Saliency and Human Fixations: State-of-the-art and Study of Comparison Metrics

Nicolas Riche, Matthieu Duvinage, Matei Mancas, Bernard Gosselin and Thierry Dutoit
TCTS Lab, University of Mons (UMONS)
20, Place du Parc, 7000, Mons, Belgium
{firstname.surname}@umons.ac.be

Abstract

Visual saliency has been an increasingly active research area in the last ten years with dozens of saliency models recently published. Nowadays, one of the big challenges in the field is to find a way to fairly evaluate all of these models. In this paper, on human eye fixations ,we compare the ranking of 12 state-of-the art saliency models using 12 similarity metrics. The comparison is done on Jian Li's database containing several hundreds of natural images. Based on Kendall concordance coefficient, it is shown that some of the metrics are strongly correlated leading to a redundancy in the performance metrics reported in the available benchmarks. On the other hand, other metrics provide a more diverse picture of models' overall performance. As a recommendation, three similarity metrics should be used to obtain a complete point of view of saliency model performance.

1. Introduction

In the field of computer vision, a wide variety of models that aim at mimicking the visual attention cognitive process exists [15] [30]. By outputing saliency maps that estimate the probability of each image area to grab our attention, those models allow to automatically predict the most relevant regions from images. In practice, rare, novel or surprising information is considered as salient.

Since the early 2000s, an increasing amount of saliency models have been proposed mainly splitting into two approaches. Some of them focus on salient object detection while others deal with predicting human eye fixations. In terms of validation, the first category uses one gold standard: Precision/Recall/F-measure metrics while the other category uses a lot of different metrics.

Due to the diversity of available metrics for eye fixation prediction assessment, several benchmarks were proposed. In 2011, Toets proposed in [27] to compare saliency models based on the Spearman's rank correlation coefficient. In 2012, Borji built a benchmark [5] where three evaluation scores (CC, NSS and AUC) are used. Finally, Judd [14] proposed a platform using three different metrics: AUC-Judd, a Similarity metric, and the Earth Mover's Distance (EMD). Lemeur in [16] also reported about methods for comparing scanpaths and saliency maps. Although these benchmarks are major contributions, none of those studies deeply discussed the relevance of their similarity metrics mix.

Therefore, the redundancy of these metrics is discussed in the paper which is organized as follows. Section 2 contains a review of all similarity metrics based on human eye fixations. Section 3 describes the methods and experiments used to study metrics based on Kendall concordance scores.The results are presented in Section 4 while section 5 provides a discussion.

2. Literature Review of Similarity Metrics

In this section, all the similarity metrics that have been used to assess saliency models are presented. We propose here a two-dimensional taxonomy to classify all similarity metrics. The first dimension is related to the nature of the similarity metric and can be divided into three categories: "value-based", "location-based" and "distribution-based" (Tab. 1). On a second dimension, among those metrics, some are general metrics which were not specifically built for saliency models and are called "common metrics", others are existing similarity measures which were adapted to the field of saliency model evaluation and can be called "hybrid metrics" and finally some metrics were specifically built for saliency models assessment and are called "specific metrics" in our taxonomy (Tab. 1). In the next sections, the similarity metrics are briefly described within the first three categories of the proposed taxonomy, namely value-based, location-based and distribution-based.

	Value-based Metrics	Location-based Metrics	Distribution-based Metrics
Common Metrics		AUC-Judd	KL-DiV / CC / Spear / Similarity
Hybrid Metrics		AUC-Zhao / AUC-Li	EMD
Specific Metrics	NSS / Percentile / Pf	AUC-Borji	

Table 1. Proposed two-dimensional taxonomy for visual salience metrics

2.1. Value-based metrics: focus on saliency map values at eye gaze positions

This first category of metrics compares the saliency amplitudes with the corresponding eye fixations maps.

2.1.1 Normalized Scanpath Saliency (NSS)

The Normalized Scanpath Saliency (NSS) metric was introduced in 2005 by Peeters and Itti [23]. The idea is to quantify the saliency map values at the eye fixation locations and to normalize it with the saliency map variance:

$$NSS(p) = \frac{SM(p) - \mu_{SM}}{\sigma_{SM}} \quad (1)$$

where p is the location of one fixation and SM is the saliency map which is normalized to have a zero mean and unit standard deviation. Indeed, the NSS score should be decreased if the saliency map variance is important or if all values are globally similar (small difference between fixation values and mean) because it shows that the saliency model will not be very predictive, while he will precisely point a direction of interest if the variance is small or the difference between fixation values and mean high.
The NSS score is the average of $NSS(p)$ for all fixations:

$$NSS = \frac{1}{N} * \sum_{p=1}^{N} NSS(p) \quad (2)$$

where N is the total number of eye fixations.

2.1.2 Percentile

The percentile metric is, for each pixel p on the eye fixation map, a ratio between the number of pixels in the saliency map with values smaller than the one corresponding to pixel p from the eye fixation map and the total number of pixels (Eq.3). This metric was defined by Peeters and Itti [22].

$$P(p) = \frac{|x \in X : SM(x) < SM(p)|}{|SM|} \quad (3)$$

where X is the set of all pixels of the saliency map SM, p is the location of one eye fixation and $|SM|$ indicates the total number of pixels. Similar to NSS, the global percentile score is the average of $P(p)$ for all the eye fixations.

2.1.3 Percentage of fixations into the salient region (Pf)

This metric was introduced by Torralba [29]. It aims at measuring the percentage of fixations into the salient region. First, saliency maps are thresholded at T = 0.8 where the saliency is normalized between 0 and 1. The threshold is set so that the selected image region occupies a fixed proportion of the image size. Second, the percentage of fixations in this area is computed and called Pf.

2.2. Distribution-based metrics: focus on saliency and gaze statistical distributions

In the literature, there are similarity and dissimilarity-based metrics between two distributions. Here, two dissimilarity and three similarity metrics are described.

2.2.1 Kullback-Leibler Divergence (KL-Div)

The Kullback-Leibler divergence is a commonly used metric to estimate an overall dissimilarity between two distributions. Many authors like [24], [26] and [17] already used this metric to compare saliency maps with human eyes fixations. The KL-Div is a measure of the information lost when the saliency maps probability distribution (called SM) is used to approximate the human eye fixation map probability distribution (called FM).

$$KL_{div} = \sum_{x=1}^{X} FM(x) * log(\frac{FM(x)}{SM(x) + \epsilon} + \epsilon) \quad (4)$$

where X is the number of pixels and ϵ is a small constant to avoid log and division by zero. SM and FM distributions are both normalized as in Eq.5 and 6.

$$SM(x) = \frac{SM(x)}{\sum_{x=1}^{X} SM(x) + \epsilon} \quad (5)$$

$$FM(x) = \frac{FM(x)}{\sum_{x=1}^{X} FM(x) + \epsilon} \quad (6)$$

when the two maps are strictly equal, the KL-divergence value is zero.

2.2.2 Earth Mover's Distance (EMD)

Earth Mover's Distance metric is a measure of the distance between two probability distributions over a region. Judd

[14] used this metric in her benchmark which is now available online. She uses a fast implementation of EMD provided by Pele and Werman [20] [21], but without a threshold. It computes the minimal cost to transform the probability distribution of the saliency maps SM into the the one of the human eye fixations FM.

$$EMD = (\min_{f_{ij}} \sum_{i,j} f_{ij} d_{ij}) + |\sum_i FM_i - \sum_j SM_j| \max_{i,j} d_{ij}$$

$$s.t. f_{ij} \geq 0, \sum_j f_{ij} \leq FM_i, \sum_i f_{ij} \leq SM_j, \quad (7)$$

and

$$\sum_{i,j} f_{ij} = \min(\sum_i FM_i - \sum_j SM_j)$$

where each f_{ij} represents the amount transported from the i_{th} supply to the j_{th} demand. d_{ij} is the ground distance between bin i and bin j in the distribution. Starting from zero, a larger EMD indicates a larger overall difference between the two distributions.

2.2.3 Linear Correlation Coefficient (CC)

The linear correlation coefficient also named Pearson correlation coefficient is used by some authors like [19] or [17]. The linear CC output range is between -1 and 1. When the correlation value is close to -1 or 1, there is almost a perfect linear relationship between the two variables:

$$CC = \frac{cov(SM, FM)}{\sigma_{SM} * \sigma_{FM}} \quad (8)$$

2.2.4 Spearman's rank correlation coefficient

The Spearman's rank correlation coefficient metric [27] is defined as the CC metric (Eq. 8) but on ranked variables. This can be viewed as a non linear correlation. Toets uses this metrics in [27] to evaluate 13 models.

2.2.5 Similarity metric

The similarity metric [14] also uses the normalized probability distributions of the saliency map SM and human eye fixation map FM. The similarity is the sum of the minimum values at each point in the distributions. Mathematically, the similarity between two maps SM and FM is:

$$S = \sum_{x=1}^{X} min(SM(x), FM(x)) \quad (9)$$

where

$$\sum_{x=1}^{X} SM(x) = \sum_{x=1}^{X} FM(x) = 1. \quad (10)$$

A similarity score of one indicates that the distributions are the same. A similarity score of zero indicates that they do not overlap at all and are completely different.

2.3. Location-based Metrics: focus on location of salient regions at gaze positions

Location-based metrics are based on the notion of AUROC (Area under the Receiver Operating Characteristic curve) coming from signal detection theory. Here, four main different implementations are available dealing with some limitations of the classical approach.

2.3.1 Judd implementation (AUC-Judd)

In [14], Judd proposed a *classical* AUC. First, fixations pixels were counted once and the same number of random pixels are extracted from the saliency map. For one given threshold, saliency pixels can be treated as a classifier, with all points above threshold indicated as 'fixation' and all points below threshold as 'background' For any particular value of the threshold, there is some fraction of the actual fixation points which are labelled as True Positives (TP), and some fraction of points which were not fixation but labelled as False Positive (FP). This operation is repeated one hundred times. Then the ROC curve can be drawn and the Area Under the Curve (AUC) computed. An ideal score is one while random classiffication provides 0.5.

2.3.2 Zhao implementation (AUC-Zhao)

In [32], Zhao used a normalized AUC (nAUC). The idea is that no saliency algorithm can perform better (on average) than the area under the ROC curve dictated by inter-subject variability for each image. Zhao computes an ideal AUC by measuring how well the human fixations of one subject can be predicted by those of the other $n-1$ subjects, iterating over all n subjects and averaging the result. Finally, the AUC of the saliency map is normalized by this ideal AUC.

2.3.3 Borji implementation (AUC Borji)

In [5], Borji applied to saliency maps validation a suitable AUC metric called *shuffled* AUC. In *classical* AUC, saliency map values from random points from the image are addressed to create a binary mask. In the *shuffled* AUC metric, saliency values and fixations from another image (instead of random) of the same dataset are taken into account. In that way, the more or less centered distribution of the human fixations of the database is taken into account. This point is important because the AUROC scores can dramatically increase if a saliency map is weighted by a centred Gaussian. Indeed, human eye fixations are rarely near the edges of general test images and the amateur photographer often places salient objects in the image centre.

2.3.4 Li implementation (AUC-Li)

In [18], Li set the border cuts for all models to be of equal size and avoids in that way to artificially increase the AUC scores. This allows a fair comparisons between models which already do this pre-processing in comparisons with those which do not. The border cut post-processing affecting the fairness during the assessment is thus eliminated.

3. Experimental Setup

The 12 metrics presented in section 2 have two inputs: the saliency map SM and the human fixation maps FM. The output is a similarity (or dissimilarity) score (scalar). The purpose of this section is to introduce these two issues as well as the design of the study.

3.1. Database and Eye Fixation Maps

The human eye fixation maps used in this paper are those in the database published by Li [18]. This database provides eye fixation ground truth (collected with an eye tracker) for 235 colour images, of a size of 480 x 640 pixels. The fixation data was recorded by the eye tracker Tobii T60 on around 20 viewers. The images are quite different with objects of interest of different sizes (small, medium and large) to avoid a size-based bias of the saliency models.

3.2. Saliency maps from 12 models

Twelve state-of-the-art models are used to obtain different saliency maps in this study. They represent the updated version of the online available models from Borji's review paper [4] where models are sorted based on their mechanism to obtain saliency map. So, we use a wide method-based range of recently published saliency models.
Itti's model [12] represents the cognitive approach. SUN [31] and Torralba [28] are Baysian models. AIM [6], DVA [10] and RARE [25] are into the information theoretic category. SR [9], PFT [8], PQFT [8] and Achanta [1] use a spectral analysis approach to compute their saliency map. Two additional models are used: HFT [18] and AWS [7]. AWS predicts very well human gaze into different databases with various metrics. HFT is also an efficient model which is proposed by the database author Jian Li [18].

3.3. Analysis method

The goal of the experiment part is twofold. First, it shows which metrics are close to each other. Second, it intends to reduce the dimensionality of the used metrics and see which metrics should be used to do an efficient benchmark. Indeed, it is important to decide which metrics should be used together because they are complementary and which ones it is useless to compute together because they will provide redundant information.

For this purpose, the 12 models from section 3.2 produce a saliency map for each of the 235 images of the database and these saliency maps are compared with the corresponding human eye fixation map. The comparison is achieved using all the 12 previously described metrics (section 2). A mean score can be computed on the whole database for each model using the different metrics which leads to 12 different rankings of the 12 models for each comparison metric.

During the following study, we will use the ranking between models and not their mean score values. This is due to the fact that the output of the metrics can be very different in terms of range of score value and some of them should be maximized (correlation measures) while others should be minimized (divergence measures). Therefore, a direct score value comparison does not make a lot of sense. By contrast, the relative rank of the different models is a consistent measure common to all metrics and its range is here between 1 and 12 (from the best model to the weakest).

To compare models rank according to the different metrics, Kendall's W concordance measure [11] is used (as defined as Eq. 11)

$$W = \frac{12 * S}{m^2 * (n^2 n)} \tag{11}$$

where n is the number of models and m, the number of metrics. So here $m = n = 12$. S, the sum of squared deviations, is defined as in Eq. 12:

$$S = \sum_{i=1}^{n} (R_i - \bar{R})^2. \tag{12}$$

where R_i is the total rank given to model i and $\bar{R}$ the mean value of these total ranks.

Kendall's W concordance is a coefficient measuring the degree of agreement between metrics. The value ranges from 0 (no agreement between model ranks) to 1 (full agreement, same models ranking). Furthermore, some rules of thumb are provided [11] to allow the researcher to interpret this measure as depicted in Tab. 2.

Table 2. Interpretation of Kendall's W coefficient

W	Interpretation	Rank Confidence
0.5	Moderate agreement	Fair
0.7	Strong agreement	High
0.9	Unusually strong agreement	Very High
1	Complete agreement	Very High

In our study, the ranking range of 1 to 12 is small therefore, higher threholds are required to keep on the interpretation. That is why we decided to be much more selective than in Tab. 2: we interpret the Kendall coefficient as in Tab. 3. Indeed this interpretation shows that W=0.98 means that only 2 or 3 models are switched between the rankings.

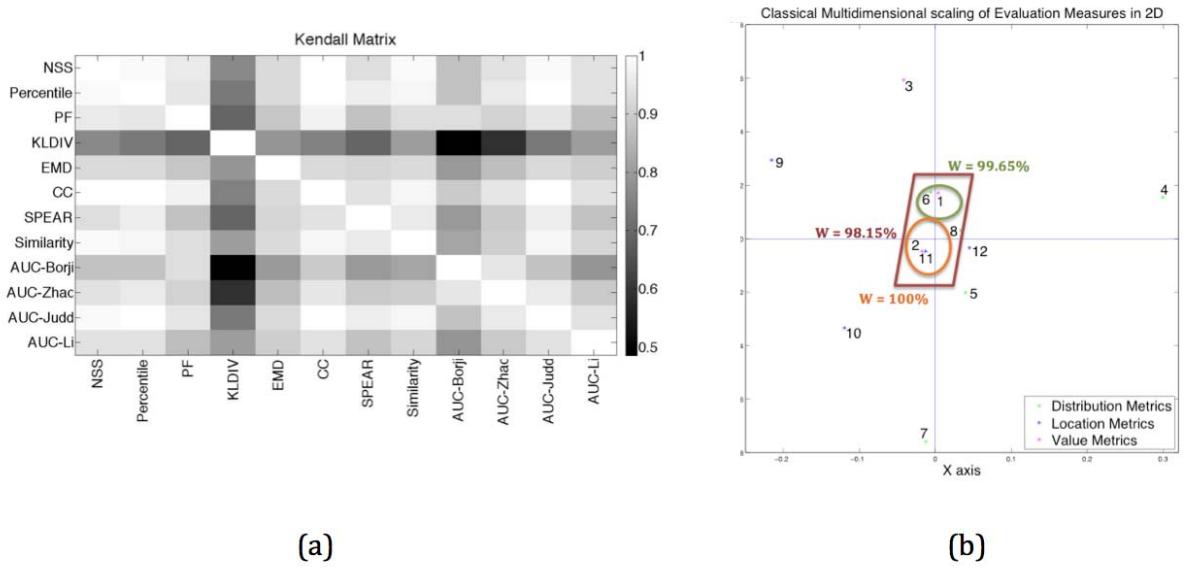

Figure 1. Kendall's analysis. (a) Kendall's Matrix on the 12 metrics. (b) Kendall's Measure on group of metrics with Classical Multidimensional scaling of Evaluation Measures in 2D: 1. NSS / 2. Percentile / 3. PF / 4. KL-DiV / 5. EMD / 6. CC / 7. SPEAR / 8. Similarity / 9. AUC-Borji / 10. AUC-Zhao / 11. AUC-Judd / 12. AUC-Li.

Table 3. Interpretation of Kendall's W coefficient on mean scores

W	Interpretation	Rank Confidence
0.7	Moderate agreement	Fair
0.85	Strong agreement	High
0.93	Very Strong agreement	High
0.98	Unusually strong agreement with 2 or 3 switched models	Very High
0.99	Unusually strong agreement with 1 or 2 switched models	Very High
1	Complete agreement	Very High

Table 4. Kendall's W coefficient of Intra-Group Metrics

Group of Metrics	W
Amplitude	.9534
Distribution	.7869
Location	.8488

4. Experimental Results

This section is organized as follows. Experiment 1 analyses the concordance between metrics. The second reduces the metrics dimensionality. The third uses previous results to provide a fair assessment of the 12 saliency models.

4.1. Experiment 1: Analysis of metrics consistency

4.1.1 Intra-Group Metrics

The concordance is computed between all metrics into the three categories: value-based (amplitude), location-based and distribution-based metrics (Tab. 4).

The concordance shows a moderate agreement for location-based and distribution-based metrics. This means that these metrics provide some complementary information: they might provide different results for the same saliency map, thus one of those metrics cannot just be ignored without a possible information loose about model ranking. However, one can see that the concordance between the amplitude metrics is high, which means that those measures are close and can therefore be summarized by a small subset of value-based metrics.

4.1.2 Inter-Group Metrics

Contrary to the intra-group study that does not achieve enough concordance, the inter-group suggests that some metrics are very close as it is shown in the Kendall matrix of Fig. 1 (a). NSS, Percentile, Correlation and AUC-Judd seem to be very close. On the opposite side, the KL-DiV metric seems like an outlier in this matrix and it is different from most of the other metrics in terms of model ranking. This result is explained by the KL-Div measure itself which is physically different from the others as it does not take space into account but only the map statistical distributions.

To provide a better representation of the proximity in terms of model ranking among metrics, we apply, on Kendall's coefficient, a classical multidimensional scaling

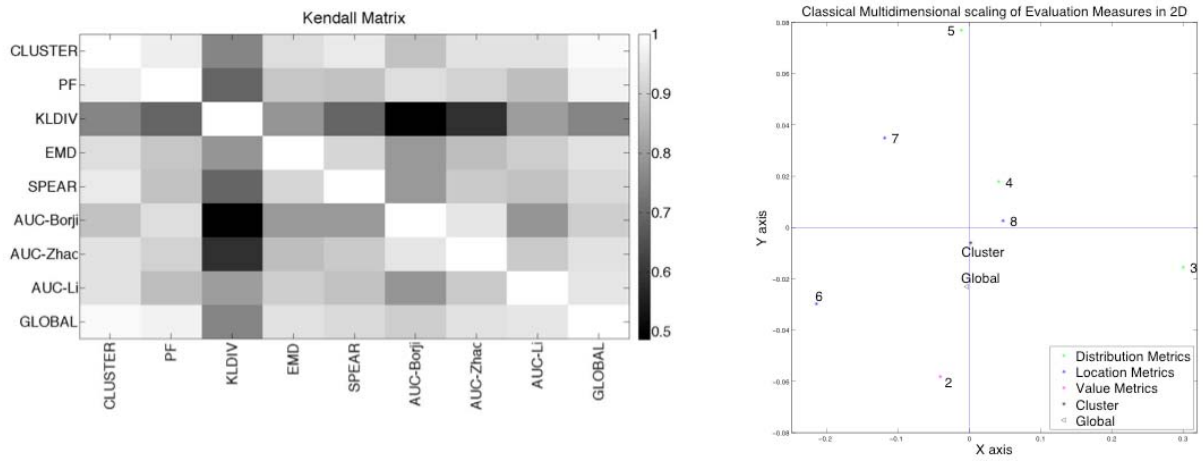

Figure 2. Kendall's analysis. (a) Kendall's Matrix on cluster, global and 7 metrics. (b) Kendall's Measure on group of metrics with Classical Multidimensional scaling of Evaluation Measures in 2D: 2. PF / 3. KL-DiV / 4. EMD / 5. SPEAR / 6. AUC-Borji / 7. AUC-Zhao / 8. AUC-Li

(CMDS). Also known as Principal Coordinates Analysis, CMDS uses statistical techniques to visualize and explore similarities or dissimilarities in data. The results are displayed in Fig. 1 (b). In this representation, X axis (equivalent to a first eigenvector) is more important than Y axis (equivalent to a second eigenvector). From the figure, one can see, for example, that PF and NSS are closer than PF and AUC-Li. This representation will be used in the next experiment.

4.2. Experiment 2: Study of the dimensionality

Based on the representation of Fig. 1, we decide to use a concordance of 98 % as a threshold to fuse metrics (in terms of rank). This threshold means that only the rank of 2 or 3 couples of models can be inverted on the 12 models which means that the differences in terms of classification of the saliency models are really minor. By using this threshold, five metrics can be fused to create a new metric called $Cluster$. In Fig. 1 b), the concordance of the metrics contained into the $Cluster$ is of 98.15 %. The ranking of $Cluster$ is defined as the mean ranking of all the metrics composing it.

For model validation, this $Cluster$ means that one measure from those included in this set is enough and the computation of the others inside this $Cluster$ is useless in terms of new information about models ranking. In this case the five metrics can be summarized well enough by any of them.

4.3. Experiment 3: Model assessment

A $Global$ metric is also computed as the mean of the ranking of all metrics. Kendall's matrix and a classical multidimensional scaling is again applied on the reduced number of metrics: $Cluster$ replaces its 5 redundant metrics. $Global$ is also added and acts like all metrics' barycentre.

In the last part of the experiment, Tab. 5 shows an assessment of each of the 12 saliency models with different metrics. The resulting $Cluster$ and the $Global$ metrics are included. One can see the rankings of the different saliency models from the best (1) to the weakest (12) on the 8 remaining metrics after dimensionality reduction. The $Global$ metric is the final ranking result after mixing all the remaining metrics (rank mean between the different rankings). Tab. 5 shows that the best saliency models in predicting human eye-tracking on Li's database are RARE, AWS and PQFT.

5. Discussion

Our approach is able to fairly compare saliency models rankings and not values, thus it is not possible here to test if the differences between the saliency models are statistically significant. For example, we observe that RARE is better than AWS, but we do not know if RARE is significantly better than AWS. To obtain this information, one should go back to the metrics values.

Our study also shows that one metric is not enough to evaluate the saliency model ranking on eye fixation data.

Metrics	Ranks											
	1	2	3	4	5	6	7	8	9	10	11	12
CLUSTER	RARE	PQFT	AWS	DVA	ITTI	AIM	TOR	PFT-SUN		SR	HFT	FT
Pf	AWS	RARE	PQFT	DVA	AIM	SUN	TOR	ITTI	PFT	SR	FT	HFT
KL-DiV	DVA	RARE	PQFT	HFT	AWS	SR	PFT	ITTI	AIM	FT	SUN	TOR
EMD	RARE	PQFT	AIM	DVA	AWS	TOR	ITTI	HFT	SR	PFT	SUN	FT
SPEAR	RARE	PQFT	DVA	AWS	ITTI	AIM	PFT	TOR	SUN	SR	HFT	FT
AUC-Borji	AWS	RARE	AIM	SUN	TOR	PQFT	ITTI	DVA	PFT	SR	FT	HFT
AUC-Zhao	RARE	AWS	ITTI	AIM	PQFT	TOR	SUN	DVA	PFT	SR	HFT	FT
AUC-Li	RARE	AWS	PQFT	ITTI	DVA	HFT	TOR	AIM	SUN	PFT	SR	FT
GLOBAL	RARE	AWS	PQFT	DVA	AIM	ITTI	TOR	SUN	PFT	HFT	SR	FT

Table 5. Model assessment for $Cluster$, $Global$ and all metrics not included in $Cluster$. The 3 bold-marked similarity metrics are enough to provide a fair saliency models rank comparison in terms of similarity with human eye-tracking data.

The minimal set of similarity metrics which should be used is a) one of the metrics composing the $Cluster$, b) AUC-Borji and c) KL-Div. The use of those three metrics is enough to cover most of the space in Fig. 2 (b) and provide a fair ranking result (Tab. 5).

The $Global$ and $Cluster$ metrics are very close (Fig. 2) which means that $Cluster$ must be used to assess any saliency model. However, the $Cluster$ metric alone shows some limitations by giving unfair advantages to some models on others. To represent the $Cluster$ metric any of its component metrics namely Percentile or AUC-Judd or CC or NSS or Similarity metrics can be used (Fig. 1).

Another interesting metric is the KL-Div which is an outlier (Fig. 1) providing really complementary results. Nevertheless, despite the fact that a lot of authors already used it alone, the KL-Div is not sufficient by itself. Indeed, two similar distributions can have very different spatial configurations. Thus, we propose to use it along with location and amplitude-based metrics, which will also verify the spatial coherence between the maps. Of course, one interest of KL-divergence is robust to gaze position errors during acquisition or the exact pointing of objects (not location-dependent), which may vary a little between participants.

The third recommended metric is AUC-Borji which is complementary in terms of information to the two others by bringing comparison between the eye-tracking data and saliency maps peak locations. Moreover, this metric is known to eliminate the effect of centred Gaussians. As only some models use centred Gaussians, eliminating this effect provides a more fair comparison.

Therefore, some saliency model benchmarks existing online as Borji [3] or Judd [13] use partly redundant similarity measures. Indeed in the case of Borji benchmark the use of NSS and AUC-Borji makes a lot of sense, but the use of the third metric (CC) does not bring any additional valuable information. Concerning Judd benchmark, while the choice of AUC-Judd and EMD metrics make sense, the use of the Similarity metric is redundant with AUC-Judd. In Judd benchmark the final ranking is obtained by averaging the three metrics, while AUC-Judd and EMD should have bigger weights compared to Similarity.

6. Conclusions

In this paper, we reviewed 12 state-of-the-art similarity metrics for visual saliency models validation and compared them over Jian Li's human eye-tracking fixations database with 12 recently published saliency models. The comparison is based on the ranking between models using Kendall's W coefficient (PCA provides poor results).

The conclusion of our comparison study is that evaluating a saliency model with human fixations using only one similarity metric is not enough to be fair. On the 12 similarity metrics, 5 measures can be grouped together in a single measure $Cluster$ which can be adequately summarized by any of them. In addition to $Cluster$, we suggest that 2 others metrics (AUC-Borji and KL-Div) with complementary interpretation should be used to fairly compare saliency models based on human eye-tracking data.

An implementation of the codes used in this paper to fairly assess his own saliency models on Li database is provided online [2] in the website project section.

7. Acknowledgments

M. Mancas is funded by the Walloon Region Predattor project. M. Duvinage is a FNRS (Fonds National de la Recherche Scientifique) Research Fellow and the corresponding author for experiment design, statistic based contribution and interpretation.

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s).

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *2009, IEEE Conference on CVPR*, pages 1597–1604. IEEE, 2009.

[2] U. Attention Group. Computational Attention Website. `http://tcts.fpms.ac.be/attention`.

[3] A. Borji. Evaluation measures for saliency maps. `https://sites.google.com/site/saliencyevaluation/home`.

[4] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on PAMI*, 2012.

[5] A. Borji, D. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE TIP*, 2012.

[6] N. Bruce and J. Tsotsos. Saliency based on information maximization. *Advances in neural information processing systems*, 18:155, 2006.

[7] A. Garcia-Diaz, X. Fdez-Vidal, X. Pardo, and R. Dosil. Decorrelation and distinctiveness provide with human-like saliency. In *Advanced Concepts for Intelligent Vision Systems*, pages 343–354. Springer, 2009.

[8] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *2008, IEEE Conference on CVPR*, pages 1–8. IEEE, 2008.

[9] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *2007, IEEE Conference on CVPR*, pages 1–8. IEEE, 2007.

[10] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. *Advances in neural information processing systems*, 21:681–688, 2008.

[11] D. C. Howell. *Statistical methods for psychology*. Wadsworth Publishing Company, 2012.

[12] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on PAMI*, 20(11):1254–1259, 1998.

[13] T. Judd. saliency benchmark. `http://people.csail.mit.edu/tjudd/SaliencyBenchmark/`.

[14] T. Judd, F. Durand, and A. Torralba. A benchmark of computational models of saliency to predict human fixations. *MIT tech report*, 2012.

[15] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–27, 1985.

[16] O. Le Meur and T. Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, pages 1–16, 2011.

[17] O. Le Meur, P. Le Callet, D. Barba, et al. Predicting visual fixations on video based on low-level visual features. *Vision research*, 47(19):2483–2498, 2007.

[18] J. Li, M. D. Levine, X. An, X. Xu, and H. He. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Transactions on PAMI*, 35:996–1010, 2012.

[19] N. Ouerhani, R. Von Wartburg, H. Hugli, and R. Muri. Empirical validation of the saliency-based model of visual attention. *Electronic letters on computer vision and image analysis*, 3(1):13–24, 2004.

[20] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *Computer Vision–ECCV 2008*, pages 495–508. Springer, 2008.

[21] O. Pele and M. Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th ICCV*, pages 460–467. IEEE, 2009.

[22] R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception (TAP)*, 5(2):9, 2008.

[23] R. J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. *Vision research*, 45(18):2397–2416, 2005.

[24] U. Rajashekar, L. K. Cormack, and A. C. Bovik. Point-of-gaze analysis reveals visual search strategies. In *Proceedings of SPIE*, volume 5292, pages 296–306, 2004.

[25] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit. Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642 – 658, 2013.

[26] B. W. Tatler, R. J. Baddeley, I. D. Gilchrist, et al. Visual correlates of fixation selection: Effects of scale and time. *Vision research*, 45(5):643–659, 2005.

[27] A. Toet. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE Transactions on PAMI*, 33(11):2131–2146, 2011.

[28] A. Torralba. Modeling global scene factors in attention. *JOSA A*, 20(7):1407–1418, 2003.

[29] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. 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, 2006.

[30] J. K. Tsotsos, S. M. Culhane, W. Y. Kei Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1):507–545, 1995.

[31] 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*, 8(7), 2008.

[32] Q. Zhao and C. Koch. Learning a saliency map using fixated locations in natural scenes. *Journal of vision*, 11(3), 2011.