

Une nouvelle mesure de complexité pour les images basée sur l'attention visuelle.

Matthieu PERREIRA DA SILVA, Vincent COURBOULAY

Laboratoire Informatique Image et Interaction - Université de La Rochelle,
Avenue M. Crépeau, 17042 La Rochelle Cedex 01, France
mperreir@univ-lr.fr, vcourbou@univ-lr.fr

Résumé – Les images numériques peuvent être analysées à de nombreux niveaux allant de la valeur de chaque pixel à la sémantique de la scène. Dans ce contexte, la définition d'un estimateur de la complexité visuelle d'une image est une tâche difficile. Dans cet article, nous proposons une définition de la complexité perceptuelle basée sur l'attention visuelle. Nous comparons les performances des mouvements oculaires humains ainsi que de différents modèles computationnels d'attention par rapport à une vérité terrain basée sur la mesure du temps d'observation lors d'une tâche de description d'images. Les résultats obtenus montrent que malgré leur manque de prise en compte de la sémantique, les focalisations attentionnelles sont un bon estimateur de la complexité des images.

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 Contexte

L'étude de la complexité visuelle est un phénomène relativement récent. Les précurseurs dans ce domaine sont les psychologues Snodgrass et Vanderwart [14] qui, au début des années 1980 ont établi un classement de la complexité d'un jeu d'images. Les recherches effectuées à cette époque visaient avant tout à étudier les processus cognitifs liés à l'analyse humaine des *stimuli* visuels. La complexité était alors déterminée expérimentalement par un certain nombre de sujets devant noter, sur une échelle donnée, la complexité des images de traits qu'on leur présentait. Afin d'obtenir une mesure plus objective de la complexité, des mesures plus « mathématiques » ont ensuite été introduites (nombre de segments de droites, de croisements de lignes, etc.). Celles-ci n'étaient cependant pas calculées de manière informatique.

Plus récemment, avec le développement des techniques de traitement d'images, la communauté informatique s'est également intéressée à la notion de complexité visuelle. Les applications potentielles sont en effet variées. Tout d'abord, connaître la complexité d'une image permet de mesurer qualitativement les performances relatives d'un algorithme de vision ou de traitement d'images en fonction de la complexité des images qui lui sont présentées [11]. Celle-ci peut aussi être une caractéristique intéressante pour un système de recherche d'images par le contenu [8] : on pourra alors rechercher des images possédant les mêmes attributs de formes, couleur, etc. et de même complexité

que l'image requête. La complexité peut également servir à estimer la capacité de watermarking d'une image [16] : plus celle-ci sera complexe, plus il sera facile d'y insérer de l'information sans en dégrader la qualité. La liste des applications possibles est bien évidemment plus longue, mais ces quelques exemples permettent de souligner la diversité des possibilités.

L'estimation de la complexité d'une image est donc intéressante, tant d'un point de vue de l'étude psychologique des phénomènes de perception, d'analyse et de mémorisation de l'information visuelle, que d'une exploitation plus informatique. Cependant, son calcul se heurte à un problème majeur : la définition même de la complexité. En effet, celle-ci est relativement floue. Le « Petit Robert » nous donne par exemple la définition suivante : « Complexe : qui se compose d'éléments différents, combinés d'une manière qui n'est pas immédiatement saisissable. »

Les mesures de complexité découlant d'une telle définition sont alors nombreuses. Lloyd [6], en propose une liste non exhaustive, classée en trois catégories : les mesures de difficulté de description, les mesures de difficulté de création, et les mesures de difficulté d'organisation. Dans le domaine de la complexité des images, les propositions sont moins nombreuses, mais tout aussi variées. Les méthodes utilisées peuvent par exemple être fractales [4], floues [12] ou, plus classiquement liées à la théorie de l'information [13]. Concernant ce dernier point, des études récentes montrent que les méthodes liées à la complexité de Kolmogorov (mesure du taux de compression GIF ou JPEG des images) permettent d'obtenir une estimation

automatique assez fidèle de l'évaluation de la complexité d'images faites par des humains [7, 2].

1.2 Hypothèse

Les méthodes évoquées plus haut évaluent la complexité des images, indépendamment de leur perception par notre système visuel et attentionnel. Nous proposons de construire une mesure basée non plus sur la complexité « directe » de leur contenu, mais sur leur complexité de perception, au travers d'un filtre attentionnel[9]. Ainsi, nous mesurons la complexité des trajectoires de focalisation de notre système attentionnel afin de déterminer si une image est complexe à parcourir (figure 1). Cette mesure n'a pas pour objectif de concurrencer ou remplacer les mesures traditionnelles, mais plutôt de fournir un complément autour de la complexité attentionnelle des images.

Dans la section suivante, nous décrivons quelques modèles attentionnels fondateurs ainsi que notre contribution par rapport à [9]. En section 3, nous décrivons comment est évalué le potentiel de ces modèles. Enfin, en section 4 nous présentons les résultats de cette évaluation.

2 Des modèles computationnels

Dans l'objectif de valider notre hypothèse selon laquelle la complexité de l'image est corrélée avec l'attention qu'elle génère, nous avons choisi trois modèles d'attention classiques. Nous les avons sélectionnés pour plusieurs raisons, leurs performances avérées, leur côté *open source*[3] [1] mais surtout le fait qu'ils sont liés à des bases d'expérimentations oculométriques [1] [5]. Nous avons aussi évalué différentes versions de notre algorithme [10]. Les caractéristiques principales de ces modèles sont résumées ci-après :

- Le modèle de L. Itti [3] est l'un des plus populaire. Il est basé sur une approche hiérarchique multi-résolution et un filtrage centre périphérie. Il fournit une représentation centralisée au travers d'une carte de saillance globale.
- Le modèle d'O. Le Meur [5] améliore le précédent en se basant sur une construction plus réaliste (mais plus complexe) des cartes d'intérêt. Il améliore également le processus de fusion.
- Enfin, N. Bruce [1] propose une approche basée sur la théorie de l'information en combinant une analyse par composantes indépendantes avec une mesure d'information de façon à estimer la saillance de chaque pixel.

Tous les modèles précédemment cités utilisent une représentation centrale de la saillance. Notre modèle [10, 9], dérivé partiellement [3] de propose un processus alternatif de détermination du focus d'attention utilisant une fusion dynamique permettant une représentation distribuée de l'attention. Les avantages d'un tel système sont nombreux, parmi eux :

- il représente naturellement la compétition entre les diverses sources d'informations (intensité, couleurs,

orientations, etc.), ainsi nous pouvons nous passer d'étapes de normalisation des différentes cartes de singularité;

- il est intrinsèquement dynamique : l'évolution du focus d'attention est gérée par les équations du système proies/prédateur ;
- il est adaptable : grâce aux différents paramètres du modèles nous pouvons simuler plusieurs comportements. En particulier la quantité de bruit injectée dans la modèle permet de dose l'exhaustivité de l'exploration de l'image.
- il est rapide : nous avons accéléré le calcul des différentes cartes par exemple en utilisant des images intégrales.

Une extension du modèle de base consiste en l'intégration d'une rétroaction (*feedback*). Celle-ci permet d'adapter la dynamique du comportement d'exploration de la scène à un contexte spécifique. Ce bouclage se fait grâce à une carte représentant l'historique des focalisations attentionnelles précédentes (leur contribution étant inversement proportionnelle à leur ancienneté). Un paramètre de bouclage $f \in [-1, 1]$ permet de :

- maximiser l'exploration de l'image si $f \in [-1; 0]$;
- stabiliser les points de fixation et la dynamique du système si $f \in [0; 1]$.

Des expérimentations menées précédemment montrent que le système possède un comportement proche de la dynamique humaine si $f = 0.2$. Cette valeur sera la valeur choisie dans les expérimentation de la section 3. L'objectif de ces expérimentations est double. Premièrement elles permettent de montrer l'intérêt des modèles basés saillance dans leur capacité à estimer la complexité des images. Secondement, elles indiquent que notre modèle dynamique fournit une information supplémentaire améliorant l'estimation de la complexité.

3 Expérimentations

3.1 Définition d'une vérité terrain

En considérant vrai l'hypothèse selon laquelle la complexité d'une image est reliée à sa difficulté de description, nous avons mesuré le temps d'observation moyen sur un jeu de $M = 148$ images appartenant à deux bases ouvertes proposées par Bruce [1] et Le Meur [5]. Il est à noter que chacune de ces bases proposent des enregistrements oculométriques. Nous avons demandé à $N = 12$ participants de regarder chacune de ces images un temps « suffisant » pour pouvoir les décrire avec un niveau de détail raisonnable. Ces mesures ont toutes été normalisées, moyennées et classées afin d'obtenir un temps d'observation moyen, temps noté *MOR* (Mean Observer Ranking).

De façon à évaluer la variation inter observateurs, nous avons construit un classement personnel à chaque observateur, son OR^i avec $i \in [1...N]$. Pour chaque classement $k \in [1...M]$, nous sommes ainsi à même de définir une distance par rapport au classement moyen *MOR* telle que :



FIGURE 1: Il est difficile de déterminer subjectivement si (a) est plus complexe que (b). Par contre, l’analyse de leurs cartes d’attention respectives (c) et (d) semble indiquer que (a) est perceptuellement plus simple que (d).

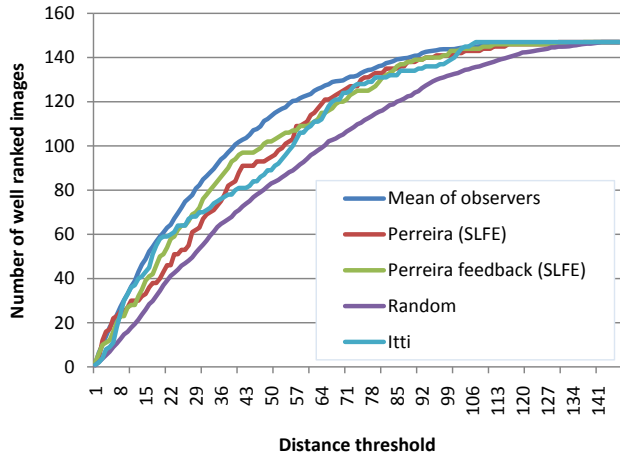


FIGURE 2: Utilisation de l’aire sous la courbe ROC en tant qu’estimateur de la complexité des images

$dOR[k] = | OR^i[k] - MOR[k] |$. Une fois cette distance définit, nous pouvons facilement construire une courbe ROC à partir d’un classifieur binaire $dOR[k] < S$ avec $S \in [1...M]$. Cette courbe ROC représente alors le nombre de bonne prédictions du classement relatif à un seuil S . L’aire sous la courbe fournit un estimateur de la qualité de l’estimation de la complexité. La moyenne des aires sous la courbe de toutes les observations est quant à elle reliée à la variation inter observateurs, elle représente la borne haute à laquelle les modèles computationnels vont pouvoir être comparés (figure 2).

3.2 Modèles et mesures

Nous avons évalué les méthodes suivantes afin d’obtenir une base de mesures de complexité sur les ensembles d’images testés :

- *Notation aléatoire* des images. Cela représentera notre borne basse de l’évaluation ;
- *Ratio de compression JPEG des images originales*. Ce ratio fournit un estimateur de type Kolmogorov directement basé sur la valeur des pixels ;
- *Ratio de compression JPEG des cartes d’expériences oculométriques* fournies par [1] [5]. Cela nous permet d’estimer comment les focalisations de l’oeil humain peuvent être corrélée avec la complexité de l’image ;
- *Ratio de compression JPEG des cartes de saillance* modélisée par les méthodes présentées section 2 ;
- *Ratio de compression JPEG de la carte d’attention*

modélisée par notre système [10] après 300 itérations.

Puisque notre système est dynamique, nous avons voulu compléter notre estimation par la génération de deux mesures exploitant la trajectoire du focus d’attention :

- *DCR : Deflate Compression Ratio* [6] traduit le degré de redondance dans l’évolution spatiale des points de la trajectoire en exploitant le principe du *minimum description length* sur les coordonnées de focalisations attentionnelles.
- *SLFE : Saccade Length Fourier Entropy* [15] mesure lui la redondance temporelle dans l’évolution de la distance s_k entre deux points consécutifs (x_{k-1}, y_{k-1}) et (x_k, y_k) .

$$SLFE = \sum_k p(S_k) \log(p(S_k))$$

avec $S = \mathcal{F}(s)$ la transformée de Fourier de la série de distances s , et $p(S_k) = \frac{S_k}{\sum_k S_k}$ la distribution statistique des fréquences de cette série.

4 Résultats et discussions

Nous pouvons voir dans les résultats présentés Tableau 1 que l’estimateur de complexité basé sur ratio de compression JPEG des cartes d’attention est meilleur que celui basé simplement sur la compression JPEG des images originales [7]. Cette assertion est vrai pour les cartes générées avec les algorithmes computationnels mais aussi pour les cartes construites avec les données oculométriques. A noter également, l’estimation de complexité obtenue par notre modèle d’attention confirme l’estimation obtenue par les acquisitions oculométriques, cependant la compression JPEG des cartes de saillance générées par notre système dynamique (tableau 2) fournit un résultat légèrement inférieur que ceux fournis par les autres. Ce dernier point est certainement dû à l’utilisation du bruit dans nos équations dynamiques, ce qui injecte de la complexité dans les trajectoires générées.

Cependant, la même mesure effectuée sur notre modèle avec bouclage (estimateur SLFE), permet d’améliorer les performances et même dépasser celles obtenues avec les autres cartes de saillance.

Données	Alléatoire	Images raw	Eye tracking	Variation intra-observateur	Bruce	Itti	Le Meur
Estimateur de complexité	-	JPEG	JPEG	-	JPEG	JPEG	JPEG
Aire sous la courbe ROC	0.666	0.691	0.713	0.782	0.729	0.734	0.730

TABLE 1: Evaluation des estimateurs de complexité basés attention

Données	Carte de saillance		Système dynamique			
	Perreira	Perreira bouclé	Perreira	Perreira bouclé	Perreira	Perreira bouclé
Estimateur de complexité	JPEG	JPEG	DCR	DCR	SLFE	SLFE
Aire sous la courbe ROC	0.709	0.708	0.711	0.710	0.735	0.745

TABLE 2: Evaluation de la complexité avec notre modèle d'attention

5 Conclusion

Dans cet article, nous avons proposé une nouvelle méthode d'estimation de la complexité des images basée sur l'attention. Nous avons montré que les cartes attentionnelles, de saillance ou de focus d'attention, peuvent servir à estimer correctement la complexité des images. Nous avons également montré que l'utilisation de la dynamique du focus d'attention générée par notre modèle d'attention distribuée améliore l'estimation de cette complexité. Nous avons proposé une expérimentation visant à déterminer une vérité terrain. Enfin, puisque notre modèle est capable de gérer les flux vidéos, nous travaillons à la transposition de nos travaux sur les images à la mesure de complexité vidéo. Une utilisation pourrait être par exemple l'ajustement de la compression en fonction de la complexité des séquences. Les premiers résultats sont plus que prometteurs.

Références

- [1] N. D. B. Bruce and J. K. Tsotsos. Saliency, attention, and visual search : An information theoretic approach. *Journal of Vision*, 9(3) :5, 2009.
- [2] A. Forsythe. Visual Complexity : Is That All There Is? *Complexity*, pages 158–166, 2009.
- [3] L. Itti, C. Koch, E. Niebur, and Others. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11) :1254–1259, 1998.
- [4] N. Lam, H. Qiu, D. Quattrochi, and C. Emerson. An Evaluation of Fractal Methods for Characterizing Image Complexity. *Cartography and Geographic Information Science*, 29(1) :25–35, January 2002.
- [5] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5) :802–817, 2006.
- [6] S. Lloyd. Measures of complexity : a nonexhaustive list. *IEEE Control Systems Magazine*, 21(4) :7–8, August 2001.
- [7] G. Mulhern and M. Sawey. Confounds in pictorial sets : The role of complexity and familiarity in basic level picture processing. *Behavior Research Methods*, 40(1) :116–129, 2008.
- [8] J. Perkiö and A. 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*, pages 1–11, Limassol, Cyprus, 2009. Springer Berlin / Heidelberg.
- [9] M. Perreira Da Silva, V. Courboulay, A. Prigent, and P. Estraillier. Attention visuelle et systèmes proies/prédateurs. In *Images), XXIIe Colloque GRETSI (Groupe d Etudes du Traitement du Signal et des Images*, Dijon, France, 2009.
- [10] M. 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*, pages 275–282, Angers, 2010. INS-TICC.
- [11] I. Peters and R.N. Strickland. Image complexity metrics for automatic target recognizers. In *Automatic Target Recognizer System and Technology Conference*, pages 1–17. Citeseer, 1990.
- [12] M. Petrou and M.E. Tabacchi. On the Evaluation of Images Complexity : A Fuzzy Approach. In *Fuzzy logic and applications : 6th international workshop, WILF 2005, Crema, Italy, September 15-17, 2005 : revised selected papers*, page 305. Springer-Verlag New York Inc, 2006.
- [13] J. Rigau, M. Feixas, and M. Sbert. An information-theoretic framework for image complexity. *Computational Aesthetics 2005*, page 177, 2005.
- [14] J. 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*, 6(2) :174–215, March 1980.
- [15] A. Toh, R. Togneri, and S. Nordholm. Spectral entropy as speech features for speech recognition. *Proceedings of PEECS*, (1), 2005.
- [16] F. Yaghmaee and M. Jamzad. Estimating Watermarking Capacity in Gray scale Images based on Image Complexity. *EURASIP Journal on Advances in Signal Processing*, 2010.