LES RENCONTRES SCIENTIFIQUES
— COLAS —

Les Rencontres Scientifiques Colas

« L'œil humain va-t-il être détrôné par l'ordinateur ?»

21 mars 2006

avec **Olivier FAUGERAS,**
Membre de l'Académie des Sciences, Directeur de Recherche à l'INRIA

et **Frédéric GUICHARD**,
Directeur de la Recherche, DxO Labs, Boulogne

Conférence animée par **Marie-Odile MONCHICOURT**

La vision humaine, sa précision, sa pérennité, l'intelligence qui la sous-tend ont toujours été un sujet de fascination. On essaie depuis très longtemps d'imiter cette vision ou de la transcender par des procédés très élaborés. Ainsi en est-il de la photographie aérienne ou satellitaire, de l'imagerie médicale, de l'astronomie et de bien d'autres domaines. Comment, depuis quarante ans, la vision algorithmique a-t-elle réussi à donner un oeil à l'ordinateur ? Comment, chemin faisant, a-t-elle pu mobiliser une grande partie de l'algorithmique contemporaine et a-t-elle aussi mis au point, pour ses besoins propres, une foule d'outils raffinés jusqu'au hardware dédié pour atteindre le temps réel ? Où en sommes-nous ? Quels progrès pourraient émerger du rapprochement avec la vision humaine ?

Intervention de Monsieur Olivier FAUGERAS

La vision biologique

En préambule à cette intervention, il semble nécessaire de dresser un rapide historique des recherches menées dans le domaine de la vision et tout d'abord de la vision biologique, celle de l'œil humain.
Herman Von Helmholtz est un des premiers chercheurs à s'y intéresser de près au XIX^{e} siècle. Sa démarche est à la fois celle d'un physicien et celle d'un psycho-physicien. Son idée principale est que notre perception visuelle est conduite par un modèle cérébral. L'esprit essaye de faire coïncider cette modélisation avec ce qui est perçu par les sens, en l'occurrence par la rétine. L'imperfection de ce processus de « placage » est trahie par les fameuses illusions optiques et donc le fait que le cerveau effectue un traitement particulier pour donner du sens à ce que lui communique la rétine. Connu depuis longtemps, ce phénomène est pourtant toujours aussi mal compris et demeure donc mystérieux. Il y a donc bien évidemment une différence entre la perception et la physique de formation de l'image. Le cerveau va essayer de plaquer un modèle tridimensionnel.

Ces recherches sont approfondies au XX^{e} siècle par l'école de la « Pattern Theory » ou théorie de la forme, avec notamment les chercheurs Richard Gregory et Ulf Grenander qui développent des idées d'inférence statistique du cerveau sur le stimulus (principe de maximum, test d'hypothèses).

On peut encore citer l'apport de mathématiciens tels que David Manford, lui aussi originaire de l'Université de Brown aux Etats-Unis.

S'intéresser à la perception, c'est aussi s'intéresser à la question de l'attention, problème mal connu, voire méconnu. Or, il y a une focalisation consciente et inconsciente de l'attention car nous ne sommes pas en mesure de décrypter en temps réel la totalité des bits d'une image, d'un film. Helmholtz utilise à ce sujet la métaphore du « rayon lumineux » qu'on viendrait braquer sur une partie de l'image pour mieux l'analyser.

Et nous disposons précisément d'outils nous permettant d'observer le cerveau aux différentes échelles à travers lesquelles l'information est transmise dans notre système biologique, et d'appréhender une partie du fonctionnement du néo-cortex. Les zones visuelles, particulièrement complexes et encore une fois mal connues, font l'objet d'une spécialisation (qualifiée par les spécialistes à l'aide d'indicatifs : V1, V2, V3). Elles échangent des flux d'informations mais chacune possède une fonction spécifique : l'une va ainsi permettre la reconnaissance de scène, d'autres la reconnaissance de visage.

Au-delà de cette spécialisation, c'est aussi le mécanisme permettant de rapprocher l'objet de la vision du lieu où il se trouve qui pose problème et n'est pas encore compris. Au moins, les techniques à notre disposition et notamment l'IRM fonctionnel nous permettent de mettre en évidence, avec plus ou moins de précision certes, les zones impliquées dans la perception visuelle. On voit ainsi très nettement lorsqu'on les soumet à un stimulus les parties du cerveau qui s'animent.

La vision computationnelle

Beaucoup plus récente que l'étude de la vision biologique, celle-ci fait l'objet depuis les années 60 de l'attention d'informaticiens, de spécialistes du traitement du signal mais aussi de spécialistes des neurosciences et de mathématiciens à partir des années 70.

David Mar expose sa théorie dans « Vision » et s'opère avec lui une rupture de paradigme. Il développe notamment deux idées particulièrement intéressantes qui bien qu'en partie erronées, vont permettre de faire avancer notre compréhension du domaine. Selon lui, la perception visuelle est un problème de traitement de l'information et est indépendante de l'organisme sur lequel s'effectue ce traitement.

Cela nous semble bien sûr aujourd'hui faux - l'organisme interfère- mais cela nous a permis de nous abstraire du substrat biologique et d'étudier séparément les fonctions visuelles.

Quelques exemples de l'état de l'art en vision computationnelle. On trouve, sur le plan des techniques mises en œuvre, ce qu'on qualifie de « segmentation d'images ». Cette technique permet notamment de dissocier la forme du fond. On procède par déplacement dans l'image d'un contour de façon à minimiser une certaine énergie (« lagrangien » en termes mathématiques). On espère qu'à l'équilibre le contour est dans un minimum absolu et on détecte de manière précise à ce moment-là le contour de l'objet. Technique récente basée sur la théorie des équations dérivées partielles et la géométrie riemannienne.
Deux autres techniques : la vision 3D (plus technologique) - à partir d'un ensemble d'images on détermine la forme tridimensionnelle d'un objet - et aussi l'idée de suivre un modèle et de reconnaître un élément à partir de données insérées dans un programme.

Intervention de Monsieur Frédéric Guichard

Après cette « vision » académique, je vais vous proposer d'aborder le sujet sous l'angle industrie à travers deux volets : la capture de l'image d'une part et la compréhension de l'image d'autre part.
Prenons l'exemple du domaine des appareils photographiques. Un appareil photographique se compose avant tout de trois éléments : l'objectif, le capteur, le dispositif de traitement d'image. Avec le passage au numérique, on observe qu'au sein de ces appareils, le niveau traitement d'image prend une place de plus en plus importante et surtout permet de minimiser ou d'économiser les autres éléments. Entre deux appareils photos munis de lentilles similaires, de capteurs identiques, il peut pourtant exister des différences très importantes de qualité de la photo si le dispositif de traitement d'image embarqué varie.
En dehors de l'amélioration de la qualité d'image, le perfectionnement du traitement permet aussi de diminuer très sensiblement la place du hardware au sein de l'appareil grâce à une grande ouverture de diaphragme.
Ce marché est aujourd'hui mature. Le premier « caméraphone » est apparu en 2002 ; plus de 450 millions d'unités ont été vendues dans le monde en 2005…Nokia est aujourd'hui le premier fabricant d'appareils photos et cumule deux fois plus de ventes que Canon, Nikon et Olympus réunis…

La technique ne bride plus le développement du traitement de l'image puisque nous disposons aujourd'hui d'une puissance de calcul largement suffisante pour ce que nous savons en faire aujourd'hui (10 milliards d'opérations seconde pour les prochains caméraphones). Un tera d'opérations par seconde n'occasionnerait qu'un surcoût de 50 dollars.
En conclusion de cette partie de l'intervention consacrée au traitement d'images, on peut donc dire qu'il est aujourd'hui mature au point de repousser les limites du hardware.

Abordons maintenant la question plus difficile de la compréhension de l'image par l'ordinateur. Je prendrais pour cela l'exemple d'un logiciel que nous avons créé à usage des piscines publiques, le dispositif « Poséidon ». Ce système capture en permanence des images, les analyses et tente de détecter les possibles noyades au moyen d'alertes envoyées aux maîtres nageurs. Le système paraît assez simple en termes de fonctionnalité et pourtant, ce sont déjà des requêtes complexes en raison de certains facteurs : variabilité des scènes (comment parmi la multiplicité des scenarii possibles indiquer lequel est porteur de danger) mais aussi variabilité des images en fonction des moments de la journée (évènements extérieurs avec variabilité de la lumière pénétrant la piscine et induisant des ombres, mise en fonctionnement de spots et d'éclairages internes qui modifient la perception, etc).

Face à ces difficultés, comment fonctionne le système Poséidon ? Poséidon fait intervenir un réseau de caméras qui vont permettre de faire de la reconstruction 3D. On va ainsi pouvoir discriminer entre les objets et les ombres. Deux caméras vont en effet distinguer une ombre de la même manière (surface « plate ») tandis qu'elles auront une perception différente en deux points différents d'un objet. Cela suppose bien sûr d'intégrer un modèle 3D précis du fond de la piscine et de l'emplacement/orientation exacte des caméras.
Au final, on peut estimer qu'il s'agit d'une application très simple pour laquelle on est obligé de mettre en œuvre une débauche technologique…Aujourd'hui, environ 100 piscines sont équipées de Poséidon.

La principale difficulté dans la traduction en langage informatique réside dans le fait que notre démarche est conceptuelle tandis que celle de l'ordinateur ne l'est pas. Comment dire à l'ordinateur « il s'agit d'une chaise », « une chaise à 4 pieds », autant de concepts que ne possède pas l'ordinateur, On parvient donc à traduire en règles informatiques quelques

concepts mais la démarche reste aujourd'hui sectorielle et non globale, pour des produits de niches et des applications très circonscrites et simples.