

IMAGE COMPLEXITY MEASURE BASED ON VISUAL ATTENTION

Matthieur Perreira Da Silva, Vincent Courboulay, Pascal Estraillier

L3i, University of La Rochelle
Avenue M. Crépeau
17042 La Rochelle cedex 01
France

ABSTRACT

Digital images can be analyzed at wide range of levels going from pixel arrangement to semantics. As a consequence, finding a visual complexity estimator is a difficult task. In this article we propose a definition of attention based perceptual complexity. We study the performance of human eye movements and of different models of computational attention against a ground-truth of image complexity based on the observation time of an image description task. The results obtained show that besides its lack of semantic processing, attentional behavior is a good estimator of image complexity.

1. INTRODUCTION

1.1. Context

The study of visual complexity is particularly relevant to both cognitive and computer science studies. Yet, it is still ill-defined.

The pioneers in this field are the psychologists Snodgrass and Vanderwart[1] which, in the early 1980's established a classification of the complexity of a set of black and white line art drawings. This classification was based on subjective observers ranking. Later, in order to obtain a more objective measure of complexity, more "algorithmic" measures were introduced (number of line segments, lines crossings, etc.). These, however, were not calculated computationally.

In the 90's, the image processing community tackled this question in order to solve different kinds of problems. Firstly, evaluating image complexity allows measuring the performance of algorithms relative to the kind of images treated[2]. Another domain which exploits complexity is content based image retrieval (CBIR) [3]: instead of searching for images with the same attributes (shape, color, etc.) it can be interesting to request images with the same complexity as the query image. Image complexity can also be used in watermarking [4]: the more complex the image is, the easier it is to insert information without altering its quality. The list of possible applications is obviously much longer (image recognition, compression, etc.), but these few examples highlight the diversity of possibilities.

To sum up, estimating the complexity of an image is interesting for a wide range of applications (from psychology to computer science). However, its estimation is facing a major problem: the definition of complexity. Webster's dictionary (1986) defines a complex object as "*an arrangement of parts, so intricate as to be hard to understand or deal with*". According to W. Li, the definition of complexity should be very close to certain measures of difficulty (of construction, description, etc.) concerning the object or the system studied [5].

The measures of complexity arising from such a definition are numerous. Lloyd [6], proposes a non exhaustive list, classified into three categories: how hard is it to describe, how hard is it to create, and what is its degree of organization?

In the field of visual complexity, proposals are more scarce but equally varied. To name few: fractal theory [7], fuzzy theory [8] or, more frequently information theory [9]. Regarding this last point, and as previously mentioned, recent work [10, 11] presents the strong correlation between human evaluation of complexity and Kolmogorov complexity (JPEG and GIF compression ratios).

1.2. Hypothesis

The methods mentioned above evaluate the complexity of images, regardless of their perception by our visual and attentional systems. Intuitively, the heatmaps generated by a computational model of attention seem to vary (spreading, patterns) according image complexity (figure 1).

In order to verify this hypothesis, we propose to construct a measure of complexity based, no longer on the "direct" analysis of their content, but from this perceptual complexity through the attentional filter. This measure does not aim to compete or to replace the traditional measures, but rather to provide an additional tool for estimating the complexity of images.

In the following section, we describe a few computational models of attention as well as our contributions to the modeling of visual attention. In section 3, we describe how we have evaluated the potential of these models. Section 4 provides the results of this evaluation.



Fig. 1: The subjective evaluation of visual complexity is a difficult task. Is (a) more complex than (b)? This question may be partially answered by observing their corresponding attention maps (c) and (d).

2. COMPUTATIONAL ATTENTION

In order to validate the hypothesis that attention is correlated to image complexity, we have chosen to evaluate three well known bottom-up attention models. They were chosen for their proven performances in the field of attention modeling, their public availability [12] [13] and / or their link with publicly available eye-tracking experiments databases [13] [14]. We have also evaluated different versions of our own computational model of attention [15], partially derived from Laurent Itti’s work. The main characteristics of these models are summarized in the following subsections.

2.1. Some existing models

Itti’s model [12] is one of the most popular attention models. It is based on a hierarchical approach using multi-resolutions decomposition and center-surround filtering. It provides a centralized representation of attention through the generation of a global saliency map.

Le Meur’s model [14] is an improvement of [12], built on a more realistic (yet more complex) generation of features and conspicuity maps as well as an improved conspicuity maps fusion scheme.

Bruce [13] proposes an alternative approach based in information theory. It combines independent component analysis (ICA) with a measure of self-information in order to estimate the saliency of each pixel of an image.

2.2. Our contributions

All of the previously cited models are based on a central representation of saliency. The model proposed in [15] provides an alternative processing of visual attention, based on competition, leading to a distributed representation of attention. This competition is achieved by using a prey / predators dynamical system.

The advantages of such a system are numerous:

- it handles naturally the competition between the different kinds of visual information (intensity, color, orientation, etc.): thus, there is no need for additional normalization steps;

- it is inherently dynamical: the evolution of the focus of attention is generated naturally by the evolution of the prey / predators system;
- it is adaptable: thanks to its different parameters it can be tuned to many different applications. In particular, it provides a noise parameter allowing to balance between focusing exclusively on a few salient points and exploring the whole image;

In addition to this original competitive maps fusion process, we have improved the computation of conspicuity and singularity maps so that they are faster to generate. As a consequence, images can be analyzed using more scales, typically 6 against 3 for the default implementation of [12].

An extension of our basic system introduces a feedback mechanism which allows adapting the dynamics and scene exploration behavior of our system to a specific context. It modulates the activity of the dynamical system with a map built from previous attentional focuses (with a weight inversely proportional to their “oldness”). This mechanism is controlled by a feedback parameter $f \in [-1, 1]$ which allows to:

- maximize scene exploration and accelerate the system dynamics if $f \in [-1; 0]$;
- stabilize the focus points on previously visited locations and slow down system dynamics if $f \in [0; 1]$.

Experiments ran on the test databases described in section 3 shows that our system is closer to human attentional behavior if $f = 0.2$. This feedback value is used in the feedback based version of our system.

The aim of the experiments described in the next section is twofold. Firstly, it should validate that saliency based models of attention can help predicting the complexity of an image. Secondly, it should determine if the dynamical behavior of our attentional model can provide additional information leading to a better evaluation of visual complexity.

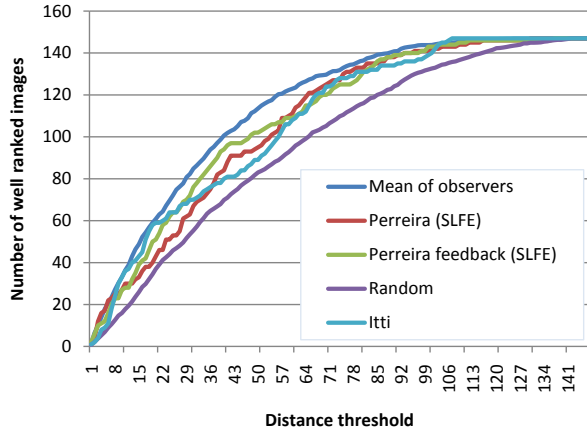


Fig. 2: ROC curves as an image complexity estimator.

3. EXPERIMENTS

3.1. Building a ground-truth of image complexity

Considering that image complexity is linked to its difficulty of description, we have measured the mean observation time of a set of $M = 148$ images belonging to two databases (containing images and eye-tracking data) made publicly available by Bruce [13] and Le Meur [14].

We asked $N = 12$ participants to watch each image of the database enough time to be able to describe it with a medium level of details. Observation times were normalized, averaged, and sorted in order to obtain a mean observer ranking MOR .

In order to evaluate the intra-observers variation, we built each observer’s own ranking OR^i with $i \in [1..N]$. For every rank $k \in [1..M]$ we can define a distance from the mean ranking MOR such as: $dOR[k] = |OR^i[k] - MOR[k]|$. We can then build a ROC curve by considering a binary classifier which correctly estimates image complexity if $dOR[k] < S$ with $S \in [1..M]$.

The ROC curve represents the number of good predictions relative to threshold S . The area under ROC curve provides a single estimator of the complexity estimation quality. The mean area under ROC curve of all observers is linked to the intra-observer variation: it represents the highest reference score to which computational models can be compared.

3.2. Models and measures

We evaluated the following methods in order to obtain a complexity based ranking of the database images:

- *Random ranking* of the images. This represents our lower bound of complexity evaluation;
- *JPEG compression ratio of the original images*. Provides a default Kolmogorov complexity estimator, di-

rectly based on image pixels;

- *JPEG compression ratio of the eye-tracking heatmaps* which are part of the databases provided by [13] [14]. This gives an estimate of how human eye movements can predict complexity;
- *JPEG compression ratio of the saliency maps* generated by several computation models of attention [12] [13] [14];
- *JPEG compression ratio of the heatmaps* built from simulated focus points obtained using [15] after 300 iterations;

Since our system is dynamical, we also built two specific measures in order to directly exploit the focus points coordinates:

- *DCR: the deflate compression ratio* [6] of the focus points coordinates measures if there are spatial redundancy or patterns in the evolution of the focus point.
- *SLFE: the saccade length Fourier entropy* [16] measures if there are temporal redundancies in the evolution of the distance between two consecutive focus points.

4. RESULTS AND DISCUSSION

Independently from our own attention model, we can see in table 1 that JPEG compression ratio of attentional maps seems to be a better complexity estimator than simple JPEG compression ratio of the original images [10]. This is true for computationally generated maps, but also for maps built from eye-tracking experiments.

Interestingly, complexity estimation obtained using computational models of attention outperforms estimation obtained using heatmaps generated from eye-tracking experiments (ground truth attention maps).

JPEG compression of the heatmaps generated by our dynamical attention model (table 2) provides slightly lower results than the other attention models but compares well to eye tracking, This is probably due to the use of randomness in its dynamical equations (which injects additional complexity to the heatmaps generated).

However, if we measure the complexity of the trajectories generated by the model and its feedback based refined version (SLFE estimator), the performances increase and even exceed those obtained by the best saliency map based models.

5. CONCLUSION

This article proposes a new method for estimating the complexity of natural images. It is a difficult task that has been addressed by only few previous works.

We show that attentional maps, and in particular saliency maps and heatmaps generated by computational models of attention, can serve as a good estimator of image complexity.

Data	Random	Raw images	Eye tracking	Intra-observer variation	Bruce	Itti	Le Meur
Complexity estimator	-	JPEG	JPEG	-	JPEG	JPEG	JPEG
Area under ROC curve	0.666	0.691	0.713	0.782	0.729	0.734	0.730

Table 1: Evaluation of attention based complexity estimators.

Data	Heatmap		Dynamical system			
	Perreira	Perreira feedback	Perreira	Perreira feedback	Perreira	Perreira feedback
Complexity estimator	JPEG	JPEG	DCR	DCR	SLFE	SLFE
Area under ROC curve	0.709	0.708	0.711	0.710	0.735	0.745

Table 2: Evaluation of complexity estimation using our visual attention model.

We also show that the dynamics of the focus points generated by our non centralized attention model can provide a better estimator of visual complexity.

Since our computational model of attention is capable of handling video inputs, future experiments will focus on studying the validity of our approach for video complexity estimation with promising applications in, for example, automatic bit-rate adjustment of video streams during their broadcast.

References

- [1] J G Snodgrass and M Vanderwart, "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity.," *Journal of experimental psychology. Human learning and memory*, vol. 6, no. 2, pp. 174–215, Mar. 1980.
- [2] I. Peters and R.N. Strickland, "Image complexity metrics for automatic target recognizers," in *Automatic Target Recognizer System and Technology Conference*. 1990, pp. 1–17, Citeseer.
- [3] Jukka Perkiö and Aapo Hyvärinen, "Modelling image complexity by independent component analysis , with application to content-based image retrieval," in *19th International Conference on Artificial Neural Networks: Part II*, Limassol, Cyprus, 2009, pp. 1–11, Springer Berlin / Heidelberg.
- [4] Farzin Yaghmaee and Mansour Jamzad, "Estimating Watermarking Capacity in Gray scale Images based on Image Complexity," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. Article ID 851920, 2010.
- [5] Wentian Li, "On the Relationship Between Complexity and entropy for markov chains and regular languages," *Complex Systems*, vol. 5, pp. 381–399, 1991.
- [6] S. Lloyd, "Measures of complexity: a nonexhaustive list," *IEEE Control Systems Magazine*, vol. 21, no. 4, pp. 7–8, Aug. 2001.
- [7] Nina Siu-Ngan Lam, Hong-lie Qiu, Dale a. Quattrochi, and Charles W. Emerson, "An Evaluation of Fractal Methods for Characterizing Image Complexity," *Cartography and Geographic Information Science*, vol. 29, no. 1, pp. 25–35, Jan. 2002.
- [8] Maurizio Cardaci, Vito Di Gesù, Maria Petrou, and Marco Elio Tabacchi, "A fuzzy approach to the evaluation of image complexity," *Fuzzy Sets and Systems*, vol. 160, no. 10, pp. 1474–1484, May 2009.
- [9] J Rigau, M Feixas, and M Sbert, "An information-theoretic framework for image complexity," *Computational Aesthetics 2005*, p. 177, 2005.
- [10] Gerry Mulhern and Martin Sawey, "Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing," *Behavior Research Methods*, vol. 40, no. 1, pp. 116–129, 2008.
- [11] Alexandra Forsythe, "Visual Complexity : Is That All There Is ?," *Complexity*, pp. 158–166, 2009.
- [12] Laurent Itti, Christof Koch, E. Niebur, and Others, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [13] N D B Bruce and J K Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 5, 2009.
- [14] O. Le Meur, Patrick Le Callet, Dominique Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.
- [15] Matthieu Perreira Da Silva, V Courboulay, A Prigent, and P Estraillier, "Evaluation of preys / predators systems for visual attention simulation," in *VISAPP 2010 - International Conference on Computer Vision Theory and Applications*, Angers, 2010, pp. 275–282, INSTICC.
- [16] AM Toh, Roberto Togneri, and Sven Nordholm, "Spectral entropy as speech features for speech recognition," *Proceedings of PEECS*, , no. 1, 2005.