Visual attention
Models, Performances and applications

O. Le Meur

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

January 29, 2013

Outline

Outline

COMPLEX VISUAL SCENES

(a) Prey vs Predator

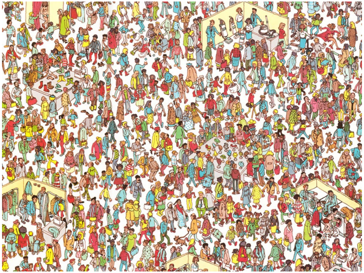
(b) King of the world?

(c) Salvador Dali

(d) René Magritte

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?

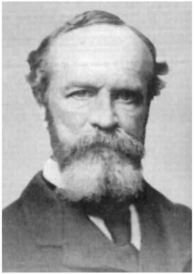
Do people look always at the same areas?

They are all connected to VISUAL ATTENTION.

Outline

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 thoughts... 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.

Selective Attention is the natural strategy for dealing with this bottleneck.

Visual Attention definition

As Posner proposed [Posner, 1980], visual attention is used:

- to select important areas of our visual field (alerting);
- to search a target in cluttered scenes (searching).

There are two kinds of visual attention:

- Overt visual attention: involving eye movements;

  Often compared to a window on the brain...[Henderson et al., 2007]

- Covert visual attention: without eye movements (Covert fixations are not observable). Attention can be voluntarily focussed on a peripheral part of the visual field (as when looking out of the corner of one's eyes).

  Covert attention is the act of mentally focusing on one of several possible sensory stimuli.

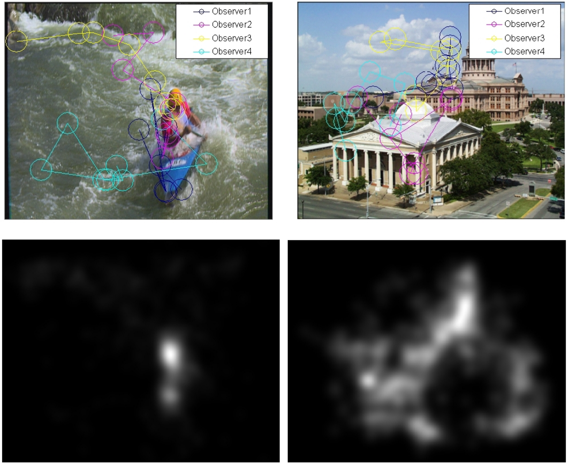
Overt attention - Eye Movements

There exists different kinds of eye movements:

- **Saccade**: quick eye movements from one fixation location to another. The length of the saccade is between 4 to 12 degrees of visual angle;
- **Fixation**: phase during which eyes is almost stationnary. The typical duration is about 200 / 300 ms [Findlay & Gilchrist, 2003]. But it depends on a number of factors (depth of processing [Velichkovsky et al., 2002]; ease or difficulty to perceive something [Mannan et al., 1995]);
- **Smooth pursuit**: voluntary tracking of moving stimulus;
- **Vergence**: coordinated movement of both eyes, converging for objects moving towards and diverging for objects moving away from the eyes.

A scanpath is a sequence of eye movements (fixations - smooth pursuit - saccades...).

Overt attention - Eye Movements



From [Le Meur & Chevet, 2010].

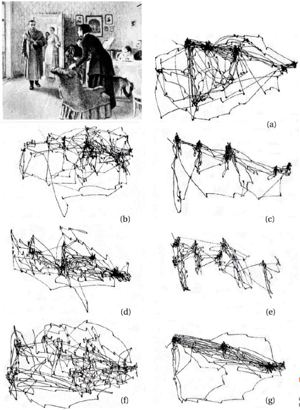
Yarbus [Yarbus, 1967]

(1914-1986)

A. Yarbus [Yarbus, 1967] demonstrated how eye movements changed depending on the question asked of the subject.

1. No question asked
2. Judge economic status
3. What were they doing before the visitor arrived?
4. What clothes are they wearing?
5. Where are they?
6. How long is it since the visitor has seen the family?
7. Estimate how long the unexpected visitor had been away from the family.

Each recording lasted 3 minutes.



Yarbus [Yarbus, 1967]

A. Yarbus showed that our attention depends on bottom-up features and on top-down information:

- Bottom-up attention (also called exogenous): some things draw attention reflexively, in a task-independent way...
  - → Involuntary;
  - → Very quick;
  - → Unconscious.

- Top-down attention (also called endogenous): some things draw volitional attention, in a task-dependent way...
  - → Voluntary (for instance to perform a task, find the red target);
  - → Very slow;
  - → Conscious.

In this talk, we are interested in bottom-up saliency which is a signal-based factor that dictates where attention is to be focussed.

Outline

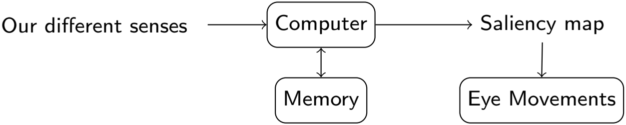
Computational models: two schools of thought

- One based on the assumption that there is an unique saliency map [Koch et al., 1985, Li, 2002]:

Definition (saliency map)

A topographic representation that combines the information from the individual feature maps into one global measure of conspicuity. This map can be modulated by a higher-level feedback.

A comfortable view for the computational modelling...



- There exist multiple saliency maps (distributed throughout the visual areas) [Tsotsos et al., 1995].

  Many candidate locations for a saliency map:
  - → Primary visual cortex[Li, 2002]
  - → Lateral IntraParietal area (LIP) [Kusunoki et al., 2000]
  - → Medial Temporal cortex [Treue et al., 2006]

The brain is not a computer [Varela, 1998].

Computational models

The following taxonomy classifies computational models into 3 classes:

Biologically inspired

- [Itti et al., 1998]
- [Le Meur et al., 2006, Le Meur et al., 2007]
- [Bur et al., 2007]
- [Marat et al., 2009]

Probabilistic models

- [Oliva et al., 2003]
- [Bruce, 2004, Bruce & Tsotsos, 2009]
- [Mancas et al., 2006]
- [Itti & Baldi, 2006]
- [Zhang et al., 2008]

Beyond bottom-up models

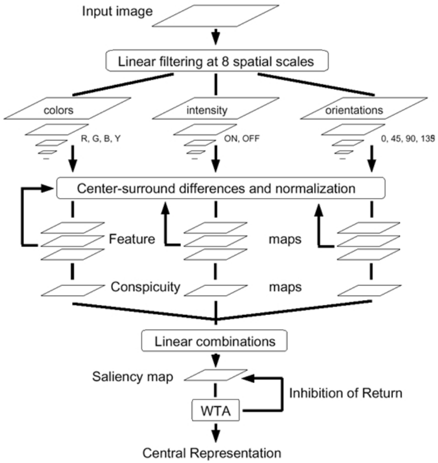
- Goal-directed models: [Navalpakkam & Itti, 2005]
- Machine learning-based models: [Judd et al., 2009]

A recent review presents a taxonomy of nearly 65 models [Borji & Itti, 2012].

Biologically inspired models

Itti's model [Itti et al., 1998], probably the most known...

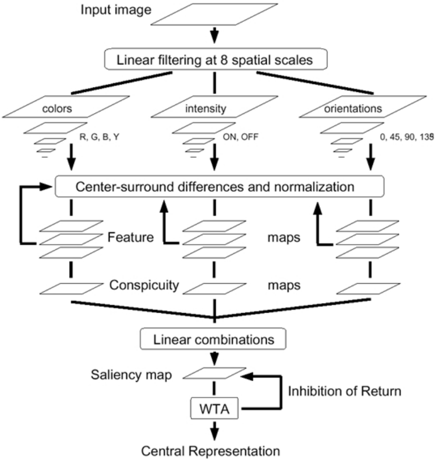
- Based on the Koch and Ullman's scheme;
- Hierarchical decomposition (Gaussian);
- Early visual features extraction in a massively parallel manner;
- Center-surround operations;
- Pooling of the feature maps to form the saliency map.



Biologically inspired models

Itti's model [Itti et al., 1998], probably the most known...

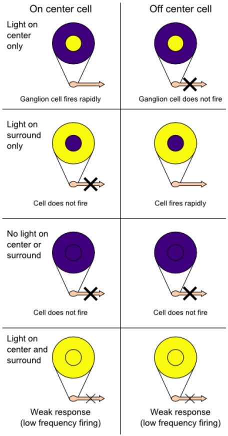
- Hierarchical decomposition (Gaussian):



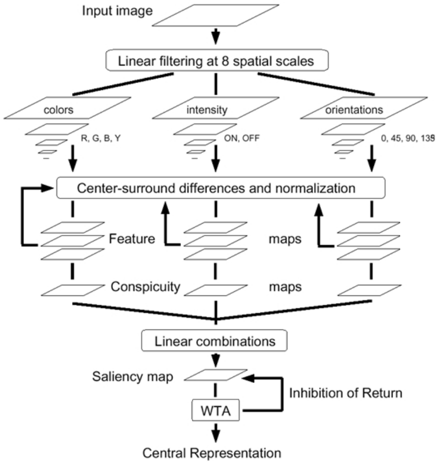
Biologically inspired models

Itti's model [Itti et al., 1998], probably the most known...

- Center-surround operations:



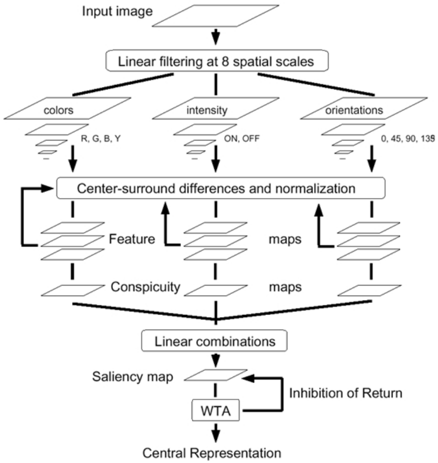
From Wikipedia.



Biologically inspired models

Itti's model [Itti et al., 1998], probably the most known...

- Normalization and combination:
  - → **Naive summation**: all maps are normalized to the same dynamic and averaged;
  - → **Learning linear combination**: depending on the target, each feature map is globally multiplied by a weighting factor. A simple summation is then performed.
  - → **Content-based global non-linear amplification**:
    1. Normalize all the feature maps to the same dynamic range;
    2. For each map, find its global maximum M and the average $\overline{m}$ of all the other local maxima;
    3. Globally multiply the map by $(M - \overline{m})^2$.



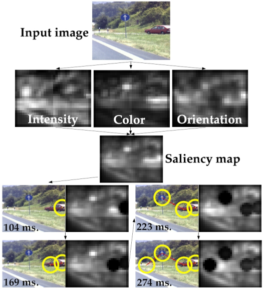
Biologically inspired models

Itti's model [Itti et al., 1998], probably the most known...

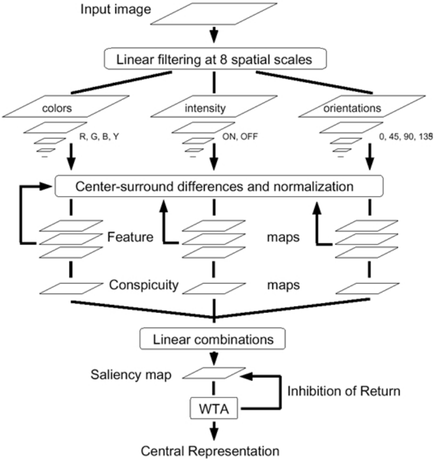
- Winner-Take-All:

  The maximum of the saliency map ⇒ the most salient stimulus ⇒ focus of attention

  Inhibitory feedback from WTA [Koch et al., 1985]



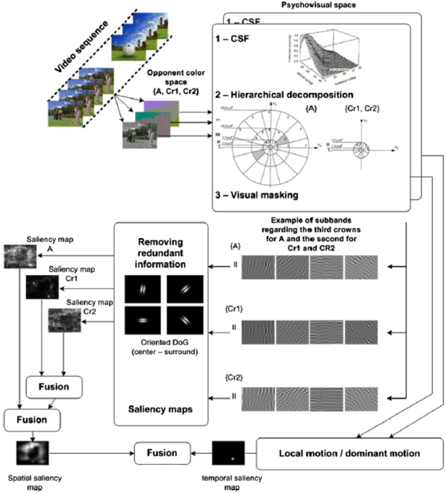
From Itti's Thesis.



Biologically inspired models

Le Meur's model [Le Meur et al., 2006], an extension of Itti's model...

- Based on the Koch and Ullman's scheme
- Light adaptation and Contrast Sensitivity Function
- Hierarchical and oriented decomposition (Fourier spectrum)
- Early visual features extraction in a massively parallel manner
- Center-surround operations on each oriented subband
- Enhanced pooling [Le Meur et al., 2007] of the feature maps to form the saliency map



Other models in the same vein: [Marat et al., 2009], [Bur et al., 2007]...

Probabilistic models

Such models are based on a probabilistic framework taken their origin in the information theory.

Definition (Self-information)

Self-information is a measure of the amount information provided by an event. For a discrete X r.v defined by $\mathcal{A} = \{x_1, ..., x_N\}$ and by a pdf, the amount of information of the event $X = x_i$ is given by:

$$I(X = x_i) = -log_2 p(X = x_i), \text{ bit/symbol}$$

Properties:

- if $p(X = x_i) < p(X = x_j)$ then $I(X = x_i) > I(X = x_j)$
- $p(X = x_i) \rightarrow 0$, $I(X = x_i) \rightarrow +\infty$

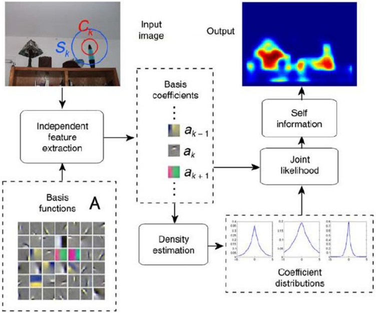
The saliency of visual content could be deduced from the self-information measure.

Self-information $\equiv$ rareness, surprise, contrast...

Probabilistic models

- First model resting on this approach has been proposed in 2003 [Oliva et al., 2003]: $S(\mathbf{x}) = \frac{1}{p(\mathbf{v}_l(\mathbf{x}))}$, where $\mathbf{v}_l$ is a feature vector (48 dimensions), the probability density function is computed over the whole image.

- Bruce in 2004 [Bruce, 2004] and 2009 [Bruce & Tsotsos, 2009] modified the previous approach by using the self-information locally on independent coefficients (projection on a given basis).



From [Bruce & Tsotsos, 2009]

Other models in the same vein: [Mancas et al., 2006], [Zhang et al., 2008]

Probabilistic models

The support regions used to calculate the pdf are not the same:

- Local spatial surround [Gao et al., 2008, Bruce & Tsotsos, 2009]. Note also that Gao et al. [Gao et al., 2008] uses the mutual information to quantify the saliency;
- Whole image [Oliva et al., 2003, Bruce & Tsotsos, 2006];
- Natural image statistics [Zhang et al., 2008] (self-information);
- Temporal history, theory of surprise [Itti & Baldi, 2006].

Machine learning-based models

Judd's model [Judd et al., 2009] combines bottom-up and top-down information:

- Low-level features:
  - → Local energy of the steerable pyramid filters (4 orientations, 3 scales);
  - → Three additional channels (Intensity, orientation and color) coming from Itti's model;
- Mid-level features: Horizon line coming from the gist;
- High-level features:
  - → Viola and Jones face detector;
  - → Felzenszwalb person and car detectors;
  - → Center prior.

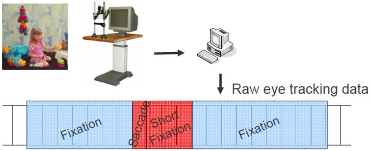
The liblinear support vector machine is used to train a model on the 9030 positive and 9030 negative training samples.

Some examples

(a) Original (b) Itti (c) Le Meur (d) Bruce (e) Zhang (f) Judd

Outline

From a fixation map to a saliency map



- Discrete fixation map f^i for the i^{th} observer (M is the number of fixations):

$$f^i(\mathbf{x}) = \sum_{k=1}^{M} \delta(\mathbf{x} - \mathbf{x}_{f(\boldsymbol{k})}) \tag{1}$$

- Continuous saliency map S (N is the number of observers):

$$S(\mathbf{x}) = \left(\frac{1}{N} \sum_{i=1}^{N} f^i(\mathbf{x}) \right) * G_\sigma(\mathbf{x}) \tag{2}$$

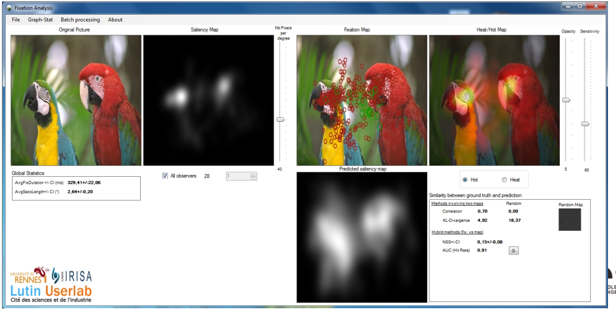
(g) Original (h) Fixation map (i) Saliency map (j) Heat map

Three principal methods

Three principal methods to compare two saliency maps:

- Correlation-based measure;
- Divergence of Kullback-Leibler;
- ROC analysis.

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



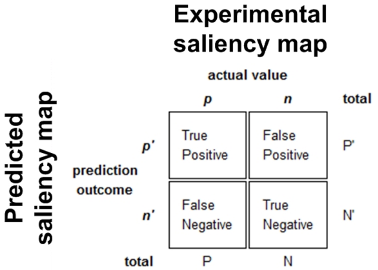
Three principal methods

ROC analysis (1/2)

Definition (ROC)

The Receiver Operating Characteristic (ROC) analysis provides a comprehensive and visually attractive framework to summarize the accuracy of predictions.

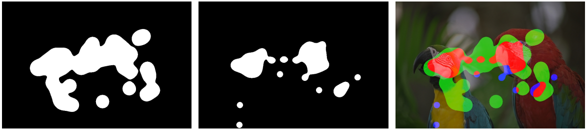
The problem is here limited to a two-class prediction (binary classification). Pixels of the ground truth as well as those of the prediction are labeled either as fixated or not fixated.



- Hit rate (TP)
- ROC curve
- AUC (Area Under Curve)
  - $\rightarrow$ AUC=1 $\Leftrightarrow$ perfect;
    AUC=0.5 $\Leftrightarrow$ random.

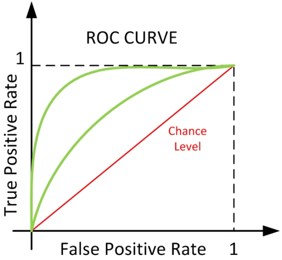
Three principal methods

ROC analysis (2/2)



(a) Reference (b) Predicted (c) Classification

A ROC curve plotting the false positive rate as a function of the true positive rate is usually used to present the classification result.



Advantages:

- + Invariant to monotonic transformation
- + Well defined upper bound

Drawbacks:

- − ...

Hybrid methods

Four methods involving scanpaths and saliency maps :

- Receiver Operating Analysis;
- Normalized Scanpath Saliency [Parkhurst et al.,2002, Peters et al., 2005];
- Percentile [Peters & Itti, 2008];
- The Kullback-Leibler divergence [Itti & Baldi, 2006].

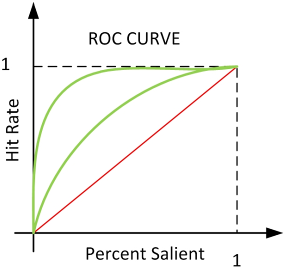
Hybrid methods

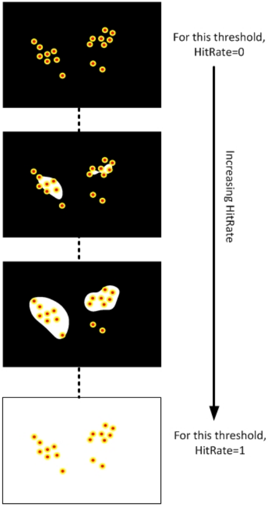
Receiver Operating Analysis

ROC analysis is performed between a continuous saliency map and a set of fixations.

- Hit rate is measured in function of the threshold used to binarize the saliency map [Torralba et al., 2006, Judd et al., 2009]:

ROC curve goes from 0 to 1!

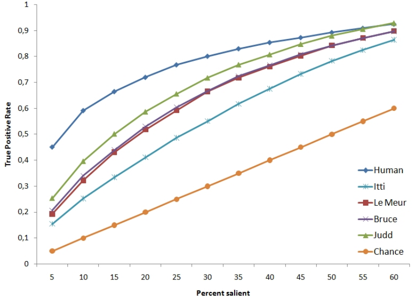




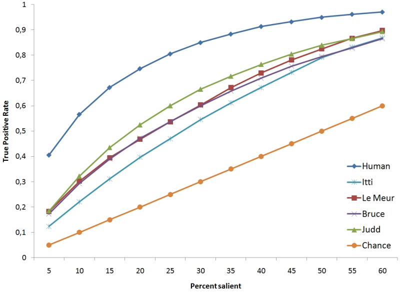
Performances

Comparison of four state-of-the-art models (Hit Rate) by using two dataset of eye movement:

N. Bruce's database:



O. Le Meur's database:



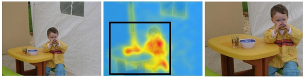
The inter-observer dispersion can be used as to the define the upper bound of a prediction

Outline

Attention-based applications

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

- Visual dispersion
- Memorability
- Quality assessment
- Advertising and web usability
- Robot active vision
- Non Photorealistic rendering
- Retargeting / image cropping:

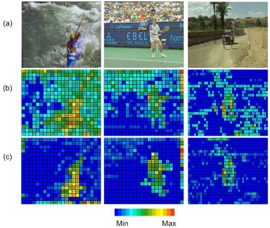


From [Le Meur & Le Callet, 2009]

- Region-of-interest extraction:

From [Liu et al., 2012]

- Image and video compression:



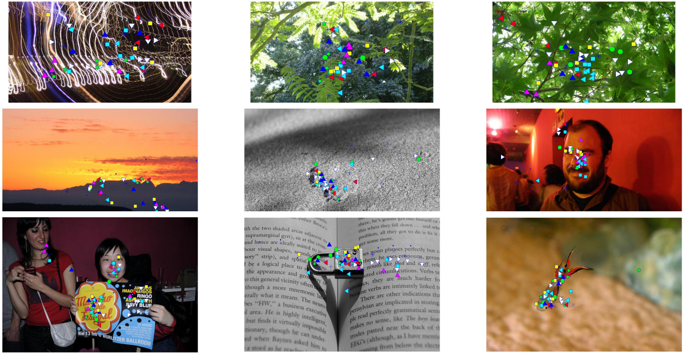
Distribution of the encoding cost for H.264 coding without and with saliency map.

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?

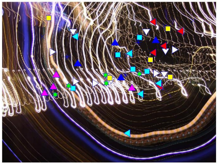


Problematic and context

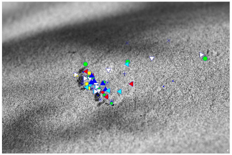
Two issues:

1. Measuring the IOVC by using eye data

LOW



HIGH



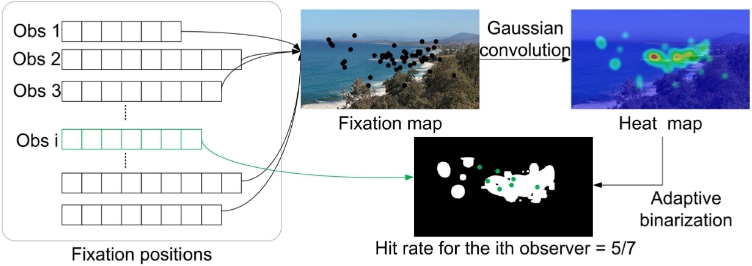
2. Predicting this value with a computational model

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]:

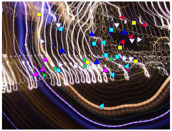
one-against-all approach



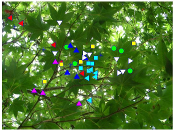
- $IOVC = 1$, strong congruency between observers.
- $IOVC = 0$, lowest congruency between observers.

Measuring the IOVC

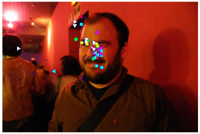
Examples based on MIT eye tracking database



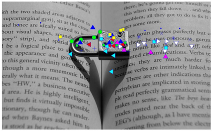
(a) 0.29



(b) 0.23

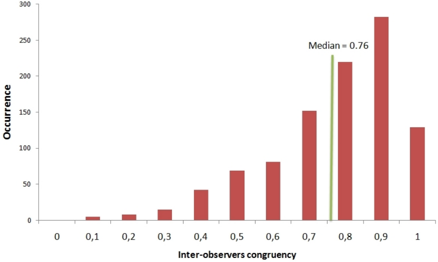


(c) 0.72



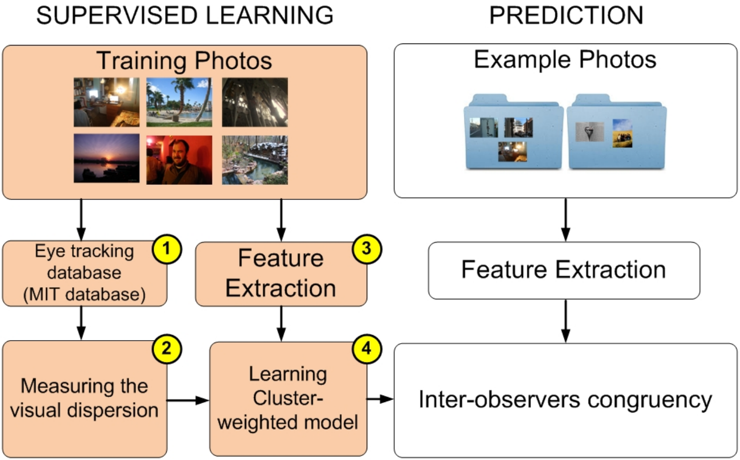
(d) 0.85

For the whole database:



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**

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

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

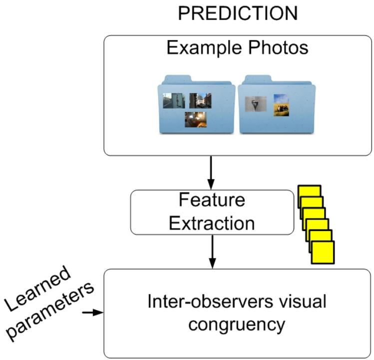
Feature extraction

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

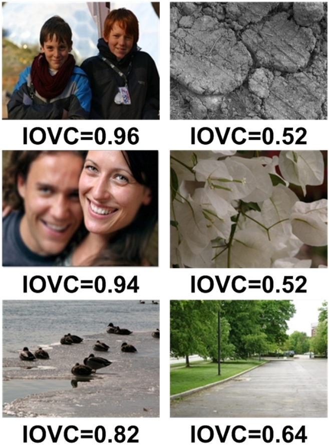
(a) Original (b) DoF=0.12 (c) Original (d) DoF=0.75

Feature extraction

After the learning, we can now predict:



Predicted IOVC:



Performance

Two tracking databases are used to evaluate the performance:

- MIT's dabase (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:

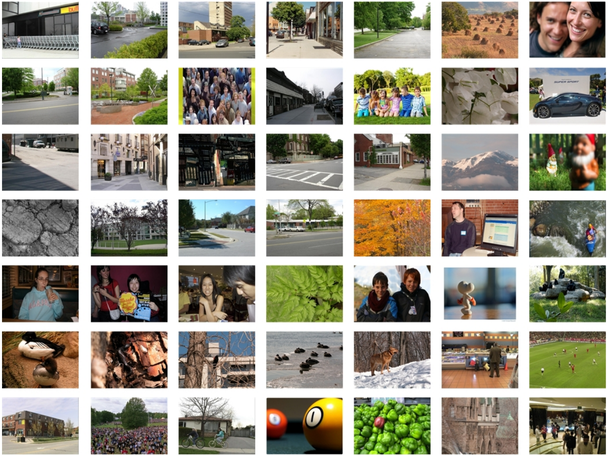
- MIT's dabase:
  $r(2004) = 0.34,\ p < 0.001$
- Le Meur's database:
  $r(54) = 0.24,\ p < 0.17$

(b) 0.89,0.94

(c) 0.91,0.91

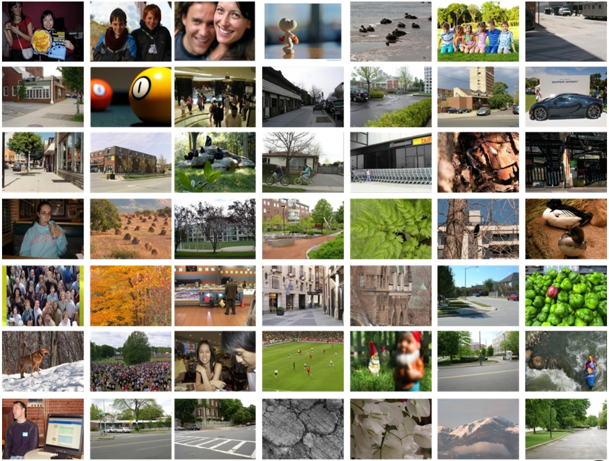
Application to image ranking

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

We do not perceive everything at once...

- Visual attention:

Overt	vs	Covert
Bottom-up	vs	Top-Down

- Computational models of visual attention:

  Biologically plausible Probabilistic beyond Bottom-up

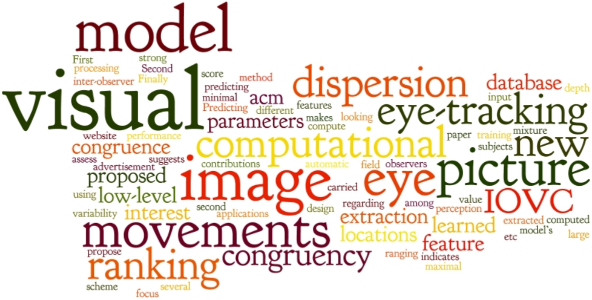
- Performances:

  between a set of fixations and a saliency map
  between two saliency maps

- Saliency-based applications.

Conclusion

Thanks for your attention



A special session has been accepted to WIAMIS'2013 (Paris):

O. Le Meur and Z. Liu, Visual attention, a multidisciplinary topic: from behavioral studies to computer vision applications

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.

[Baccino & Manunta, 2005] T. Baccino & Y. Manunta. Eye-fixation-related potentials: insight into parafoveal processing. Journal Of Psychophysiology, 19(3), pp. 204-215, 2005.

[Borji & Itti, 2012] A. Borji & L. Itti. State-of-the-art in Visual Attention Modeling, IEEE Transactions on Pattern Analysis and Machine Intelligence, In press, 2012.

ECOLE SUPERIEURE D'INGENIEUR DE RENNES

UNIVERSITÉ DE RENNES 1

[Bruce, 2004] N.D.B. Bruce. Image Analysis through local information measures. International Conference on Pattern Recognition, 2004.

[Bruce & Tsotsos, 2006] N.D.B. Bruce & J.K. Tsotsos. Saliency Based on Information Maximization. Advances in Neural Information Processing Systems, 18, pp. 155-162, June 2006.

[Bruce & Tsotsos, 2009] N.D.B. Bruce & J.K. Tsotsos. Saliency, attention and visual search: an information theoretic approach. Journal of Vision, 9(3), pp. 1-24, 2009.

[Bur et al., 2007] A. Bur, H. Hügli. Dynamic visual attention: competitive versus motion priority scheme. Proc. ICVS Workshop on Computational Attention & Applications, 2007.

[Chamaret & Le Meur, 2008] C. Chamaret & O. Le Meur. Attention-based video reframing: validation using eye-tracking. ICPR, 2008.

[Fan et al., 2003] X. Fan, X. Xie, W. Ma, H. Zhang, H. Zhou. Visual attention based image browsing on mobile devices, ICME, pp. 53-56, 2003.

[Fecteau et al., 2006] J.H. Fecteau, D. P. Munoz. Salience, relevance, and firing: a priority map for target selection, Trends in cognitive sciences, 10, pp. 382-390, 2006.

[Findlay & Gilchrist, 2003] J.M. Findlay & I. D. Gilchrist. Active vision: the psychology of looking and seeing. Oxford: Oxford University Press, 2003.

[Follet et al., 2009] B. Follet, O. Le Meur & T. Baccino. Relationship between coarse-to-fine process and ambient-to-focal visual. ECEM, 2009.

[Foulsham et al., 2009] T. Foulsham, A. Kingstone and G. Underwood, Turning the world around: Patterns in saccade direction vary with picture orientation, Vision Research, vol. 48, pp. 1777-1790, 2008.

[Gao et al., 2008] D. Gao, V. Mahadevan and N. Vasconcelos. On the plausibility of the discriminant center-surround hypothesis for visual saliency, Journal of Vision, 8(7):13, pp. 1-18, 2008.

[Henderson et al., 2007] J.M. Henderson, J.R. Brockmole, M. S. Castelhano, M. van R. Mack, M. Fischer, W. Murray and R. Hill. Visual saliency does not account for eye movements during visual search in real-world scenes. Eye movements: A window on mind and brain. (pp. 537-562). Oxford, UK: Elsevier.

[Itti et al., 1998] L. Itti, C., Koch, E., Niebur. A model for saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, 20, pp. 1254-1259, 1998.

[Itti & Koch, 2000] L. Itti and C., Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, 40 (10-12), pp. 1489-1506, 2000.

[Itti & Baldi, 2006] L. Itti and P. F. Baldi. Bayesian Surprise Attracts Human Attention, In: Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005), pp. 547-554, Cambridge, MA:MIT Press, 2006.

[Judd et al., 2009] T. Judd, K. Ehinger, F. Durand and A. Torralba. Learning to Predict Where Humans Look, IEEE International Conference on Computer Vision (ICCV), 2009.

[Kusunoki et al., 2000] M., Kusunoki, J. Gottlieb, M.E. Goldberg. The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. Vision Research, 40(10), pp. 1459-1468, 2000.

[Koch et al., 1985] C. Koch, S. Ullman. shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4, pp. 219-227, 1985.

[Landragin, 2004] F. Landragin. Saillance physique et saillance cognitive. Cognition, Représentation, Langage (CORELA) 2(2), 2004.

[Le Meur et al., 2006] O. Le Meur, X. Castellan, P. Le Callet, D. Barba, Efficient saliency-based repurposing method, ICIP, 2006.

[Le Meur et al., 2006] O. Le Meur, P. Le Callet, D. Barba and D. Thoreau. A coherent computational approach to model the bottom-up visual attention. IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(5), pp. 802-817, 2006.

[Le Meur et al., 2007] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. Vision Research, 47(19), pp. 2483-2498, 2007.

[Le Meur & Chevet, 2010] O. Le Meur and J.C. Chevet. Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks, IEEE Trans. On Image Processing, vol. 19, No. 11, pp. 2801-2813, November 2010.

[Levenshtein, 1966] V.I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. Doklady Physics 10, pp. 707-710, 1966.

[Li, 2002] Z. Li. A saliency map in primary visual cortex. Trends Cognitive Sciences 6(1), pp. 9-16, 2002.

[Mancas et al., 2006] M. Mancas, C. Mancas-Thillou, B. Gosselin, B. Macq. A rarity-based visual attention map - application to texture description. ICIP, 2006.

[Mancas, 2009] M. Mancas. Relative Influence of Bottom-Up and Top-Down Attention. Attention in Cognitive Systems, Lecture Notes in Computer Science, 2009.

[Marat et al., 2009] S. Marat, T. Ho Phuoc, L. Granjon, N. Guyader, D. Pellerin, A. Guerin-Dugue. Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos, International Journal of Computer Vision, 82(3), pp. 231-243, 2009.

[Mannan et al., 1995] S. Mannan, K.H. Ruddock and D.S. Wooding. Automatic control of saccadic eye movements made in visual inspection of briefly presented 2D images. Spatial Vision, 9, pp. 363-386, 1995.

[McLeod et al., 1988] P. McLeod, C. Heywood, J. Driver and S. Zihl. Selective deficit of visual search in moving displays after extrastriate damage. Nature, 339, 466-467, 1988.

[Minsky, 1974] M. Minsky. A Framework for Representing Knowledge. MIT-AI Laboratory Memo 306, June, 1974.

[Nadenau, 2000] M. Nadenau. Integration of human color vision models into high quality image compression. PhD thesis, EPFL, 2000.

[Navalpakkam & Itti, 2005] V. Navalpakkam & L. Itti. Modeling the influence of task on attention. Vision Research, vol. 45, pp. 205-231, 2005.

[Oliva and Schyns, 2000] A. Oliva & P.G. Schyns. Colored diagnostic blobs mediate scene recognition. Cognitive Psychology, 2000.

[Oliva et al., 2003] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson. Top-down control of visual attention in object detection. ICIP, vol. 1, pp. 253-256, 2003.

[Oliva, 2004] A. Oliva. The gist of a scene. In the Encyclopedia of Neurobiology of Attention. L. Itti, G. Rees, and J.K. Tsotsos (Eds.), Elsevier, San Diego, CA (pages 251-256), 2004.

[Oliva, 2005] A. Oliva. Gist of a scene. In Neurobiology of Attention. Eds. L. Itti, G. Rees and J. Tsotsos. Academic Press, Elsevier, 2005.

ECOLE SUPERIEURE D'INGENIEUR DE RENNES

UNIVERSITÉ DE RENNES 1

[Pare and Munoz, 2001] M. Pare and D.P. Munoz. Expression of a recentering bias in saccade regulation by superior colliculus neurons. Experimental Brain Research, 137 pp. 354-368, 2001.

[Parkhurst et al.,2002] D. Parkhurst, K. Law, E. Niebur. Modelling the role of salience in the allocation of overt visual attention. Vision Research, 42(1), pp. 107-123, 2002.

[Peters et al., 2005] R. Peters, A. Iyer, L. Itti, C. Koch. Components of bottom-up gaze allocation in natural images. Vision Research, 2005.

[Peters & Itti, 2008] R. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. ACM Transactions on Applied Perception, Vol. 5, 2008.

[Posner, 1980] M. I. Posner. Orienting of Attention. Quarterly Journal of Experimental Psychology 32, 1, 3-25, 1980.

[Potter & Levy, 1969] M.C. Potter and E. I. Levy. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 81, 10-15, 1969.

[Salvucci et al.,2000] D.D. Salvucci, J.H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. ETRA, pp. 71-78, 2000.

[Schyns & Oliva, 1994] P.G. Schyns & A. Oliva. From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. Psychological Science, 5, 195-200, 1994.

[Tatler, 2007] B.W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision 7(14):4, 1-17, 2007.

[Thorpe et al., 1996] S. Thorpe, D. Fize, C. Marlot. Speed of processing in the human visual system. Nature 381:520-2, 1996.

[Treisman & Gelade, 1980] A. Treisman and G. Gelade. A feature integration theory of attention. Cognitive Psychology Vol 12, pp. 97-136, 1980.

[Treue et al., 2006] S. Treue, J. C. Martinez-Trujillo. Visual search and single-cell electrophysiology of attention: Area MT, from sensation to perception. Visual Cognition, 14(4), pp. 898-910, 2006.

[Tsotsos et al., 1995] J.K., Tsotsos, S. Culhane, W., Wai, Y., Lai, N., Davis, F., Nuflo. Modeling visual attention via selective tuning, Artificial Intelligence 78(1-2), pp. 507-547, 1995.

[Unema et al., 2005] P. Unema, S. Pannasch, M. Joos and B. Velichkovsky. Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. Visual Cognition, 12(3), pp. 473-494, 2005.

[Velichkovsky et al., 2002] B. M. Velichkovsky. Heterarchy of cognition: The depths and the highs of a framework for memory research. Memory, 10(5), pp. 405-419, 2002.

[Yarbus, 1967] A. L. Yarbus. Eye Movements and vision. New York: Plenum, 1967.

[Varela, 1998] F. Varela. http://www.larecherche.fr/content/recherche/article?id=18406, La Recherche, N. 308, pp.109-112, 1998.

[Wolfe, 1994] J.M. Wolfe. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1, 2, pp. 202-238, 1994.

[Zhang et al., 2008] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. Journal of Vision, vol. 8, no. 7, pp. 1-20, 12 2008.