Adaptativité et interactivité

Vers un système de vision comportemental

Matthieu Perreira Da Silva — Vincent Courboulay
Armelle Prigent — Pascal Estraillier

Laboratoire Informatique Image Interactions
Université de La Rochelle
Avenue Michel Crépeau
17042 La Rochelle Cedex 1
mperreir@univ-lr.fr

RÉSUMÉ. *Les systèmes de vision actuels sont certes performants mais trop spécialisés et trop peu adaptables à leur environnement. Ceci est dû entre autres au fait que nous ne sommes, pour l'instant, pas capable de produire des systèmes à la fois adaptables et efficaces. Dans cet article nous présentons les limitations de ces systèmes et proposons une architecture de vision comportementale qui n'aura certes pas les capacités du système visuel d'un homme, mais qui pourra à terme réagir et apprendre d'un environnement visuel quelconque.*

ABSTRACT. *Current vision systems are highly capable but are too specialized and lack adaptivity. These limitations are due to the fact that we are still unable to build systems that would be adaptive and at the same time have high performances. In this article we describe the limitations of such systems and propose a behavioural vision architecture that despite not reaching the performances of the human visual system will be able to react and learn from any visual environment.*

MOTS-CLÉS : *Vision, temps réel, bio-inspiré, attention, saillance, adaptativité, apprendissage.*

KEYWORDS: *Vision, real time, bio-inspired, attention, saliency, adaptivity, learning.*

1. Introduction

De plus en plus de systèmes interactifs proposent des interfaces naturelles. Ainsi, dans le domaine du jeu, la wii et sa manette dotée d'un accéléromètre permettent au joueur d'interagir directement via des mouvements. Dans ce contexte, une nouvelle approche de l'interaction consiste à se baser sur la vision.

Dans (Matthieu Perreira Da Silva *et al.,* 2008), nous avons développé un système utilisant la vision par ordinateur afin de rendre une application interactive réactive à l'état d'attention de l'utilisateur. Cependant, ce système de vision présente un certain nombre de limitations, partagées par une grande majorité des systèmes de vision actuels :

– son contexte d'utilisation est restreint,

– ses capacités ne sont pas évolutives.

De manière imagée, nous pouvons comparer les systèmes de vision actuels à des insectes : ils naissent avec toutes les connaissances dont ils auront besoin au cours de leur vie et effectuent les tâches pour lesquelles ils sont programmés de manière très efficace. Par contre, ils n'ont aucune capacité d'adaptation et ne pourront pas survivre en cas d'évolution brusque de leur environnement. A l'opposé des insectes, l'homme a des capacités très limitées à sa naissance et a besoin de plusieurs mois voire années avant d'atteindre ses capacités optimales. Ce *handicap* de départ est compensé par une capacité d'apprentissage et d'adaptation qui ont permis à l'homme d'atteindre un haut niveau de développement.

Par analogie, nous pensons que pour dépasser les limites des systèmes de visions actuels et les rendre plus adaptatifs et autonomes il est nécessaire de repenser la façon dont ces systèmes doivent être conçus. Une première approche en ce sens est de proposer un système de vision adaptatif et réactif. Au regard de la complexité du système cognitif humain, nous envisageons de développer un système ayant des capacités d'adaptation basiques mais avec des possibilités d'évolution en fonction des futurs matériels et/ou algorithmes.

Dans les chapitres suivants nous mettrons en avant les caractéristiques communes et distinctives des systèmes visuels humains et informatiques, puis détaillerons notre proposition de système visuel comportemental.

2. Les systèmes de vision humain et informatique

2.1. *Le système visuel humain*

Le système visuel humain est complexe. Il a été et sera encore source de nombreuses études, car bien que l'on comprenne de mieux en mieux son fonctionnement nous sommes encore loin de comprendre toutes ses interactions, rétroactions, évolutions. Il est globalement acquis actuellement que son architecture globale serait codée

par nos gènes, mais que les détails de son *câblage* seraient le fruit d'une adaptation à l'environnement (celui-ci pouvant être les signaux émis par la rétine, autant que l'activité interne et autonome du cerveau) (Bednar, 2002). La figure 1 donne une vision schématique de cette architecture. On trouvera également dans (Frintrop, 2006) une description neurobiologique détaillée du système de vision humain.

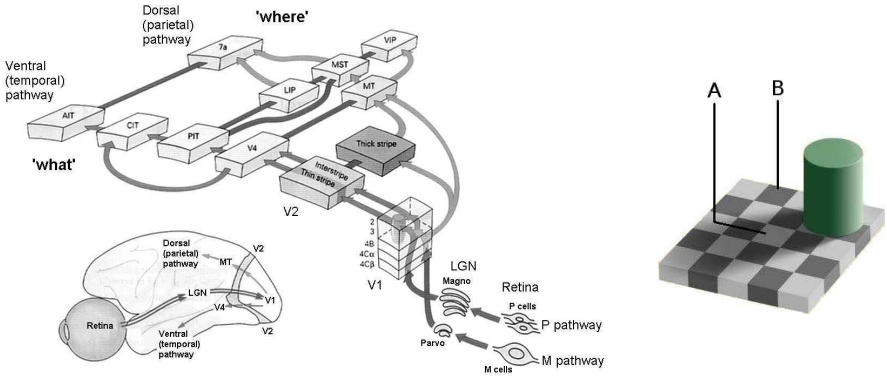

Figure 1. *Gauche : Architecture du système visuel humain ; illustration tirée de (Behnke, 2002). Droite : illusion de perception des intensités ; image originale d'Edward H. Adelson. Malgré ce qu'essaie de nous faire croire notre cerveau, les cases A et B du damier possèdent la même intensité lumineuse.*

Le rôle des différentes aires du système visuel (LGN, V1, V2, V3, V4, MT) est plus ou moins bien identifié. Le cortex visuel primaire (V1) est l'aire la plus étudiée et la mieux connue (Bednar, 2002).

L'étude de ces différentes aires corticales a permis de déterminer les principales caractéristiques du système visuel (Behnke, 2002) :

– la rétine n'a pas une résolution spatiale uniforme. Elle est bien plus précise en son centre qu'à sa périphérie,

– il existe deux voies de traitement des données issues de la rétine :

- la voies dorsale (voie où) gère les aspects spatio-temporels des signaux reçus. Elle travaille rapidement sur des signaux de faible résolution,

- La voie ventrale (voie quoi) est impliquée dans la reconnaissance des objets.

– l'information visuelle est séparée en canaux chromatiques (couple d'opposition Rouge/Vert et Jaune/Bleu) et achromatique (Luminance),

– l'information visuelle est également séparée en différents canaux sensibles à une orientation et une fréquence spécifique,

– l'attention visuelle est utilisée comme mécanisme de réduction de la quantité d'information à traiter. La focalisation du système visuel sur des éléments *saillants* permet un traitement plus efficace de ceux-ci,

– un certain nombre *d'éléments discriminants* regroupée en cartes de caractéristiques sont utilisées pour la détection et la classification des objets observés,

– les objets d'intérêt sont segmentés en utilisant des règles de groupement perceptuel. A ce sujet voir les différentes contributions concernant la théorie de la forme [ref Stephen E. Palmier : Les théories contemporaines de la perception de Gestalt].

Différents modèles théoriques du système visuel ont été développés, allant de la modélisation de V1 (Bednar, 2002) à une théorie unifié des systèmes de vision et de cognition (Grossberg, 2007). Ces modèles permettent de mieux comprendre le fonctionnement du système visuel mais ne permettent pas (par leur complexité ou leur abstraction) la mise en oeuvre pratique d'un système de vision temps réel au comportement réaliste.

Il est à noter que notre système visuel, malgré sa grande complexité et son impressionnante efficacité n'est pas sans défaut. On peut ainsi mettre en évidence différentes illusions mettant en avant certains *disfonctionnements* de notre vision. Ces illusions sont également une grande source d'études car elles permettent de mieux comprendre les mécanismes du système visuel (figure 1).

Dans ce chapitre nous avons évoqué les principales caractéristiques du système visuel humain. Dans le chapitre suivant nous allons présenter très succinctement les principaux principes des systèmes de vision par ordinateur *classiques*.

2.2. *La vision par ordinateur*

Parallèlement au développement de modèles théoriques du système visuel, des systèmes informatiques capables de réaliser des tâches visuelles avec des niveaux de performance approchant ceux de l'humain ont été conçus. Cependant, dans cet optique, il est nécessaire de concevoir des systèmes hautement spécialisés et devant opérer dans des environnements spécifiques.

Le paradigme de la vision par ordinateur proposé par David Marr (Marr, 1982) (figure 2), encore largement utilisée de nos jours, était bio-inspirée. Cependant cette approche totalement ascendante se basait sur les connaissances de l'époque qui étaient moins complètes que celles dont nous disposons actuellement.

Des approches plus récentes ont tiré partie de notre meilleure compréhension du système visuel. Ainsi, de nombreux systèmes tirent avantage de l'attention visuelle pour diriger leurs tâches de reconnaissance. C'est le cas du système robotique VOCUS (Frintrop, 2006), du système de détection de nouveauté de (Ban *et al.,* 2006) ou de l'architecture attentionnelle de (Itti, 2000).

Les aspects traitements massivement parallèles et hiérarchique de notre cerveau ont également été la source d'inspiration de systèmes tels que le neocognition (Fukushima, 1988) ou les réseaux à convolution (LeCun *et al.,* 2001), permettant de résoudre plus efficacement certains problèmes de reconnaissance d'objet.

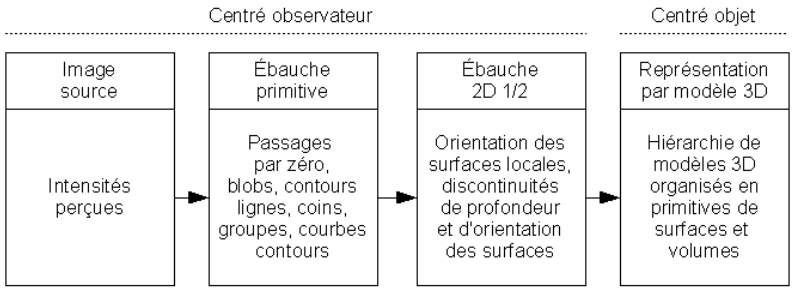

Figure 2. *L'architecture de vision* classique *proposée par David Marr.*

Cependant, ces différentes approches restent spécialisées et négligent une partie importante de notre système visuel (au sens large) : la plasticité de notre chaîne de traitement des informations visuelles, ses nombreuses rétroactions ainsi que notre grande capacité d'apprentissage, tous deux guidés par notre *instinct du savoir* (Perlovsky, 2007). Dans la partie suivante, nous proposons un système de vision comportemental dont l'objectif est d'essayer de combler ces lacunes.

3. Un système de vision comportemental temps réel réactif et adaptatif

3.1. *Pour quoi faire ?*

Les approches existantes sont soit des modèles essayant d'expliquer le fonctionnement de tout ou partie du système visuel humain (sans souci du temps réel), soit des systèmes prévus pour réaliser une tâche spécifique en exploitant certaines propriétés de la vision humaine permettant d'effectuer plus efficacement leur traitement (parfois temps réel).

Nous souhaitons développer un système complet qui puisse s'inspirer des mécanismes mis en place par l'évolution pour le système visuel humain afin de créer un être virtuel qui puisse réagir en temps réel à des stimuli visuels. L'objectif n'est en aucun cas de créer un clone du système visuel humain complet. En effet, outre la complexité et la capacité de traitement nécessaires, notre démarche est centrée sur les mécanismes adaptatifs de la vision humaine.

Nous proposons un système capable de générer ses propres représentations du monde à partir des stimuli visuels auxquels il aura été confronté. Il pourra également réagir à ces stimuli et s'adapter en fonction de leur adéquation avec ses *instincts* primaires / perceptions antérieures. Les caractéristiques communes avec l'humain que nous avons choisi de mettre en oeuvre sont les suivantes :

– Instinct de la connaissance (Perlovsky, 2007). Le système aura comme besoin fondamental l'information. Ce sera sa *nourriture*. Cet instinct n'est pas clairement défini chez l'homme, mais de nombreux indices laissent penser que cela pourrait être un de nos instincts primaires, comparable à celui d'absorption de nourriture ou de reproduction.

– Mécanismes de réduction de la charge de traitement permettant de réagir en temps réel. Cela passe en grande partie par l'utilisation des mécanismes attentionnels permettant de focaliser les ressources disponibles sur une tâche ou un objet précis.

– Adaptation du système en fonction de l'environnement. Selon les principes du darwinisme neuronal (Saulnier, 2003), différents mécanismes de sélection permettront au système de s'adapter, notamment grâce à l'affaiblissement ou au renforcement de certains circuits de traitements en fonction des interactions avec l'environnement visuel.

– Construction d'un espace d'état permettant une représentation interne cohérente de l'environnement visuel de l'être virtuel. La construction de cet espace de représentation pourrait être guidée par les règles de la théorie de la Forme (Palmer, 1999). Cette représentation devra être cohérente, mais ne devra pas forcement correspondre à celle d'un être humain.

– Emergence de réactions basiques (étonnement, peur, intérêt, etc.) en réactions aux changements dans l'environnement visuel de notre être virtuel.

– Enfin, en l'absence de stimulus visuel : réorganisation des connaissances acquises. Le système de pouvant pas organiser toutes les informations reçues en temps réel il se réorganise de manière *offline* (équivalent du sommeil chez l'homme).

La création de cet *être virtuel* répond à deux problématiques distinctes : créer un système visuel artificiel temps réel et adaptatif qui soit capable de construire une certaine représentation cohérente du monde ; exploiter cette représentation dans un système cognitif basique permettant l'émergence de comportements simples (peur, etc.).

3.2. *Bio-inspiré mais pas trop !*

Pour atteindre les objectifs que nous nous sommes fixés, il est nécessaire de s'inspirer des meilleures solutions trouvées par la nature grâce aux mécanismes de l'évolution, sans oublier que la mise en oeuvre pratique de notre système sera réalisée sur un ordinateur standard, qui est une machine principalement sérielle. En effet, bien que nos ordinateurs soient maintenant capables d'exécuter plusieurs milliards d'opérations à la seconde, leur architecture est totalement incapable de gérer la complexité des 150 millions de neurones et les 150 milliards de connexions de l'aire V1 du cortex visuel. Il sera donc plus judicieux de s'inspirer de l'architecture et de certains mécanismes biologiques plutôt que d'essayer (en vain ?) d'imiter la nature.

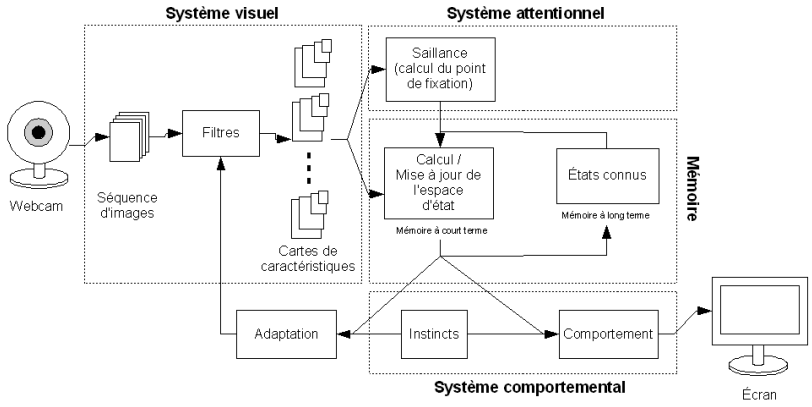

Figure 3. *Architecture du système de vision comportemental réactif et adaptatif.*

3.3. *Architecture*

La figure 3 représente l'architecture complète du système de vision. Celle-ci schématise les différents traitements allant de la capture d'image via une webcam à l'expression du comportement issu des stimuli visuels sur l'écran de l'ordinateur. Nous envisageons dans un premier temps de mettre en oeuvre cette architecture de la manière suivante :

– le calcul des cartes de caractéristiques sera effectué sur la base de l'architecture proposée par (Itti, 2000) et modifiée par (Frintrop, 2006) à laquelle nous ajouterons des cartes de symétrie et une partie adaptative (via la boucle de rétroaction *adaptation* sur le schéma) ;

– la fusion des cartes de saillance sera effectuée grâce à un algorithme Proie-Prédateur. Le point de fixation obtenu déterminera la zone à analyser afin de mettre à jour l'espace d'état (mémoire à court terme). Les espaces d'état *nouveaux* seront intégrés à la base des espaces connus (mémoire à long terme) ;

– l'espace d'état sera construit entre-autre en utilisant les règles de la théorie de la forme à partir des cartes de caractéristiques (traitement ascendant). Les états connus couplés aux instincts seront utilisés comme prégnance (traitement descendant) ;

– différents comportements (peur, surprise, etc.) pourront émerger en fonction de l'adéquation de l'espace d'état avec les états connus ainsi que de la satisfaction des instincts par l'espace d'état ;

– enfin, les filtres d'entrée permettant le calcul des cartes de caractéristiques seront adaptés en fonction de la satisfaction des instincts par l'espace d'état.

La conjonction de ces différentes interactions devrait permettre de générer un système réactif et adaptable aux différents stimuli auxquels il est soumis.

4. Conclusion

Nous avons présenté dans cet article un aperçu des limitations des systèmes actuels de vision et émettons l'hypothèse qu'il est nécessaire de changer notre manière de concevoir les systèmes de vision pour que ceux-ci puissent à terme répondre aux besoins de facilité d'interaction avec les applications interactives de demain. Ce constat nous a amené à introduire un système de vision comportemental qui sera à même d'apprendre de s'adapter et de réagir de manière autonome. Nous avons présenté quelques pistes concernant les détails de son architecture, mais l'essentiel du travail reste à faire. Il nous reste à faire de ce système une réalité permettant l'interaction en temps réel avec un utilisateur...

5. Bibliographie

Ban S.-W., Lee M., « Selective attention-based novelty scene detection in dynamic environments », *Neurocomputing*, vol. 69, n° 13-15, p. 1723-1727, 2006.

Bednar J. A., Learning to see : genetic and environmental influences on visual development, PhD thesis, University of Texas, 2002. Supervisor-Risto Miikkulainen.

Behnke S., Hierarchical Neural Networks for Image Interpretation, PhD thesis, Freie Universitat Berlin, November, 2002.

Frintrop S., VOCUS : A Visual Attention System for Object Detection and Goal-Directed Search, PhD thesis, University of Bonn, 2006.

Fukushima K., « A Neural Network for Visual Pattern Recognition », *Computer*, vol. 21, n° 3, p. 65-75, 1988.

Grossberg S., « Towards a unified theory of neocortex : laminar cortical circuits for vision and cognition », *in* , T. D. Paul Cisek, , J. F. Kalaska (eds), *Computational Neuroscience : Theoretical Insights into Brain Function*, vol. Volume 165, Elsevier, p. 79-104, 2007.

Itti L., Models of Bottom-Up and Top-Down Visual Attention, PhD thesis, California Institute of Technology, Pasadena, California, Jan, 2000.

LeCun Y., Bottou L., Bengio Y., Haffner P., « Gradient-Based Learning Applied to Document Recognition », *Intelligent Signal Processing*, IEEE Press, p. 306-351, 2001.

Marr D., *Vision : a computational investigation into the human representation and processing of visual information*, W. H. Freeman, San Francisco, 1982.

Matthieu Perreira Da Silva Vincent Courboulay A. P., Estraillier P., « Real-Time Face Tracking for Attention Aware Adaptive Games », *in* , M. V. A. Gasteratos, , J. Tsotsos (eds), *ICVS 2008*, number 5008 in *LNCS*, Springer-Verlag, p. 99-108, May, 2008.

Palmer S. E., « Les théories contemporaines de la perception de Gestalt », *Intellectica*, vol. 28, n° 1, p. 53-91, 1999.

Perlovsky L. I., Neural Dynamic Logic of Consciousness : The Knowledge Instinct, Report n° ADA472303, AIR FORCE RESEARCH LAB HANSCOM AFB MA, September, 2007.

Saulnier B., Le darwinisme neuronal de Gerald M. Edelman, Rapport, Juin, 2003.