

Methods for comparing scanpaths and saliency maps: strengths and weaknesses

O. Le Meur olemeur@irisa.fr

T. Baccino thierry.baccino@univ-paris8.fr

Univ. of Rennes 1

<http://www.irisa.fr/temics/staff/lemeur/>

July 2011

Introduction

Definition (Scanpath [Noton and Stark(1971)])

A scanpath is a particular sequence of eye movements when a particular visual pattern is viewed.

Visual scanpath is often held as a marker of attention



For the purpose of this presentation, we will consider a scanpath as being

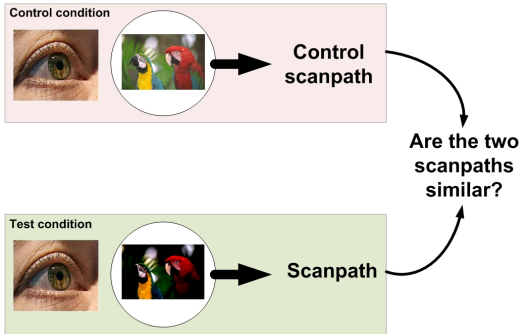
- any eye-movement data collected by eye-tracking apparatus,
- any path stemming from a computational model (saliency algo. with IOR [Koch and Ullman(1985)] for instance).

Introduction

The overall scanpath pattern is influenced and shaped by a combination of:

- ① Top-down cognitive factors (expectations, goals, memory...).
- ② Bottom-up processes involving visual sensory input.

Example: Impact of the visual quality on the deployment of visual attention



Different methods can be used to evaluate the similarity between scanpaths.

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps
- 5 Measuring a realistic upper-bound
- 6 Two populations of visual fixations?

Agenda

- 1 Introduction
- 2 **Methods involving two scanpaths**
 - String edit
 - Vector-based metric
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps
- 5 Measuring a realistic upper-bound
- 6 Two populations of visual fixations?

Three principal methods

These three methods have been described in the chapter proposal:

- String edit [Levenshtein(1966)];
- Mannan's metric [Mannan et al.(1995)];
- Vector-based metric [Jarodzka et al.(2010)].

Three principal methods

String edit-Levenshtein distance

Definition (String edit-Levenshtein distance [Levenshtein(1966)])

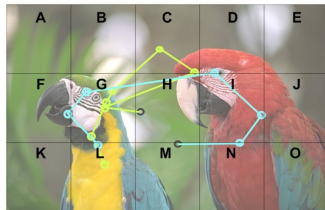
This technique was originally developed to account for the edit distance between two words. The similarity is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character.

Advantages:

- + Easy to compute
- + Keep the order of fixation

Drawbacks:

- How many viewing areas of interest should we use (7,12,15,25...)?
- It does not take into account fixation duration...



Parrot picture with a 5×3 grid overlaid

Three principal methods

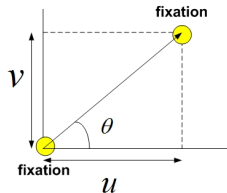
Vector-based metric (1/2)

Definition (Vector-based metric [Jarodzka et al.(2010)])

The vector-based metric represents the scanpath as a sequence of vector. For example, a scanpath with n fixations is represented by a set of $n - 1$ vectors.

This representation is interesting because it preserves:

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



Three principal methods

Vector-based metric (2/2)

The vector-based metric is composed of three steps:

- 1 Scanpath simplification:
 - small consecutive saccadic vectors are merged;
 - consecutive vectors having similar directions are merged.
- 2 Temporal alignment:
 - Similarity matrix M ;
 - Adjacency matrix A ;
 - Find the shortest path.
- 3 Scanpath comparison providing 5 measures:
 - difference in shape (vector difference);
 - difference in amplitude of saccade;
 - difference in spatial position;
 - difference in direction;
 - difference in duration.

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

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps**
 - From a fixation map to a saliency map
 - Divergence of Kullback-Leibler
 - ROC analysis
- 4 Methods involving scanpaths and saliency maps
- 5 Measuring a realistic upper-bound
- 6 Two populations of visual fixations?

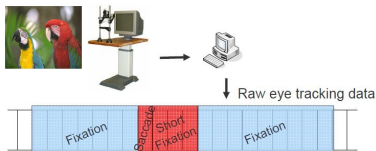
Three principal methods

These three methods have been described in the chapter proposal:

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

Three principal methods

From a fixation map to a saliency map

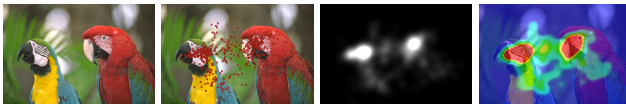


- Discrete fixation map f^i for the i th observer (M is the number of fixations):

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

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

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



(a) Original

(b) Fixation map

(c) Saliency map

(d) Heat map

Three principal methods

Divergence of Kullback-Leibler

Definition (Divergence of Kullback-Leibler)

The Kullback-Leibler divergence is used to estimate the overall dissimilarity between two probability density functions. Let define two discrete distributions R and P with probability density functions r_k and p_k , the KL-divergence between R and P is given by the relative entropy of P with respect to R :

$$KL(R, P) = \sum_k p_k \log \left(\frac{p_k}{r_k} \right) \quad (3)$$

The KL-divergence is only defined if r_k and p_k both sum to 1 and if $r_k > 0$ for any k such that $p_k > 0$.



(a)



(b)



(c)

$$KL(c, b) = 3.33 \text{ and } KL(b, c) = 7.06.$$

Advantages:

+ Easy to use

Drawbacks:

- Not bounded

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.

Experimental saliency map

		actual value		
		p	n	
Predicted saliency map	p'	True Positive	False Positive	P'
	n'	False Negative	True Negative	N'
prediction outcome		P	N	total

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

Three principal methods

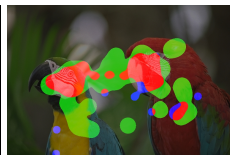
ROC analysis (2/2)



(a) Reference

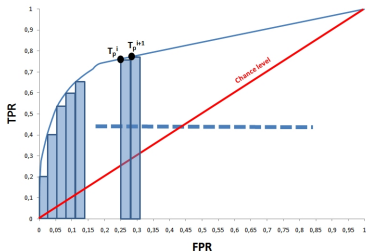


(b) Predicted



(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:

- ...

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps**
 - Receiver Operating Analysis
- 5 Measuring a realistic upper-bound
- 6 Two populations of visual fixations?

Four principal methods

These four methods have been described in the chapter proposal:

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

Four principal methods

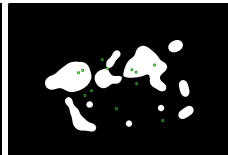
Receiver Operating Analysis (1/3)

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

- Human fixations only [Torralba et al.(2006), Judd et al.(2009)]:
 - In this case, the hit rate is measured in function of the threshold used to binarize the saliency map.



(a) HitRate=100%



(b) HitRate=50%

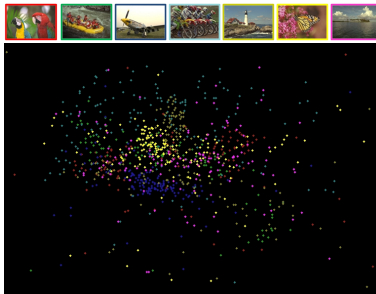
This method is not sensitive to the false alarm rate.

Four principal methods

Receiver Operating Analysis (2/3)

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

- Human fixations plus a set of control points
[Einhäuser and König(2003), Tatler et al.(2005)]:
 - by selecting the control points from a uniform or random distribution;



Four principal methods

Receiver Operating Analysis (3/3)

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

- Human fixations plus a set of control points
[Einhäuser and König(2003), Tatler et al.(2005)]:
 - by selecting locations randomly from a distribution of all fixation locations for that observer that occurred at the same time, but on other images.



This method accounts for center bias, same systematic tendency...
It underestimates the salience of areas which are more or less centered in the image...

Agenda

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps
- 5 **Measuring a realistic upper-bound**
- 6 Two populations of visual fixations?

Measuring a realistic upper-bound (1/3)

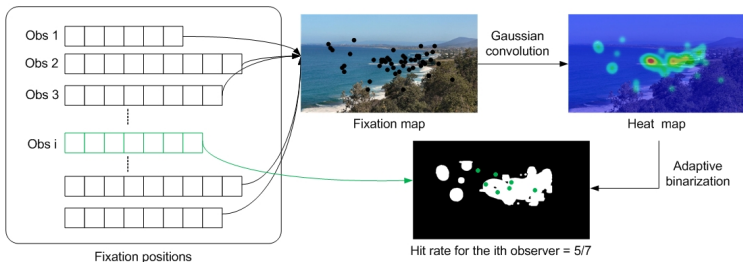
No saliency prediction can perform better than inter-observers dispersion.

Dispersion between observers:

- prior knowledge, experience, task, cultural difference...
- face, text, low-level visual features...

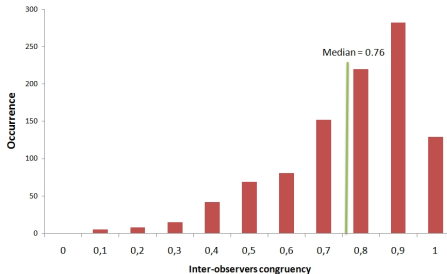
The dispersion can be evaluated by a one-against-all or leave one out.

Example: inter-observers congruency based on the Hit Rate metric [Torralba et al.(2006)]



Inter-observers congruency for Judd's database [Judd et al. (2009)]:

- 1000 pictures, 15 observers;
- congruency based on the hit rate.



Low
congruency



High
congruency

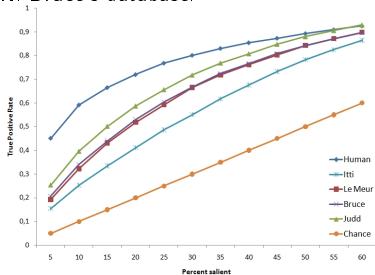
Measuring a realistic upper-bound (3/3)

The inter-observer dispersion can be used as:

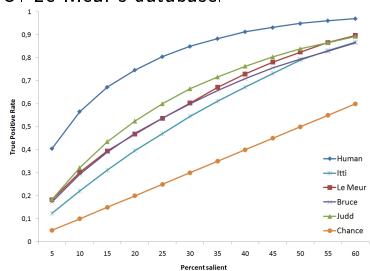
- to define the upper bound of a prediction
- to normalize the metric ($nAUC$ as proposed by [Zhao and Koch(2011)]).

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

N. Bruce's database:



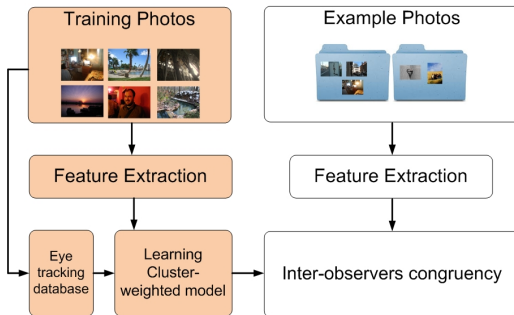
O. Le Meur's database:

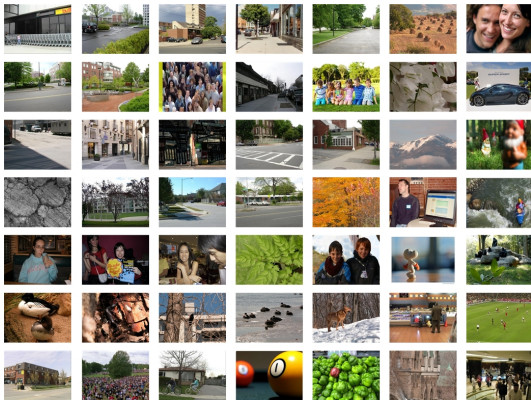


Predicting the dispersion between observers

There exist two computational models to predict the dispersion between observers:

- Visual Clutter [Rosenholtz et al.(2007)] based on entropy of Wavelet subbands;
- IOVC (inter-Observers Visual Congruency) [Le Meur et al.(2011)]:
 - Face detection;
 - Color Harmony;
 - Depth of Field;
 - Scene Complexity (entropy, number of regions, contours).





Pictures with the highest predicted congruency



Pictures with the lowest predicted congruency



Agenda

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps
- 5 Measuring a realistic upper-bound
- 6 **Two populations of visual fixations?**
 - Focal-ambient dichotomy

Two populations of visual fixations?

Recent findings about two distinct populations of fixations

- Velichkovsky and his colleagues
[Velichkovsky(2002), Unema et al.(2005), Pannasch et al.(2011)] conjointly analyzed the fixation duration with the subsequent saccade amplitude.
 - (short) Fixations with subsequent large-amplitude saccades ⇒ **Ambient mode**
 - (long) Fixations with subsequent small-amplitude saccades ⇒ **Focal mode**
 - Ad hoc threshold to classify the fixations (5 degrees).
 - Larger proportion of focal fixations
- Automatic classification of visual fixations based on K-means [Follet et al.(2011)]:
 - Two populations of fixation similar to previous studies
 - Automatic classification gives a threshold of 6 degrees
 - 70% of focal fixations and 30% of ambient fixations

Two populations of visual fixations?

Automatic classification of visual fixations based on K-means [Follet et al.(2011)]



(a) Focal



(b) Ambient



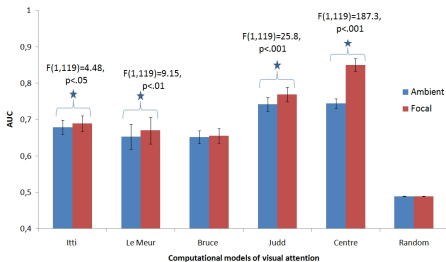
(c) Focal



(d) Ambient

Focal and Ambient fixation-density maps

Is there a correlation between model-predicted saliency and these maps?



- Both are correlated to model-predicted saliency;
- Focal maps are more bottom-up than ambient ones;
- Ambient maps are less correlated to center map.

Agenda

- 1 Introduction
- 2 Methods involving two scanpaths
- 3 Methods involving two saliency maps
- 4 Methods involving scanpaths and saliency maps
- 5 Measuring a realistic upper-bound
- 6 Two populations of visual fixations?

Introduction

Methods involving two scanpaths

Methods involving two saliency maps

Methods involving scanpaths and saliency maps

Measuring a realistic upper-bound

Two populations of visual fixations?

Conclusion

References



Einhauser, W., König, P., 2003.

Does luminance-contrast contribute to a saliency for overt visual attention?
European Journal of Neuroscience 17, 1089–1097.



Follet, B., Le Meur, O., Baccino, T., 2011.

Features of ambient and focal fixations on natural visual scenes, in: ECEM.



Itti, L., Baldi, P., 2005.

Bayesian surprise attracts human attention, in: Cambridge, M.M.p. (Ed.), Advances in Neural Information Processing Systems, pp. 1–8.



Jarodzka, H., Holmqvist, K., Nyst, M., 2010.

A vector-based, multidimensional scanpath similarity measure, in: Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications.



Judd, T., Ehinger, K., Durand, F., Torralba, A., 2009.

Learning to predict where people look, in: ICCV.



Koch, C., Ullman, S., 1985.

Shifts in selective visual attention: towards the underlying neural circuitry.
Human Neurobiology 4, 219–227.



Le Meur, O., Baccino, T., Roumy, A., 2011.

Prediction of the Inter-Observers Visual Congruency (IOVC) and application to image ranking.
ACM TO BE PUBLISHED xx, xx–xx.



Levenshtein, 1966.

Binary codes capable of correcting deletions, insertions, and reversals.
Soviet Physics Doklady 10, 707–10.



Mannan, S., Ruddock, K.H., Wooding, D.S., 1995.

Automatic control of saccadic eye movements made in visual inspection of briefly presented 2D images.
Spatial Vision 9, 363–386.



Noton, D., Stark, L., 1971.

Scanpaths in saccadic eye movements while viewing and recognizing patterns.
Vision Research 11, 929–942.



Pannasch, S., Schulz, J., Velichkovsky, B., 2011.

On the control of visual fixation durations in free viewing of complex images.
Attention, Perception & Psychophysics 73, 1120–1132.



Parkhurst, D., Law, K., Niebur, E., 2002.

Modeling the role of salience in the allocation of overt visual attention.
Vision Research 42, 107–123.



Peters, R., Itti, L., 2008.

Applying computational tools to predict gaze direction in interactive visual environments.
ACM Transactions on Applied Perception 5.



Peters, R.J., Iyer, A., Itti, L., Koch, C., 2005.

Components of bottom-up gaze allocation in natural images.
Vision Research 45, 2397–2416.



Rosenholtz, R., Li, Y., Nakano, L., 2007.

Measuring visual clutter.
Journal of Vision 7, 1–22.



Tatler, B.W., Baddeley, R.J., Gilchrist, I.D., 2005.

Visual correlates of fixation selection: effects of scale and time.
Vision Research 45, 643–659.



Torralba, A., Oliva, A., Castelano, M., Henderson, J., 2006.

Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search.

Psychological review 113, 766–786.



Unema, P., Pannasch, S., Joos, M., Velichkovsky, B.M., 2005.

Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration.

Visual Cognition 12, 473–494.



Velichkovsky, B., 2002.

Heterarchy of cognition: The depths and the highs of a framework for memory research.

Memory 10, 405–419.



Zhao, Q., Koch, C., 2011.

Learning a saliency map using fixated locations in natural scenes.

Journal of Vision 11, 1–15.