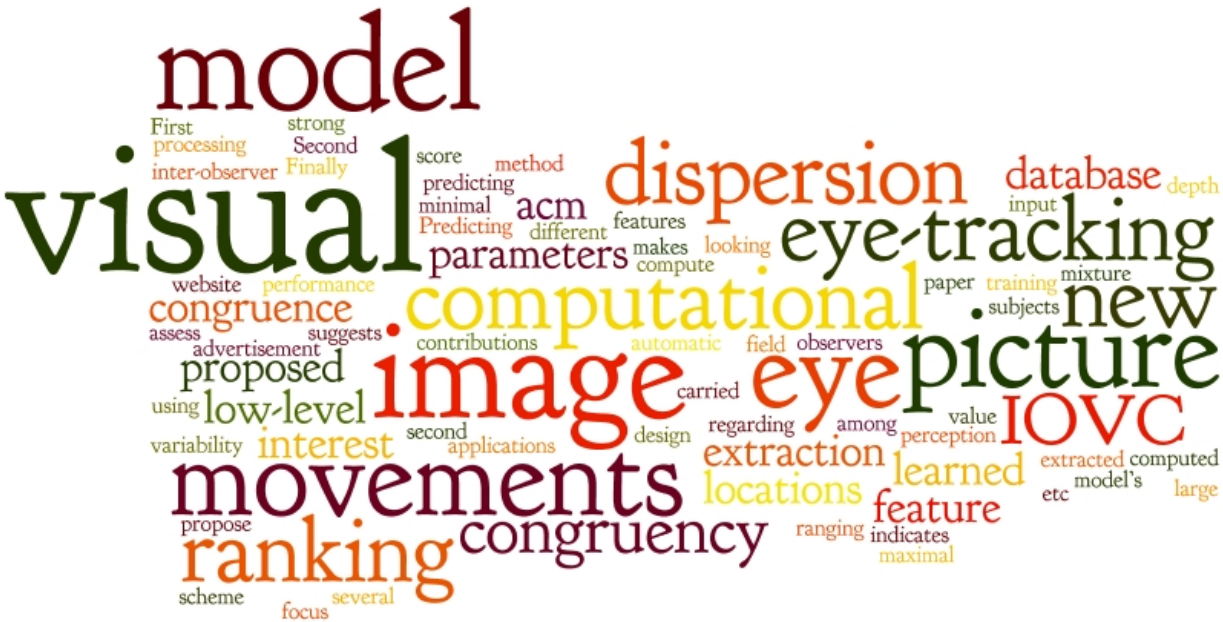
From [Le Meur & Le Callet, 2009]

- Region-of-interest extraction

From [Liu et al., 2012]

A special session has been submitted to WIAMIS'2013 (Paris):
O. Le Meur and Z. Liu, Visual attention, a multidisciplinary topic: from behavioral studies to computer vision applications

Conclusion



Any questions?

References

[Cohen et al., 2006] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. In ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH), volume 56, pages 624-630, 2006.

[Isola et al., 2011] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 145?152, 2011.

[James, 1890] W. James. The Principles of Psychology, 1890.

[Le Meur & Le Callet, 2009] O. Le Meur and P. Le Callet, What we see is most likely to be what matters: visual attention and applications, ICIP 2009.

[Le Meur & Baccino, 2012] O. Le Meur and T. Baccino, Methods for comparing scanpaths and saliency maps: strengths and weaknesses, Behavior Research Method, 2012

[Lienhart & Maydt, 2002] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In ICIP, volume 1, pages 900-903, 2002.

[Liu et al., 2012] Z. Liu and R. Shi, et. al., Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut, IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 1275-1289, Aug. 2012.

[Luo et al., 2008] Y. Luo and X. Tang. Photo and video quality evaluation: focussing on the subject. In ECCV, pages 386-399, 2008.

[Rosenholtz et al., 2007] R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. Journal of Vision, 7(2), March 2007.

[Torralba et al., 2006] A. Torralba, A. Oliva, M. Castelhano, & J. Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological review, 113(4):766-786, 2006.

esir ECOLE SUPERIEURE D'INGENIEUR DE RENNES UNIVERSITÉ DE RENNES 1