

Attention visuelle et systèmes proies / prédateurs

Matthieu PERREIRA DA SILVA, Vincent COURBOULAY, Armelle PRIGENT, Pascal ESTRAILLIER

Laboratoire Informatique Image et Interaction

Université de La Rochelle, Avenue M. Crépeau, 17042 La Rochelle cedex 0, France

mperreir@univ-lr.fr, vcourbou@univ-lr.fr, aprigent@univ-lr.fr, pestrail@univ-lr.fr

Résumé – Cet article présente une méthode utilisant un système d'équations proies / prédateurs afin de simuler l'évolution en temps réel du focus d'attention. Cette méthode permet une approche plus réaliste de l'attention visuelle sélective pour les systèmes de vision par ordinateur.

Abstract – This article presents a method based on preys / predators systems for real time focus of attention simulation. This method represents a more realistic approach for computer vision based selective visual attention.

1 Introduction

Dans [8], nous avons présenté une architecture de vision comportementale destinée à servir de base à un *compagnon virtuel*. Ce compagnon devant être interactif, son système de vision se soit d'être temps réel. Afin de satisfaire cette contrainte nous avons décidé d'utiliser un mécanisme d'attention visuelle permettant de détecter les éléments les plus intéressants de la scène et ainsi réduire les ressources nécessaires à l'étude de l'environnement. Différents modèles informatiques de l'attention visuelle ont été développés tel que [10] ou [2]. Malheureusement, ils sont souvent trop complexes pour une exécution temps réel et/ou prennent en compte de manière simpliste l'évolution temporelle du focus d'attention.

L'approche que nous proposons ici permet d'étudier l'évolution temporelle du focus d'attention en se basant sur une version modifiée des algorithmes classiques de [4]. L'article originel propose une architecture comprenant des calculs *bas niveau* permettant d'extraire trois cartes de singularité (conspicuity maps). Ces cartes de singularité représentent l'importance respective des différents éléments de la scène observée par rapport aux principaux canaux perceptuels humains (intensité, couleur, orientation). Dans notre architecture, ces calculs bas niveau sont effectués en utilisant des algorithmes dérivés des travaux de [2] eux même basés sur [4]. L'approche utilisée afin d'accélérer les calculs (*i.e.* l'utilisation d'images intégrales) a été reprise et étendue afin que toutes les cartes soient calculées en temps réel. Cependant, afin de pouvoir mieux comparer nos résultats avec ceux de [4], les cartes de singularité utilisées dans cet article sont calculées grâce à la *Saliency Toolbox* [11].

La deuxième partie de l'architecture de [4] propose un système *moyen niveau* permettant de fusionner les cartes singularité puis de générer un parcours d'attention visuelle à partir d'un système de *winner takes all* et *inhibition of return*. C'est cette deuxième partie que nous proposons de remplacer par un système proies / prédateurs afin d'introduire un paramètre temporel permettant de générer des saccades, fixations et trajectoires de manière plus plausible (figure 1).

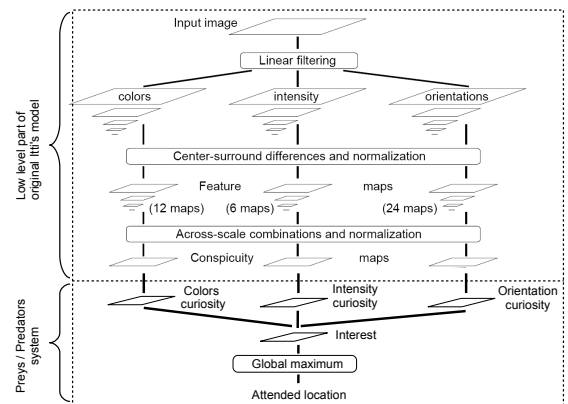


FIGURE 1 – Architecture du système de modélisation d'attention visuelle. Le schéma a été repris et modifié à partir de [4].

Le choix des systèmes proies/prédateurs est justifié par les arguments suivants :

- les systèmes proies / prédateurs sont des systèmes dynamiques. Ils gèrent donc naturellement l'évolution temporelle de leur activité ce qui est primordial pour générer l'évolution temporelle du focus d'attention à partir d'une image fixe ou d'une séquence d'images.
- en l'absence d'objectif (flux d'information top-down ou *prégnance*), le choix d'une méthode de fusion des informations issues des différentes cartes de singularité est difficile à faire. Les systèmes proies / prédateurs proposent une solution permettant la mise en place d'un équilibre naturel entre ces différentes cartes *via* un mécanisme de compétition (*prédation*).
- les systèmes dynamiques discrétisés peuvent avoir un comportement chaotique. Bien que cette propriété ne soit généralement pas recherchée, c'est une qualité importante dans notre cas puisqu'elle permet l'émergence de parcours originaux, permettant d'explorer la scène visuelle, même dans des zones faiblement saillantes.

Dans la section suivante, nous présentons comment ces systèmes peuvent être utilisés pour simuler l'évolution temporelle des phénomènes d'attention visuelle.

2 Une simulation de l'évolution temporelle de l'attention visuelle

2.1 Les systèmes proies / prédateurs

Les systèmes proies prédateurs sont des systèmes d'équations utilisés pour simuler l'évolution et l'interaction de différentes colonies de proies et de prédateurs. Plus de détails sur ces systèmes peuvent être trouvés dans [7]. Dans notre système, nous nous sommes inspirés de [6] afin de représenter l'évolution temporelle de l'intérêt (*i. e.* du focus d'attention) lié, dans un premier temps (cf. section 5, à une image statique.

Classiquement, l'évolution d'un système proies / prédateurs 2D est régie par quelques règles simples inspirées des systèmes d'équation de Volterra-Lotka [7] :

1. les proies C ont un taux de croissance proportionnel à leur population et un facteur de croissance b ;
2. les prédateurs I ont un taux de croissance proportionnel à leur taux de prédation CI (taux de rencontre entre les proies et les prédateurs) et un facteur de prédation s ;
3. les proies et les prédateurs se déplacent grâce à une règle de diffusion classique proportionnelle à leur laplacien Δ_C et un facteur de diffusion f ;
4. les prédateurs ont un taux de mortalité m_I proportionnel à leur population ;
5. les proies ont un taux de mortalité m_C proportionnel à leur population, additionné d'une mortalité supplémentaire proportionnelle au taux de prédation CI .

2.2 Transposition dans l'analyse de l'attention

Pour simuler l'évolution temporelle des points de fixation, nous proposons d'utiliser un système d'équations proies / prédateurs tel que décrit ci-dessus avec les caractéristiques suivantes :

- il existe trois populations de proies et une seule population de prédateurs ;
- les trois populations de proies représentent la répartition spatiale de la curiosité engendrée par les trois cartes de singularité (intensité, couleur, orientation) ;
- la population de prédateurs représente l'intérêt généré suite à la consommation de la curiosité liée aux différentes cartes ;
- le maximum global d'intérêt (maximum de la carte des prédateurs) représente le focus d'attention à l'instant t .

L'application de ces caractéristiques à un système d'équations proies / prédateurs permet d'obtenir les équations qui sont décrites dans la sous-section 2.3.

2.3 Simulation de l'évolution du focus d'attention par système proies / prédateurs

Pour chacune des trois cartes de singularité (couleur, intensité, orientation) calculées, l'équation de la matrice des proies C est régie par l'équation suivante :

$$\frac{dC_{x,y}^n}{dt} = hC_{x,y}^{*n} + hf \Delta_{C_{x,y}^{*n}} - m_C C_{x,y}^n - sC_{x,y}^n I_{x,y}$$

avec $C_{x,y}^{*n} = C_{x,y}^n + wC_{x,y}^{n-2}$ et $n \in \{c, i, o\}$, ce qui signifie que cette équation est valable pour les 3 matrices C^c , C^i et C^o représentant respectivement la couleur, l'intensité et l'orientation.

La matrice I des prédateurs consommant ces 3 types de proies est régie par l'équation suivante :

$$\frac{dI_{x,y}}{dt} = s(P_{x,y} + wI_{x,y}^2) + sf \Delta_{P_{x,y} + wI_{x,y}^2} - m_I I_{x,y}$$

avec

$$P_{x,y} = \sum_{n \in \{c, i, o\}} (C_{x,y}^n) I_{x,y}$$

et

$$h = b(1 - g + gG)(a * R + (1 - a) * S)$$

C représente la curiosité générée à partir de la singularité intrinsèque de l'image. Elle est créée à partir d'une combinaison h de trois facteurs :

- la *singularité S de l'image*, calculée pour cet article grâce la Saliency Toolbox [11] et dont la contribution est inversement proportionnelle au facteur a ;
- une *source R de bruit aléatoire*, simulant le haut niveau de bruit de l'activité de notre cerveau [1] et dont a définit l'intensité par rapport à S ;
- une *carte gaussienne G* permettant de simuler la préférence centrale observée lors des expérimentations psychovisuelles [9]. L'importance de cette carte est modulée par le facteur g .

De manière similaire à [6] un terme quadratique modulé par le facteur w est ajouté aux équations classiques de Volterra-Lotka afin de simuler la non linéarité (feedback positif). Cette non-linéarité permet de renforcer la dynamique du système et facilite l'émergence du comportement chaotique du système en favorisant la saturation de certaines valeurs des matrices. Enfin, nous rappelons que la curiosité C est consommée par l'intérêt I et que le point de fixation à un instant t est le maximum de la carte d'intérêt.

3 Résultats

Comme indiqué dans la section 4.2 de [2], l'évaluation d'un système d'attention visuel comportant pas ou peu d'information top-down est difficile car il n'existe pas de réelle vérité terrain. Il est certes possible d'effectuer des mesures oculométriques sur différents sujets, mais les variabilités intra et inter-sujets constatées généralement dans la littérature ne permettent pas de déterminer un parcours visuel de référence pour une image donnée. Pour contourner partiellement ce problème, on compare généralement non pas le parcours visuel en lui-même mais une carte statistique (*carte de saillance*) représentant la probabilité que le sujet regarde à un endroit donné (voir chapitre 1.3 de [5]). Dans cet article, nous utiliserons cette méthode d'évaluation pour comparer nos résultats à ceux obtenus par [4] ainsi qu'à une vérité terrain obtenue, sur une durée de 10 secondes, par [5]. Cette comparaison ne permettant pas

d'évaluer l'aspect dynamique du système d'attention visuelle, nous effectuerons également quelques remarques concernant la *plausibilité* de la dynamique des deux systèmes comparés.

3.1 Algorithme de référence

La figure 2 représente la vérité terrain issue de [5] ainsi les résultats obtenus en utilisant la Saliency Toolbox [11] (implémentant les algorithmes de [4] en Matlab). Les paramètres utilisés sont ceux par défaut à l'exception de l'échelle des cartes de singularité et de la carte de saillance qui est 3 (4 fois plus petit que l'image initiale) au lieu de 5 (16 fois plus petit). Notons également que l'on utilise bien la *normalisation itérative* permettant, en l'absence d'information top-down, d'effectuer une fusion des cartes de singularité favorisant celles contenant un pic unique (un seul ballon orange au milieu d'un champ vert par exemple).

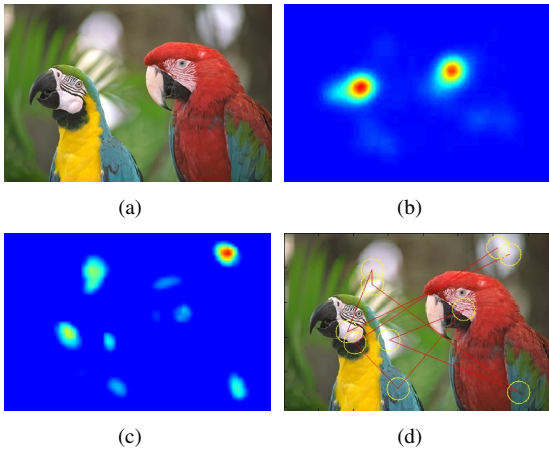


FIGURE 2 – Vérité terrain et résultat obtenus avec la Saliency Toolbox [11] : (a) Image original ; (b) Vérité terrain obtenue par [5] ; (c) Carte de saillance ; (d) Exemple des 10 premiers focus d'attention.

3.2 Système proies / prédateurs

La figure 4 présente les cartes de saillance calculées à partir de différentes durées de simulation (soit environ 1, 2 et 10 secondes, en considérant que notre système traite 30 images par secondes). Il est à noter qu'au delà de 10 secondes de simulation, on ne constate plus de changement significatif dans les cartes obtenues. Ces résultats ont été obtenus en utilisant les 3 cartes de singularité non normalisées (afin de permettre à notre système d'exploiter un maximum d'information), calculées avec la Saliency Toolbox (figure 3). On constate que bien que restant différents de la vérité terrain, les résultats obtenus dans les différents cas présentent de meilleures similitudes avec celle-ci que ceux obtenus avec la méthode de [4].

On relativisera cependant ces résultats en rappelant que lors des mesures oculométriques, il n'est pas possible de supprimer le biais sémantique (information top-down) qui pousse à regarder la tête des oiseaux tout en gardant un nombre de points d'acquisition statistiquement suffisant. On notera également que les résultats sont très dépendants de la méthode utilisée pour calculer les cartes de singularité (voir figure 5).

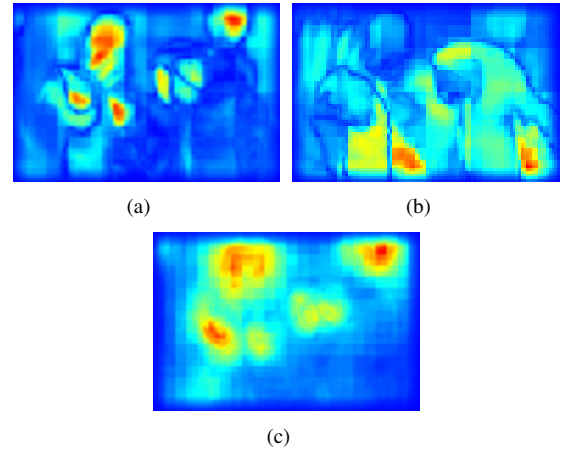


FIGURE 3 – Cartes de singularité non Normalisées obtenus avec la Saliency Toolbox [11]. Ces cartes servent de base au système proies / prédateurs : (a) Carte de singularité du canal *intensité* ; (b) Carte de singularité des canaux *couleur* ; (c) Carte de singularité des canaux *orientation*.

Les différentes simulations ont été réalisées avec les paramètres suivants, déterminés empiriquement :

a	b	g	w	m_C	m_I	s	f
0.3	0.005	0.01	0.001	0.3	0.5	0.025	0.2

Ces paramètres représentent des valeurs raisonnables permettant d'obtenir un système à l'équilibre. Cependant, d'autres combinaisons sont possibles. En particulier, il est possible de faire varier les valeurs indiquées dans une fourchette de $\pm 50\%$ sans que les résultats n'en soient particulièrement affectés.

Concernant le comportement dynamique, le système génère une évolution temporelle du focus d'attention très semblable aux résultats de [4], mais sans utiliser de mécanisme explicite d'inhibition de retour. Ce mécanisme est produit naturellement par la *prédation* de la curiosité par l'intérêt. Le système effectue différentes saccades sur les éléments les plus saillants des cartes de singularité et est capable de reboucler sur des zones déjà parcourues avec des variations dont l'importance est déterminée par le facteur de curiosité a .

4 Discussion et extensions

Le système proies / prédateurs tel que décrit ci-dessus donne des résultats pertinents. Cependant certaines améliorations peuvent être apportées afin de rendre le modèle plus naturel et améliorer son comportement :

- le calcul des cartes de singularité selon [4] comporte une étape de normalisation des cartes de caractéristiques (feature maps) par un opérateur $N(\cdot)$ afin de compenser l'absence d'information top-down dans le modèle. Dans notre modèle, une partie du comportement top-down du système de vision étant simulé par les équations proies / prédateurs, on module alors le facteur h de croissance des proies de la façon suivante :

$$h = b(1 - g + gG)(a * R + (1 - a) * S)(1 - e)$$

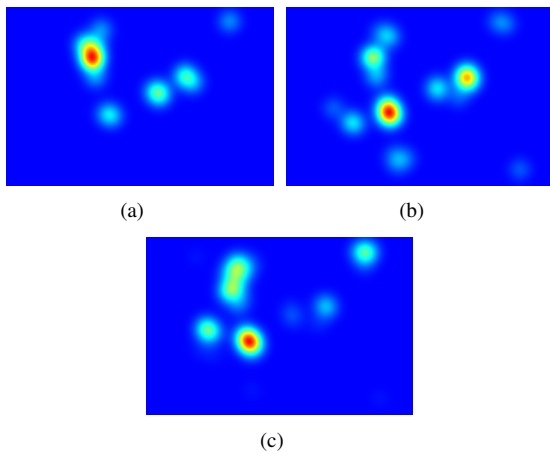


FIGURE 4 – Résultats de la simulation proies / prédateurs après différents temps de simulation : (a) Au temps $t = 30$; (b) Au temps $t = 60$; (c) Au temps $t = 300$;

avec e l'entropie de la carte de singularité (couleur, intensité ou orientation) normalisée entre 0 et 1.

La modulation par l'inverse de l'entropie permet de favoriser les cartes possédant un nombre limité de maximum locaux. Expliqué en termes de proies/prédateurs, on favorise la croissance des populations de proies les plus organisées (regroupée en un petit nombre de sites).

- il est possible de calculer des cartes de singularité plus réalistes, tout en conservant des temps de calcul raisonnables, en utilisant les travaux de [3]. Comme indiqué au début de cet article, il est nécessaire de pousser encore plus loin la simplification et l'optimisation du calcul des cartes pour que celui-ci soit temps réel. Ce n'est pas l'objet de cet article, mais il est cependant intéressant de présenter un exemple de résultat obtenu avec un calcul de carte plus élaboré, sur la même image d'exemple que précédemment (figure 5). Les paramètres utilisés sont les mêmes que ceux indiqués en section 3.2.

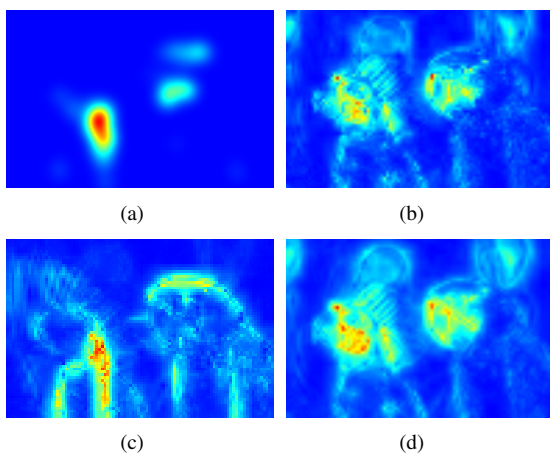


FIGURE 5 – Résultats obtenus avec des cartes de singularités plus réalistes : (a) Cartes de saillance à $t = 300$; (b) Carte de singularité du canal intensité ; (c) Carte de singularité du canal couleur ; (d) Carte de singularité du canal orientation.

5 Conclusion et perspectives

Les résultats obtenus avec ce système sont très prometteurs. Afin de mieux exploiter tout son potentiel, nous souhaitons étendre ce système en lui permettant de traiter, non plus des images statiques, mais des séquences d'images. Ainsi les interactions complexes entre les cartes ainsi que l'évolution temporelle du système pourront être mieux exploitées. On pourra également peut être mieux exploiter le potentiel des équations proies-prédateurs en effectuant des interactions plus complexes entre les cartes.

Références

- [1] Michael D. Fox, Abraham Z. Snyder, Justin L. Vincent, and Marcus E. Raichle. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, 56(1) :171 – 184, 2007.
- [2] Simone Frintrap. *VOCUS : A Visual Attention System for Object Detection and Goal-Directed Search*. PhD thesis, University of Bonn, 2006.
- [3] Simone Frintrap, Maria Klodt, and Erich Rome. A real-time visual attention system using integral images. University Library of Bielefeld, 2007.
- [4] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11) :1254–1259, Nov 1998.
- [5] Olivier Le Meur. *Attention sélective en visualisation d'images fixes et animées affichées sur écran : Modèles et évaluation de performances - Applications*. PhD thesis, Ecole Polytechnique de l'Université de Nantes, Octobre 2005.
- [6] Mike Lesser and Dinah Murray. Mind as a dynamical system : Implications for autism. In *Psychobiology of autism : current research & practice*, 1998.
- [7] James D. Murray. *Mathematical Biology : I. An Introduction*. Interdisciplinary Applied Mathematics. Springer, 2002.
- [8] M. Perreira Da Silva, V. Courboulay, A. Prigent, and P. Estraillier. Adaptativité et interactivité - vers un système de vision comportemental. In *MajecSTIC 2008*, 2008.
- [9] Benjamin 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) :1–17, 11 2007.
- [10] John K. Tsotsos, Yueju Liu, Julio C. Martinez-trujillo, Marc Pomplun, Evgueni Simine, and Kunhao Zhou. K. : Attending to visual motion. *CVIU*, 100 :3–40, 2005.
- [11] Dirk Walther. *Interactions of Visual Attention and Object Recognition : Computational Modeling, Algorithms, and Psychophysics*. PhD thesis, California Institute of Technology, Pasadena, California, March 2006. Encadré par Christof Koch et Pietro Perona, Laurent Itti, Shinsuke Shimojo, Richard Andersen.