

# La vision située : un concept pour faciliter l'autonomie des systèmes

The situated vision : a concept to facilitate the autonomy of the systems

Bertrand Zavidovique<sup>1</sup>, Roger Reynaud<sup>2</sup>

<sup>1</sup>F91405 Orsay cedex

<sup>2</sup>IEF, Bât 220, DIGITEO-labs, Université Paris Sud

Manuscrit reçu le 8 juillet 2006

## Résumé et mots clés

Cet article présente notre approche des techniques par lesquelles la vision doit contribuer à l'autonomie d'un système. Il établit un parallèle entre deux aspects de l'évolution des systèmes de perception. La première partie explique comment une conception système des problèmes de la vision (dénommée *vision active*) aide à les résoudre efficacement. Il y apparaît la nécessité d'anticiper, conduisant à une première définition de la vision située et cataloguant un ensemble des situations qui constitue la description fonctionnelle de l'exo-système. La seconde partie montre comment les descriptions architecturales des systèmes de fusion s'adaptent aux caractéristiques à extraire, et donc comment un modèle émerge progressivement au service d'un contrôle opportuniste vers plus d'intelligence (au sens capacité à utiliser conjointement des renseignements issus de diverses modalités). Nous pensons que l'essentiel des développements à court et moyen terme proviendra des avancées autour de cette approche duale et du concept de « systèmes autonomes », dans lesquels les flots internes de communication apparaissent alors plus utiles à contrôler que les tâches locales, supposées accomplies de manière satisfaisante. Un schéma de système de contrôle de flux est enfin proposé pour mettre en œuvre au plus haut niveau du modèle un contrôleur de commutations entre différentes situations. *Viser des tâches plus ambitieuses s'appuie sur des descriptions et un contrôle plus abstraits.*

**Vision active, Système autonome, contrôle opportuniste, architecture de fusion.**

## Abstract and key words

This paper is about vision to make systems more autonomous. We parallel two aspects of the current evolution of System-Perception: more ambitious yet coherent tasks tightly rely on more abstract description and control. As vision is likely to be a major and complex sensory modality for machines as it is for most animals, we concentrate our development on it. In the first part we show how thinking to systems helped to better pose vision problems and solve them in a useful manner. That is the "active vision" trend that we explain and illustrate. Along the same line, the necessity for anticipation shows further, leading to a first definition of "situated vision". The second part deals with how to design systems able to achieve such vision. We show from a few examples how architectural descriptions evolve and better fit important features to grasp – a model – in view of more efficient control towards intelligence. Inner communication flows are better be controlled than local tasks that should be assumed completed efficiently enough in all cases. We conclude with a plausible sketch of a system to be experimented on in situations that require some autonomy.

**Active vision, Autonomous Systems, opportunist control, fusion architecture.**

# 1. Systèmes autonomes : la métaphore de l'orchestre

Il est admis que des monstres préhistoriques comme le brontosaure avaient un cerveau de la taille d'une orange pour le volume corporel d'une demi-douzaine de bus ; ils étaient vraisemblablement assez lents et pas très bien coordonnés à cause des difficultés de faire passer un message de la tête aux pieds. À l'opposé, considérons un orchestre philharmonique de volume comparable : il ne suffit pas que ses membres jouent de manière satisfaisante en respectant le tempo et sans fausse note. À un moment donné, les exécutants se doivent de jouer ensemble, de véritablement faire de la musique : cela tient d'une alchimie extrêmement complexe à expliquer si ce n'est pour le rôle primordial que les communications jouent pour favoriser cette réelle communion.

Du point de vue système (Figure 1), un orchestre est un système multi-acteurs qui transforme un ensemble d'énergies humaines en un son mélodieux sous le contrôle, distribué, des musiciens et la volonté, centralisée, du chef d'orchestre. Les yeux jouent évidemment un rôle majeur à côté du capteur principal pour cette tâche qu'est l'oreille. Certains membres, *i.e.* opérateurs de cette machine parallèle, peuvent avoir des instruments plus anciens que d'autres avec des timbres différents, d'autres membres peuvent avoir des réactions plus ou moins rapides ; ils ont sûrement des ressentis différents quant au morceau joué ou à la façon de l'interpréter. L'un d'entre eux peut avoir une rage de dent ce jour-là, etc. Mais en fin de compte, ils produiront

ensemble le son unique d'un instrument incomparablement plus sophistiqué et complexe que chacune de ses composantes. Comment font-ils ? Ils mobilisent tout capteur à leur disposition non seulement pour écouter la musique produite, ou suivre la partition, mais aussi pour comprendre et compenser localement les variations qui pourraient ne pas contribuer à l'objectif : les oreilles recherchent toute perturbation dans la sortie produite localement par eux-mêmes ou un voisin de leur groupe, tout en percevant l'impression globale. Les yeux vérifient le rythme en observant l'attitude des voisins, ou la posture du chef d'orchestre et sa gestuelle. Plus précisément, ils communiquent par leur propre expression corporelle, au travers de contacts visuels ou de modulations sonores comme l'attaque ou l'intensité des sons produits : non seulement ils sont en mode de perception active dans le sens où ils poursuivent toute information pertinente, mais encore ils aident les autres à le faire. Il ne faut pas non plus oublier le rôle que jouent les répétitions dans la connaissance réciproque et dans la façon dont le chef d'orchestre code ses messages pour guider vers le but ultime du système : l'entraînement a rendu plus facile l'interprétation de tous les signaux de communication interne en construisant un langage commun plus abstrait.

Le but de cet article n'est pas uniquement de décrire les facteurs supposés du comportement intelligent de la machine (forme d'intelligence et localisation du traitement correspondant), la question est surtout de savoir décrire le système lui-même jusqu'à établir un lien explicatif entre les descriptions comportementales (description externe) et architecturale (description interne) du même système. Il faut alors répondre aux questions suivantes : comment le chef d'orchestre agit-il pour contrôler l'orchestre et sur quelle représentation se base-t-il ? Comment relier la sortie, une forme d'harmonie, à la description du sys-

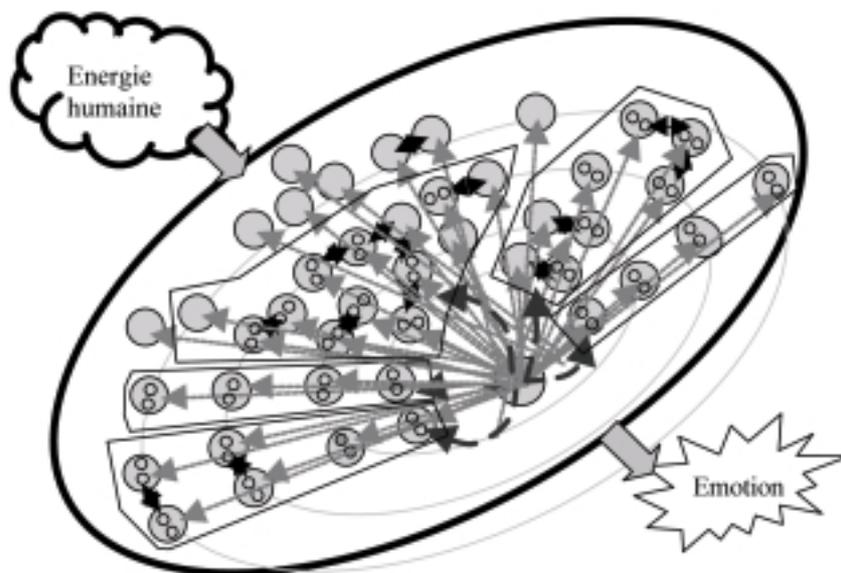


Figure 1. Une vue de l'orchestre fonctionnant en système autonome durant un concert.

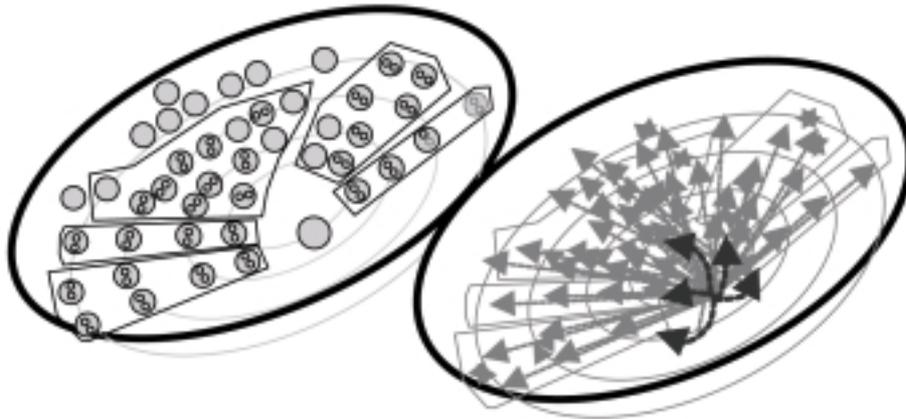


Figure 2. Représentations duales de l'architecture d'un orchestre : machine parallèle composée d'acteurs à gauche et flux<sup>1</sup> de communication à droite.

tème, contrôleurs compris. On conçoit que connaître la méthode d'apprentissage de chaque instrument soit de peu de secours, ni de comprendre la physique de génération des sons d'un cuivre, d'une percussion ou d'une corde, ni de savoir l'état de santé du second violon. À ce niveau de complexité, la description ne peut plus concerner la collection des composants du système. Une première étape vers un contrôle efficace suppose deux conditions : *chaque partie constitutive du système accomplit sa propre tâche au mieux ; et ce mieux est suffisant*. Si une des deux conditions n'est pas respectée, alors l'objectif ne sera pas atteint, et optimisation ou contrôle ne sont plus d'aucune utilité. À l'inverse une description axée sur les flux d'information (Figure 2), si elle maintient des critères d'efficacité locale, devrait être plus adaptée au contrôle du comportement global. *Plus la description complète sera concise et abstraite, plus le contrôle sera ambitieux*.

Il est donc vraisemblable que décrire un système qui offre de telles capacités de coordination au service d'une performance donnée ne puisse se faire en termes du fonctionnement correct ou non de tel ou tel composant. Sur le plan de la complexité également, cette nécessaire hypothèse que chaque partie remplit sa tâche est intéressante : elle implique que le supplément d'activité rendant le comportement perçu intelligent ne serait alors qu'une affaire de communications internes et de leur contrôle. De plus si la représentation du système devient abstraite, elle décrira mieux l'activité d'un système plus abstrait et général. Ainsi expliquer comment un individu s'efforce de survivre en toute circonstance est mieux fait en terme de courir (mouvement rapide), se cacher (fusion dans l'environnement) ou se battre (réaction vive) plutôt qu'en terme de déplacements précis des articulations de chaque membre dans l'espace géométrique de la tâche, à trois dimensions, lui-même projeté de l'espace sensoriel, virtuel à quatre dimensions.

Cet article dresse un parallèle entre deux aspects de l'évolution actuelle des systèmes pour viser des tâches plus ambitieuses en

leur associant des descriptions et un contrôle plus abstraits. Ce contrôle détermine un alignement explicite d'une situation externe (décrite en termes d'objets, d'interrelations, de changements) sur une situation interne (décrite en termes de modules d'acquisition, traitement ou décision, de réseau et de système d'exploitation). L'ensemble des alignements construit une mise en correspondance de l'exo-système {ensemble des situations externes et fonctionnalités correspondant aux spécifications de l'application} sur l'endo-système {ensemble des situations internes et configurations possibles de la machine nécessaires au caractère opérationnel de l'application}. Le présent article établit la conjecture suivante : *la mise en correspondance réalise la commande du système tandis que sa mise à jour est l'objectif même de son contrôle*. Nous en dérivons une proposition de mise en correspondance pour les systèmes de vision située, où le contrôle de haut niveau est opportuniste.

Dans le chapitre 2, l'évolution des machines de vision montre comment une approche système a permis de mieux poser les problèmes de vision pour les résoudre efficacement. Ce constat montre la pertinence de mécanismes d'anticipation, et conduit à la définition de l'exo-système ciblé dans les applications de vision située. Le chapitre 3 traite des différents points de vue utilisés pour concevoir un système et amène à poser la conjecture décrite précédemment. Le chapitre 4 aboutit à la proposition d'un schéma réel de mise en correspondance à partir de descriptions des relations exo/endo-systèmes en fusion multi capteurs proposés progressivement pour rendre explicite un modèle, au service d'un contrôle plus efficace.

1. Toutes les flèches de communications locales ne sont pas indiquées : le violoniste entend plus les violons voisins que la flûte à l'opposé.

## 2. Fonctionnalités accessibles et définition de l'exo-système

### 2.1. Perspectives historiques

Dans les années 80, les systèmes de vision étaient pour la plupart orientés reconstruction de l'environnement, essentiellement sous forme géométrique. La modélisation 3-D est alors le point central dans tous ces systèmes avant qu'une décision puisse être prise (mouvement de caméras, modification de paramètres, ...): c'est la méthode MIT [24]. Selon la théorie de la vision précoce et des croquis fondamentaux, due à David Marr, la reconstruction volumique est fondée sur la forme à partir de contour, ombre ou texture et le mouvement sur le flot optique. Notons qu'à ce jour encore aucun système opérationnel n'a jamais fonctionné sur ces principes. Tous les systèmes opérationnels connus ([1] à [22]) procèdent de manière quasi diamétralement opposée: la configuration 3-D est en quelque sorte un sous produit d'une boucle perception action, à utiliser si nécessaire. Un tel miracle apparent provient du caractère actif de la vision impliquée dans le comportement global de ces systèmes.

En effet, la plupart des problèmes de vision et donc de perception sont mal posés au sens d'Hadamard parce qu'ils supposent d'inverser de multiples projections: acquisition dans le plan focal, échantillonnage sur des entiers, segmentation selon des caractéristiques, etc. D'autres problèmes comme la navigation dans un monde rigide sont bien posés mais instables. Plutôt que de régulariser avec des contraintes telles que la variation continue bornée, hypothèse contestable pour des variables mesurées à partir d'une scène, les ingénieurs préfèrent contraindre les capteurs dans la mesure où ces derniers doivent être contrôlés de toute manière. Un tournant s'opère dans les années 90 [23]. C'est ainsi qu'Aloimonos [24] a commencé en 1988 de formaliser l'idée et tester sa validité sur quatre problèmes à la mode de cette époque: reconstitution des formes à partir de leurs ombres, contours ou texture (mal posé dans un contexte passif), et structure à partir du mouvement (instable dans un contexte passif). Un effort similaire est réalisé en 1992 par Clark [56] qui calcule une carte de profondeur par stéréo photométrie où la source de lumière bouge de façon contrôlée. Le concept de perception active considéré ici est plus proche de celui de Bajcsy développé en 1985 [25] et nos propres idées [26,27,28] de résolution de problèmes par une stratégie de contrôle des capteurs: ceci impose l'explicitation des objectifs perceptifs et le choix des critères. Aujourd'hui, cette approche englobe plusieurs tendances:

- **La vision active**: une analyse plutôt théorique du processus de vision issue des travaux d'Aloimonos pour optimiser des tâches visuelles ou guidées visuellement [24,29,30]. Comprendre une structure à partir d'un mouvement contrôlé [31, 32] appartient à

cette catégorie dans la mesure où le mouvement des caméras aide à optimiser la reconstruction 3-D.

- **La perception active**: une étude des stratégies de perception, incluant la coopération entre capteurs et traitements du signal pour parvenir à une meilleure connaissance de l'environnement [33,34,35]. Ce concept s'étend à toute la fusion de données capteur lorsque l'information rassemble différents points de vue, différents principes physiques de capteurs, et plusieurs pas de temps. D'autres exemples s'attachent à définir des stratégies [36,37], concevoir des superviseurs [38], ou étudier l'impact des incertitudes de l'acquisition sur la boucle de rétroaction [39].

- **La vision animée** [40]: basée sur l'analyse de la perception humaine [41,42], elle a pour but d'imiter la fixation et le contrôle du regard. L'objectif de Ballard est alors d'optimiser la complexité algorithmique, et dans ce but, une idée intéressante est de considérer des coordonnées exocentriques plutôt qu'égocentriques, invariantes aux mouvements de l'observateur. On trouvera dans [43,44,45,46] d'autres résultats dans cette catégorie.

- **La vision intentionnelle** [47]: un autre point de départ est l'évidence que la plupart des systèmes de vision ont à répondre à des questions précises et sur un nombre limité de tâches connues. Dans ce contexte, seule l'information pertinente relative à la tâche courante doit être extraite. Le concepteur se pose alors les questions suivantes dans l'ordre: Dans quel but le système est-il construit? Quelle connaissance est nécessaire pour atteindre ce but? Comment puis-je obtenir cette connaissance? Une analyse étendue et une discussion de ce qui semble une approche encore re-constructionniste sont dans [48,49,50].

### 2.2. Premières applications

Il n'y a pas de doute que la technologie favorise cette évolution en miniaturisant, et donc accélérant, les capteurs et les processeurs, et en faisant varier les phénomènes physiques sous-jacents pour fournir de nouveaux degrés de liberté (ex.: capacité à ajuster la géométrie et l'optique simultanément). Le rôle clé de la technologie dans le succès de l'approche de la «perception active» au cours de la dernière décennie est prouvé par l'explosion des têtes binoculaire et trinoculaire pour expérimenter dans le domaine de la localisation et le suivi par triangulation, disparité ou autre [51] à [67].

Pour comprendre mieux le niveau exact de réalisme autorisé par ces techniques, nous listons quelques exemples représentatifs dans différentes applications, rangés par complexité croissante.

**Description quantitative de surface**: le contrôle de la position de l'axe optique donne accès aux caractéristiques géométriques de la surface (courbure locale et axes principaux) à partir de vues multiples grâce au théorème de Blaschke [68,69].

**Reconstruction 3-D en présence d'occlusions**: un capteur constitué d'un laser rigidement couplé à une caméra balaye un plan. Sa géométrie connue donne accès à la profondeur de tous les objets de la scène. Les occlusions génèrent des zones

aveugles soit parce que le rayon laser n'atteint pas la surface soit parce que la réflexion de ce rayon n'atteint pas la caméra. Une rotation plane résout le premier type tandis qu'une rotation du support à l'intérieur du plan résout le second type.

**Contrôle visuel de systèmes:** la vision peut servir dans une rétroaction classique de deux façons [70] :

- Regarder/bouger : la position relative d'une caméra par rapport à un objet (amer occasionnel) est mesurée à partir de l'image traitée et prise pour entrée. La différence estimation/but est alors minimisée en bougeant la caméra.

- Servomécanisme visuel: l'objectif prend en compte directement les caractéristiques extraites des images (ex. : une caméra évolue pour aligner des points d'intérêt spécifiques). Cette technique est rapide mais nécessite la matrice d'interaction faisant correspondre les caractéristiques de l'image avec les mouvements de la caméra, laquelle peut ne pas être complètement calculable.

La recherche dans ce domaine est encore vivace en ce qui concerne l'efficacité des caractéristiques, les lois de contrôle et leur robustesse.

**Recherche indirecte :** l'autonomie impose aux robots de localiser des objets de moindre intérêt ou d'intérêt temporaire, étant donné la phase courante de la mission. Une caméra grand champ a peu d'intérêt en face de petits objets et les déforme, une caméra champ étroit peut mettre beaucoup de temps pour balayer tout le champ de vision. Wixon [71,72] a proposé un schéma de recherche général fondé sur les relations spatiales entre objets intermédiaires significatifs, faciles à pointer en basse résolution, les objets petits ou moins en évidence ne pouvant être découverts qu'en haute résolution. Trouver le premier grâce à sa taille ou sa couleur réduit l'espace de recherche du dispositif à champ étroit. En comparant les efficacités, un gain de huit en rapidité peut être obtenu. D'autres relations structurales fondées sur la physique par exemple sont considérées [73]

**Information à partir de l'interprétation:** TEA1 [74,75] et Dynamic-TEA1 répondent aux questions du type «Qui vient dîner ce soir?». Si un invité distingué est attendu, plusieurs verres sont disposés par personne sur la table. Pour confirmer une conjecture, le système rassemble sélectivement des données image en résolution variable et les traite en conséquence. La connaissance *a priori* est codée sous forme de réseau bayésien hiérarchique. L'interprétation se déroule en cinq étapes :

- Tester la liste des actions
- Exhiber la plus adaptée au sens d'un critère Gain/Coût
- Réaliser l'action
- Mettre à jour le réseau bayésien
- Itérer jusqu'à atteindre une confiance prédéfinie dans la réponse à la question.

Ce dernier exemple est parmi les plus sophistiqués. Il confirme que le meilleur compromis entre traiter et contrôler autorise aujourd'hui les systèmes à échapper à la *malédiction* de la

reconstruction préalable de l'environnement. Elle est remplacée par une boucle perception-action comme dans un servomécanisme visuel ou une recherche indirecte. L'allocation de ressources et la technologie (indépendance vis-à-vis des alimentations, miniaturisation, mise en réseau) rendent les fonctions plus versatiles. Les systèmes répondent à des questions de plus haut niveau sur la perception (trouver des verres sur une table nécessite plus que de la segmentation) et ils assurent une grande variété de décisions qui pour être locales n'en sont pas moins plus générales (trouver des sous parties, etc.).

### 2.3. Vers des applications plus difficiles

Aucun des systèmes ci-dessus n'offre de réelle autonomie : les compétences du robot semblent trop spécialisées et parcellaires, fonctionnant sans lien les unes aux autres. L'objectif reste plus de diminuer la complexité des algorithmes ou du système que d'accroître la capacité du système à gérer un nombre de modules. La « vision située » a alors pour premier objectif un tel contrôle en aménageant la communication des ressources. Conséquence directe de l'absence d'une optimisation globale, la plupart des opérations menées à bien au niveau précédent restent robot-centrées : la machine tire parti de ses propres mouvements pour connaître mieux son environnement statique.

Pour donner un exemple temporel, trouver par soi-même des amers intermédiaires est une amélioration notoire mais que se passe-t-il s'il n'y a aucun objet intéressant hormis des mobiles, ou si l'intérêt lié à la scène ou à la tâche perceptive provient du mouvement ? Les mouvements des objets font alors partie de la question, et si le robot suit une cible en mouvement, il peut ne jamais l'attraper. La différence entre la vision active et la vision située réside dans ce cas entre chercher la cible où elle se trouve ou bien analyser l'endroit et plus généralement la situation où la cible sera au moment de l'observation.

Ces dix dernières années, des modules fonctionnels prouvent une adaptation croissante, de la généricité et un certain niveau sémantique. Deux exemples parmi d'autres illustrent ce progrès : à Munich, patrie de la première voiture autonome [1,2,3,4,5,6], la génération actuelle prend en compte un contrôle du regard multi focal, à base de vision saccadée et d'estimation de la position à venir. À Oxford, l'algorithme de ConDensAtion supporte un suivi amélioré basé sur une représentation plus générale des probabilités, apparenté à un filtrage de Kalman avec densités conditionnelles multi modales [76,77]. Voir aussi [33,70,78,79] pour d'autres résultats basés sur des tentatives d'extension des probabilités (logique floue, évidence, etc.). Les composants deviennent plus adaptatifs avec la technologie, mais pas les systèmes.

Ainsi, les modules exécutent leurs tâches de manière plus satisfaisante mais en même temps une optimisation plus globale est nécessaire pour aller au-delà du niveau actuel d'intelligence. La pression est aujourd'hui sur les communications. En résumant des exemples de systèmes de fusion multi capteurs progressivement plus ambitieux, nous tentons de comprendre comment les robots pourront aborder des situations plus variées.

### 3. Conception de systèmes dits intelligents

L'intelligence des systèmes offre plusieurs facettes. Les plus communes réfèrent aux tâches que ces systèmes peuvent remplir, aux situations changeantes que ces systèmes savent maîtriser ou à leur versatilité *i.e.* leur capacité de se reconfigurer depuis la simple adaptation jusqu'à l'apprentissage véritable. Toutes ont trait aux objectifs du système et à ses fonctionnalités décrivant la capacité du système à agir sur le monde extérieur. Des caractéristiques moins évidentes d'un système intelligent réfèrent plus à des aspects technologiques et néanmoins fondamentaux, à savoir les chemins de données et la façon dont l'information peut transiter et circuler à l'intérieur du système, en étant stockée et transformée tout au long du chemin. Il s'agit donc de la capacité interne du système à être décrit pour être construit.

Nous classons ces différentes facettes de la façon suivante :

1. Le domaine envisagé est suffisamment mature (voir § 2) pour disposer d'enchaînement de tâches pouvant répondre de façon satisfaisante à toute situation spécifiée dans l'application. Un tel enchaînement est dénommé arc de traitement. L'ensemble des arcs constitue un graphe câblé ou programmé.
2. Pour assurer le fonctionnement satisfaisant, il est possible d'adapter les paramètres des traitements composant les arcs de ce graphe câblé.
3. Le contrôle des situations changeantes impose la reconnaissance des situations en cours, et d'enclencher des actions correspondant à différents arcs de traitement. Il est alors nécessaire de pouvoir commuter ou bifurquer dans le graphe des arcs. C'est particulièrement cette fonction de supervision que nous voulons mettre en relief dans cet article pour atteindre le niveau de complexité « situationnelle » inhérente aux systèmes de vision autonome.

4. La facette « Connaître *a priori* / Apprendre par cœur » consiste à proposer des mesures de la réussite et échec d'un arc de traitement, en vue d'optimiser la représentation d'une nouvelle situation en identifiant les prémisses et en programmant les tâches et actions correspondantes. En conduite autonome, l'exemple serait d'apprendre par démonstration un trajet point à point ou une trajectoire optimale dans un virage, puis de la rejouer à bon escient.

5. La facette « Comprendre les situations » consiste à condenser, abstraire, coder, réunir des situations ayant des prémisses identiques, grouper des prémisses ayant des actions identiques... pour mieux abstraire les classes.

La facette 1 reste du ressort des acteurs au travers de l'hypothèse « chacun fait de son mieux et ce mieux est satisfaisant ». Les facettes 2 et 3 sont le cœur du problème traité dans cet article, l'autonomie. Les facettes 4 et 5 ne sont pas abordées ici.

#### 3.1. Positionnement du problème

Les exemples les plus sophistiqués décrits jusqu'ici tels la dérivation d'information ou la recherche indirecte et les avancées récentes précisées ci-dessus montrent qu'il est crucial d'anticiper. On constate au stade actuel deux types de prédiction : interne et externe. La bonne nouvelle quand on parle de contrôle ou de mécanismes d'optimisation est, qu'une fois implantés, ils servent les deux types. L'optimisation interne consiste en une gestion des ressources, et l'optimisation externe concerne la tâche classique d'adaptation et d'accomplissement. Reste que de nombreux contre-exemples simples, dès le stade technologique, comme la gestion mémoire (ex. la notion de cache) prouvent que même à ce niveau rendre les systèmes plus réactifs ne les rend pas plus intelligents. Pour qu'une forme d'intelligence émerge, nous posons comme hypothèse que l'anticipation doit être au même instant l'objectif et un moyen de l'atteindre. Techniquement (Figure 3), cela veut dire que ce contrôle détermine un alignement explicite d'une situation

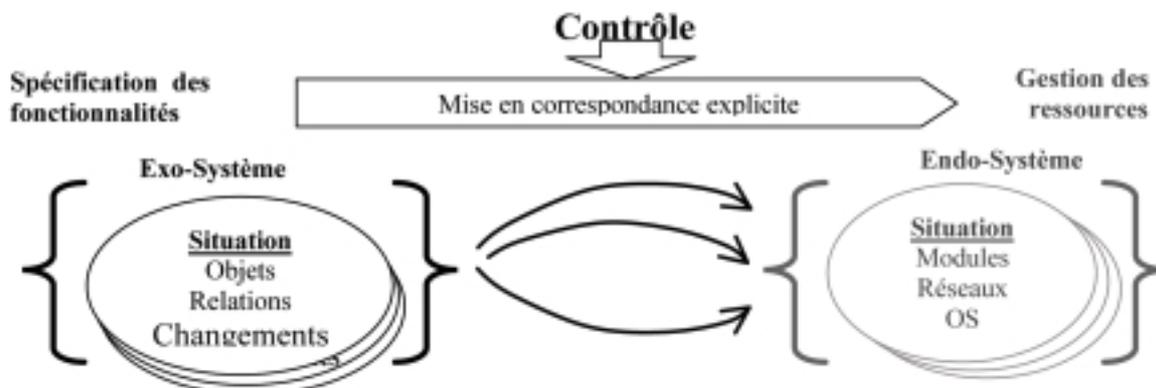


Figure 3. Mise en correspondance des situations externes sur le graphe d'activité interne par alignement pour toutes les fonctionnalités désirées. La vision située est constituée des trois blocs : la spécification de l'exo-système ; la spécification de l'endo-système ; et la mise en correspondance.

externe (décrite en termes d'objets, d'interrelations, de changements) sur une situation interne (décrite en termes de modules d'acquisition, traitement ou décision, de réseau et de système d'exploitation). L'ensemble des alignements construit une mise en correspondance de l'exo-système {ensemble des situations externes et fonctionnalités correspondant aux spécifications de l'application} sur l'endo-système {ensemble des situations internes et configurations possibles de la machine nécessaires au caractère opérationnel de l'application}. L'existence de cette mise en correspondance a trois conséquences majeures :

- Le terme *situation* induit le nom de vision située qui réfère à des graphes variables, éventuellement dynamiques, des deux côtés de l'application.
- Situé implique un contrôle sensiblement plus complexe que l'adaptatif, ne serait ce que par sa double nature endo et exo.
- L'endo-système et l'exo-système interagissent, ce qui implique que les flux d'information entre objets soient transformés en flots de données entre modules de traitement pour être également contrôlables.

Cette conséquence ultime est la traduction exacte de la conjecture ci-contre : *la mise en correspondance réalise la commande du système tandis que sa mise à jour est l'objectif même de son contrôle.*

Nous pouvons maintenant clarifier une différence majeure de contrôle entre vision active et vision située. La première maximise le volume d'information utile à la tâche en gérant les ressources du système et donc, en l'absence d'intervention physique, indirectement celles de l'environnement. La seconde traite des flux d'information internes, externes et transverses, pour maximiser une efficacité prédéfinie étant donné un résultat planifié.

### 3.2. Différentes approches

On peut distinguer quatre principales représentations des dispositifs associant supervision et planification, chacune correspondant à une approche scientifique : la systémique, l'automatique théorique, l'intelligence artificielle et les sciences cognitives. Les résultats varient en fonction de l'accent mis sur telle ou telle représentation et sur la proportion d'intervention humaine et de modèles abstraits. Mais en traçant brièvement les grandes lignes de ces quatre approches, il apparaît qu'aucune ne peut supporter complètement la vision située telle que nous sommes en train de la définir. C'est pourquoi nous proposons un type d'architecture intermédiaire dans laquelle on contrôle les flux entre des opérateurs supposés quasi optimaux et fonctionnant de manière satisfaisante.

La systémique fait l'hypothèse qu'un problème se découpe en sous problèmes à partir de réarrangements connus du type parallèle ou séquentiel, jusqu'à ce que la complexité soit suffisamment faible pour que la sous fonction résultat soit conçue

par une équipe de taille humaine. La difficulté provient alors de l'atomisation qui génère progressivement plus d'interactions imbriquées entre modules dans l'espoir de satisfaire les spécifications initiales. L'essentiel du travail consiste ainsi en la définition de standards fonctionnels et de protocoles pour atteindre un fonctionnement cohérent, en fait assez éloigné d'un comportement intelligent.

L'automatique théorique procède à partir de modèles du système et des perturbations envisagées. Le modèle associe des modules tels que l'état de chaque module contribue à l'état du système global. Un contrôleur réel s'attachera par exemple à coordonner les transitions d'états locaux. Mais alors la dimension du système (linéaire, au sens de la complexité ou à tout autre sens) empêche vite toute modélisation précise : les sous-systèmes sont dotés de modèles enrichis par des versions dégradées (ex. stochastiques) de couplage entre états. Les solutions approchées ne sont plus que garanties avec une marge d'incertitude et une robustesse suffisante pour qu'une action visée ou une séquence d'actions ait des chances de réussir.

L'intelligence artificielle est prise ici au sens de « ensemble des techniques d'inférence » pour la prise de décision. L'inférence peut-être logique autant que stochastique (ex. Bayésienne) ou à base d'heuristiques pour la supervision. Chaque technique a des avantages et inconvénients excitant l'esprit inventif du concepteur au service de l'ingénierie de systèmes complexes, mais qui restent dédiés à un champ d'application spécifique (entre autres contraintes, la soi-disant expertise du domaine).

Les sciences cognitives intéressent la discussion théorique des représentations mentales, leurs nature, propriétés et fonctions. Via le concept de représentations analogues, on accède au raisonnement, à l'apprentissage et à la résolution de problèmes. Les organisations conceptuelles fournissent autant d'architectures dites cognitives et la structuration de la connaissance permet quelques formes génériques d'implantation sur des calculateurs. Cette approche mimétique favorise de nouvelles théories en intelligence artificielle, mais ne conduit pas aisément à la conception d'une application concrète donnée.

### 3.3. Nature opportuniste du contrôle

Pour finir de poser ce problème, listons les différentes facettes d'un comportement à base d'interactions, telles qu'elles ont été déclinées au sein des groupes de recherche ISIS<sup>2</sup> et CHM<sup>3</sup>. La figure 4 décrit les mécanismes d'interaction entre acteurs d'un système (agents, modules de traitements, ...). Ils s'appliquent aux systèmes échangeant avec leur environnement. La plupart des branches ont des instances déjà implantées séparément dans des systèmes artificiels : la fusion d'information, la compétition, la coopération planifiée ou le contrôle adaptatif. Le point dur intéressant reste le contrôle opportuniste. Or concernant la

2. ISIS : Groupe de Recherche du CNRS Information Signal Image vIision.

3. CHM : Groupe de Recherche du CNRS Communication Home Machine

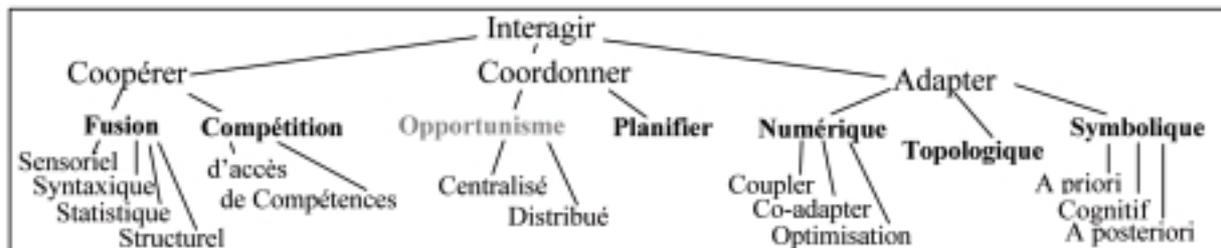


Figure 4. Taxonomie des mécanismes d'interaction.

vision située, la nature opportuniste du contrôle se traduit d'abord par la gestion d'événements qui modifient la situation courante (principalement des objets apparaissent et disparaissent de la scène conduisant à une modification de la situation), puis par la détection au plus tôt de comportements répertoriés de certains objets dans cette scène (mobilité, immobilité, direction du regard du piéton, ...).

## 4. Un modèle pour les systèmes autonomes de vision

Différents types de complexité induisent différentes approches de conception. En les confrontant globalement à la technologie disponible, on comprend qu'il soit difficile d'éviter une décomposition fonctionnelle du système perceptif en modules communiquant et en couches différentes. Couches et modules seront plus ou moins proches de l'action et cette décomposition complexifie le contrôle suivant ses deux axes principaux :

- La gestion des ressources (actionneurs, capteurs, processeurs, mémoires cache et mémoire de masse) pour atteindre un sous-objectif ou pour focaliser une attention : rendre les données nécessaires à une décision accessible au plus tôt.
- Le dimensionnement de la description de l'état interne et de la prédiction pour anticiper ce à quoi l'on doit se préparer dans un futur dont le terme est adéquat : anticiper plusieurs futurs (axe temps) ; regarder plus large, plus loin, plus précisément (axe espace) ; raisonner, imaginer plus profondément, avec des raffinements (axe base de données).

On considère donc avec profit la décomposition fonctionnelle initiale (Figure 3) comme une première architecture modulaire mettant en œuvre des flux de données qu'il s'agit alors de contrôler. Le problème que nous tentons de poser correctement au travers des exemples de la section suivante est : comment conserver le caractère opportuniste du contrôle tout en implantant la majorité des spécifications fonctionnelles présentées dans § 2 qui se traduisent par l'existence d'arcs de traitement qui vont être décrits en § 4.2

### 4.1. La fusion de capteur « délibérative » par opposition à « réactive »

Deux directions principales ont émergé en pratique de la robotique pour ce qui est de la modélisation des systèmes de vision :  
 - L'approche délibérative où un système n'est pas supposé s'évertuer à résoudre tout le problème. Il dispose d'outils pour décomposer fonctionnellement un problème en sous-problèmes en fonction de spécialités prédéfinies. Le contrôle peut alors être réalisé par un module de planification qui équilibre systématiquement la charge de calcul.

- L'approche réactive dans laquelle un système accomplit une sélection d'actions faisables. Les exemples présentés au § 2 participent de cette dernière famille.

La plupart des tentatives récentes sont hybrides et mettent un accent variable sur la planification ou le couplage entrée-sortie (*i.e.* perception/action). D'après § 2, la vision active apparaît à la fois comme une technique clé en vision et comme un processus intelligent puisqu'il consiste à chercher l'information où elle se trouve. En réalité, du point de vue contrôle et système, il n'y a pas de différence entre la fusion de capteurs et la vision active. Les deux visent une perception plus riche du système. La fusion demande une coopération optimisée entre les caméras et les autres capteurs, la vision active exploite le même mécanisme entre caméras et actionneurs. La fusion tire bénéfice de capteurs *a priori* plus variés ou agiles, et la complexité est principalement dans la gestion de ressources, la vision active met les données images en exergue et la complexité est maîtrisée par contrôle des capteurs. Les algorithmes et les techniques restent identiques dans les deux cas ainsi que l'objectif final : la maximisation de l'information nécessaire pour déclencher correctement les actions.

Nous nous concentrons de ce fait sur les systèmes de fusion de capteurs. Nous résumons un petit nombre d'architectures spécifiques à cette fusion de données pour illustrer l'évolution dans la conception des systèmes de perception (voir [89] pour une présentation plus détaillée). Ces architectures peuvent être décrites au niveau matériel, au niveau système ou au niveau fonctionnel. Les complexités tant pratique que théorique ou fonctionnelle croissent en rapport avec le développement des systèmes de fusion. Une première approche pour maîtriser la combinatoire associée à cette complexité croissante est de la

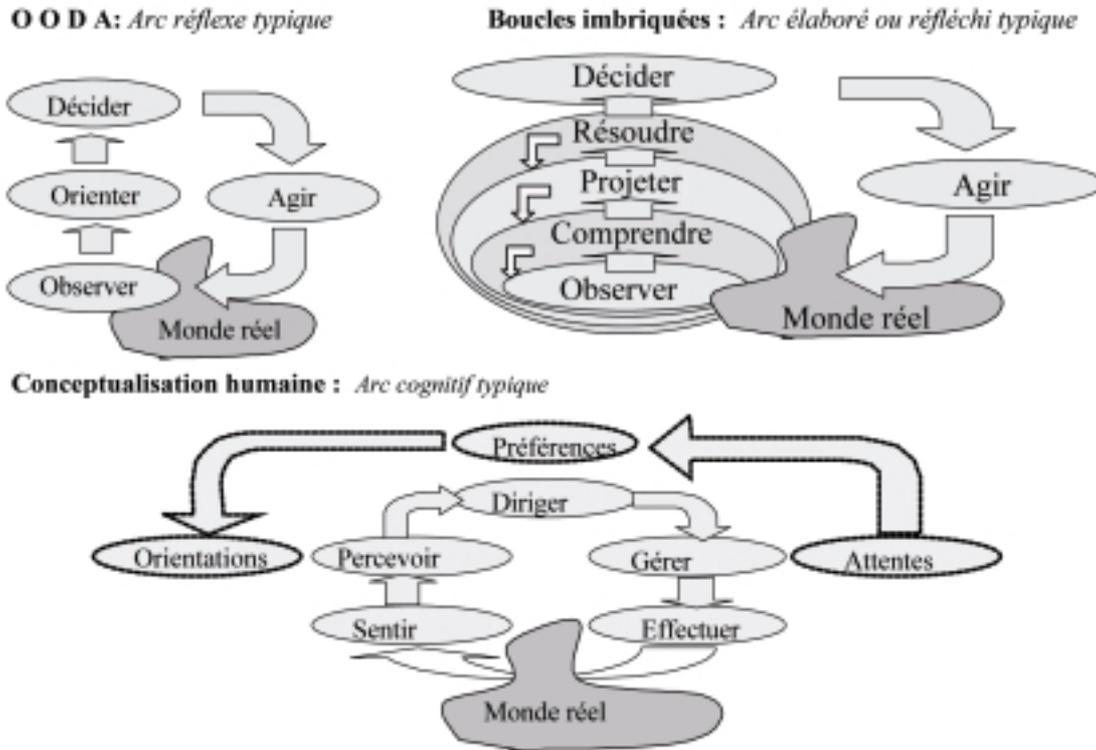


Figure 5. Arcs de traitement implantés dans différents schémas adaptés.



répartir en niveaux fonctionnels différents. Un groupe de travail parrainé par le DoD<sup>4</sup> américain a proposé un tel modèle multi niveaux de fusion. Les 6 niveaux composent un modèle d'architecture fonctionnel et prétendument générique. Les descriptions correspondantes du système sont délibératives et évoluent encore de nos jours (JDL Data Fusion Working Group [80,87]). Un travail de taxonomie de ce type (et donc non opérationnel) pourrait conduire à un typage des tâches, et donc à l'interopérabilité des sous-tâches dans le système complet, mais les flux et les interactions entre modules ne sont pas explicites. Dasarthy [81] suggère d'étiqueter aussi les flux de données et de contrôle suivant 3 niveaux d'abstraction: données, prémisses et décisions. Les données en circulation ont alors un type à l'instar des techniques de programmation déclaratives ou orientées objet. Le résultat est, à nouveau au stade taxonomique, de montrer la richesse des mécanismes de combinaison d'information. Les flots entre « blocs de combinaison » sont déjà clairement identifiés par leur sémantique, et un mécanisme de régulation contrôle les étages de niveaux inférieurs à partir de décisions partielles. Bedworth [82] adapte un autre modèle, nommé OODA qui est en fait une boucle traversant les flux de données. La boucle se referme par l'action sur le monde réel, introduisant donc de manière explicite une réactivité dans la conception de l'architecture et du contrôle. Endsley [83] et Salerno [84] appartiennent clairement à la famille des automaticiens [86,88]

avec un ensemble de quatre boucles réactives intégrées à l'intérieur de différents niveaux fonctionnels de tâches. Un état global contrôlé par un mécanisme hiérarchique a été préféré au contrôle direct des flux de données. Il en résulte un mélange entre les approches de type Bedworth/Boyd et JDL: le gain majeur sur les modèles précédents réside dans le fait d'explicitement le contrôle grâce un typage des états et des communications plutôt que typer simplement les données. Bedworth et Frankel [85] vont un pas plus loin en adaptant certaines hypothèses sur le déroulement de tâches comparables dans le cerveau: la cognition humaine supposerait de modéliser les interactions entre les différentes modalités de raisonnement. Deux flux explicites circulent dans des directions opposées: des données brutes vers la décision, de la décision vers des niveaux de contrôle inférieur. La description fait apparaître les flux de données avec un typage, circulant d'une unité fonctionnelle à une autre, mais sans encore figurer une régulation explicite. Ces unités sont empilées suivant un principe d'abstraction hiérarchique. C'est alors en quelque sorte une somme de combinaisons à la Dasarthy et à la Salerno avec flux de données. L'évocation de l'humain autorise des passerelles entre les différentes écoles impliquées dans la conception d'un système complexe (et autonome), au niveau opérationnel, théorique et calculatoire, élargissant la classe des difficultés abordées.

4. DoD: Department of Defense.

## 4.2. Construire des blocs ou construire des architectures

Les exemples choisis confirment [89] que l'évolution se concentre finalement sur les flux d'information. On note que seules les deux premières architectures présentent un caractère opérationnel, les opérateurs qui les composent sont mieux identifiés en vue de leur contrôle, et les états du système apparaissent également définis. Le modèle JDL se concentre sur les fonctionnalités tactiques limitées au monde des opérations réelles. Le modèle DFD concerne l'implantation de mécanismes pratiques de vision. Les trois dernières propositions régressent en ce que les opérateurs demanderaient à spécifier trop de détails par rapport aux objectifs. Mais en se concentrant sur les flux de données et de contrôle, ils mettent en évidence des arcs de traitement qui supportent les mécanismes de fusion à différents niveaux sémantiques.

Le développement des mécanismes d'inférences n'apparaît pas facilité. En revanche on note une tendance à la réutilisation de blocs de base similaires et associés de manière différente, ce que nous illustrons figure 5. Les associations traduisent différentes motivations. Le modèle OODA s'adapte au traitement réflexe avec sa boucle à plat. Pour supporter des raisonnements plus élaborés à base de solutions prédictives en compétition, le second type implante les mêmes fonctions mais dans des boucles imbriquées. Le dernier modèle générique fait apparaître un flot supplémentaire dans la direction opposée, en réalité une supervision distribuée, qui traverse des fonctions à nouveau similaires. Ces trois modèles s'accommodent d'empilements de telles boucles, chacune étant alors dévolue à un niveau sémantique donné.

C'est pourquoi à un certain point entre le deuxième et le troisième exemple de schéma, la combinatoire des interactions devient complexe et ne se trouve pas décrite dans le schéma lui-même. Les tâches imbriquées correspondent à des traitements avec rétroactions, l'interaction entre les doubles flux de données et de contrôle reste complexe et introduit des latences. Finalement, le fonctionnement du système se décrit à nouveau au travers du contrôle des communications entre modules qui sont supposés.

1. Avoir été conçus en vue de leur tâche individuelle.
2. Être en situation de réaliser la tâche dans des conditions connues et de préférence quasi-optimales.

Ainsi les modules n'ont-ils plus besoin d'être décrits en détail, pas plus qu'ils ne nécessitent des contrôles fréquents.

## 4.3. Un modèle informatique pour les systèmes autonomes

De l'étude précédente nous avons conclu que les meilleures solutions devraient provenir d'un compromis entre des tendances assez contradictoires :

- intégrer des sous-fonctions modulaires dans un système de fusion global et général ;

- enchaîner de façon opportuniste des mécanismes actifs dans des applications spécifiques.

La frontière entre les deux schémas de conception a tendance à devenir floue quand des réactions versatiles ou des réactions complexes mélangeant opportunité, précision et portée sont nécessaires au guidage d'un véhicule autonome. Notre propre modèle (figure 6) se veut suffisamment général pour instancier tous les modèles décrits précédemment selon la classe de situations et l'application.

L'interaction première avec l'exo-système est réalisée par des arcs de traitement en bas à droite. Ils sont construits par ailleurs (algorithmes classiques, dispositifs matériels usuels, fonctions communément admises, ou bio-inspirées, etc.) et disponibles sous forme de blocs. Chaque arc supporte un fonctionnement nominal en mode de perception active ou réactive et a fait ses preuves en tant que mécanisme opérationnel dans une gamme donnée de contextes<sup>5</sup>. Nous les insérons dans le schéma global dans un pur style OODA où la boucle se referme par une chaîne de contrôle centralisée (blocs numérotés 1 à 5 sur la gauche). Dans la plupart des cas, plusieurs étapes peuvent être court-circuitées, telles 2 et 4 en cas de détection positive de piétons, pour réagir plus rapidement. Il est à noter que cela modifie l'état des arcs actifs à ce moment-là. Autant d'arcs que nécessaire peuvent être implantés et agir de façon concurrente dans un mécanisme anticipatif et attentionnel : dû au caractère opérationnel réel, les arcs sont triés selon « l'étonnement » perceptif (essentiellement le délai de réactions de chacun des arcs).

La chaîne de contrôle constitue un module centralisé (à gauche) en une colonne de 5 séquences :

1. Mettre à jour la connaissance de la scène sur arrivée d'événements
2. Sélectionner les comportements possibles qui vont entrer en compétition
3. Redéfinir les objectifs à court et moyen terme
4. Planifier les actions
5. Enclencher les actions

De telles étapes procédurales sont presque communes (voir TEA1 par exemple). Si nouveauté il y a, elle réside ici dans l'étape 3. Le contrôle à ce niveau se concentre sur trois variables support de l'anticipation : le terme (temps), le champ (espace) et l'imagination (base de donnée). Chacune est quantifiée sur trois valeurs : court, moyen et long. Ceci est notre traduction architecturale du concept de vision située (§ 2.3 et § 3.1) généralisant celui de vision active : par exemple les séquences 2 et 4 autorisent la prédiction qui est une différence essentielle entre vision active et située, sans interférer avec les réflexes qui ignorent ces étapes. Par exemple, soit un objectif plus large comme

---

5. Soulignons qu'il s'agit d'une traduction littérale du postulat proposé avec la métaphore de l'orchestre.

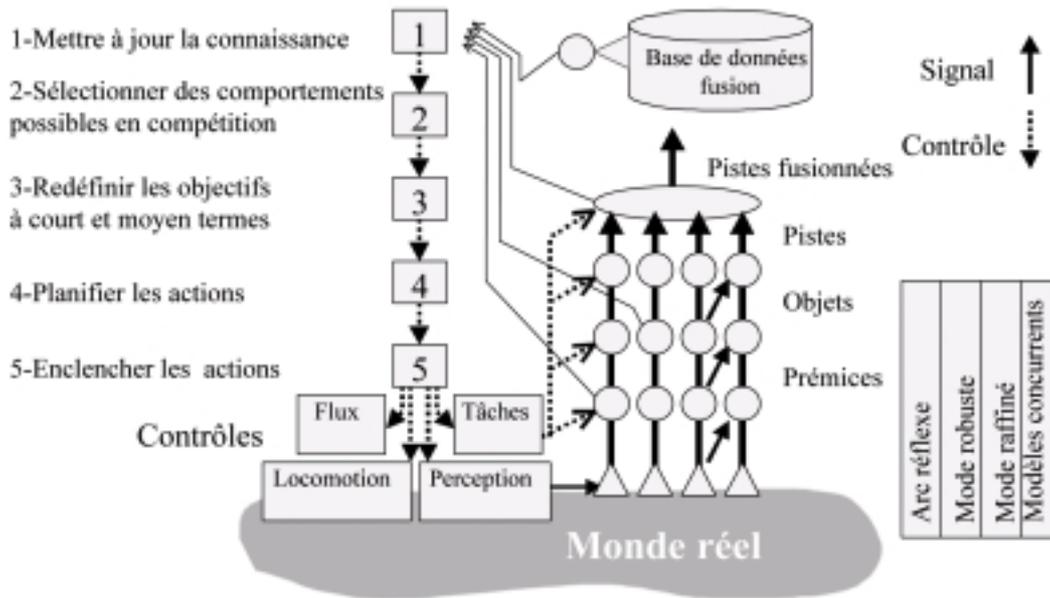


Figure 6. Instancier un double flux en vue d'un contrôle opportuniste.

pour un véhicule celui de «coller à la route», passer de «regarder où une cible est» à «regarder où une cible sera» requiert : [Étape 2] : sélectionner des comportements potentiels à partir des pistes fusionnées et de modèles locaux extraits de la base de données.

[Étape 3] : décider sur anticipation et filtrer ensuite sur les comportements importants par rapport aux objectifs programmés, puis traduire ces comportements en des requêtes intermédiaires.

[Étape 4] : compiler les requêtes dans le langage des capteurs et outils des modules disponibles

[Étape 5] : Enclencher des outils ou services adaptés aux bonnes conditions (vis-à-vis du temps, de la précision, de l'énergie).

En fin de compte, pour acter ces étapes, le module de contrôle central gère 4 entités : la perception (active vision), la locomotion (véhicule, robotique), le contrôle et la supervision (des tâches à déclencher sur les flots de données des arcs de traitement) et le dernier, mais non le moindre, le contrôle des flux. Deux sont des ressources à gérer et deux sont des moyens à contrôler. La structure complète devient active sur des événements issus des arcs de traitements ou à l'instigation de tâches opérant en continu sur la base de données de fusion.

D'autres instances du modèle proposé sont envisageables mais celle présentée ici, en implantant un contrôle centralisé déclenché sur événement, constitue une plate-forme adaptée pour l'expérimentation à la fois sur des acceptions simplificatrices de l'intelligence humaine comme promis par les sciences cognitives et sur la préoccupation architecturale de favoriser des comportements intelligents externes comportant une complexité «situationnelle» élevée. En ce sens, un modèle d'architecture versatile a été concrétisé : il réalise un compromis entre :

- des ensembles structurés de fonctionnalités haut niveau dérivées des applications, bien trop générales pour atteindre un quelconque caractère opérationnel,
- des mécanismes opératifs dérivés de la notion de flux, qui n'offrent pas d'emblée une flexibilité d'organisation suffisante pour être représentatifs d'une réelle forme d'intelligence.

Ce noyau architectural est prévu pour accepter continûment des changements d'objectifs dont l'accumulation devrait déboucher en pratique sur une vision située.

## 5. Conclusion

Au travers de la métaphore de l'orchestre, nous présentons la notion de système autonome formé d'acteurs et vérifiant le principe que *chaque acteur fait de son mieux et que ce mieux est satisfaisant*. Il est donc vraisemblable que décrire un système qui offre de telles capacités de coordination au service d'une performance donnée ne puisse se faire en termes du fonctionnement correct ou non de tel ou tel composant. Sur le plan de la complexité également, cette nécessaire hypothèse que chaque partie remplit sa tâche est intéressante : elle implique que le supplément d'activité rendant le comportement perçu intelligent ne serait alors qu'une affaire de communications internes et de leur contrôle.

Cet article se base ainsi sur un parallèle entre deux aspects de l'évolution actuelle des systèmes *pour viser des tâches plus ambitieuses en leur associant des descriptions et un contrôle plus abstraits*. L'étude de l'évolution des systèmes de vision conduit à représenter l'exo-système par l'ensemble des situa-

tions externes et fonctionnalités correspondant à toutes les spécifications issues de l'analyse des besoins en vision située § 2.3. Un tel système adresse la complexité en terme de nombre de situations prises en compte et de capacité à changer d'objectif. Dans le cadre de l'implantation d'un système complexe, il est nécessaire de réaliser une première description systémique et il est souvent judicieux de la conserver comme modèle haut niveau de l'architecture informatique. Nous avons donc considéré alors qu'un contrôle opportuniste des communications internes devait être ce haut niveau. Ce contrôle détermine l'alignement d'une situation externe sur la configuration interne. L'ensemble des alignements construit une mise en correspondance de l'exo-système sur l'endo-système. Cet article établit la conjecture suivante: *la mise en correspondance réalise la commande du système tandis que sa mise à jour est l'objectif même de son contrôle.*

Pour illustrer nos travaux, nous avons explicité une proposition de schéma informatique à contrôle centralisé opportuniste nourrie au plus tôt par des événements et issue d'un grand nombre d'arcs de traitement. Le contrôleur est constitué d'étapes procédurales qui sont presque communes. La nouveauté réside ici dans l'étape consistant à «Redéfinir les objectifs à court et moyen terme». Le contrôle à ce niveau se concentre sur trois variables support de l'anticipation: le terme (temps), le champ (espace) et l'imagination (base de donnée). Chacune est quantifiée sur trois valeurs : court, moyen et long. Ceci est notre traduction architecturale du concept de vision située.

Aucune contrainte temps réel n'a été prise en compte. Nous acceptons que le système ralentisse quand la complexité augmente de trop. Ce système est conçu pour prendre en compte le niveau d'intelligence basé sur la commutation de situation et d'objectif, une perspective est de voir dans quelle mesure les niveaux d'intelligence «Apprendre par démonstration» et «Comprendre les situations» peuvent être implantés sur cette architecture.

## Références

- [1] B. MYSLIVETZ, E.D. DICKMANS, Recursive 3D road and relative ego-state recognition, *IEEE Trans. PAMI*, spec. issue on Interpretation of 3D scenes, Feb. 1992.
- [2] E.D. DICKMANS, V. GRAEFE, Dynamic monocular machine vision and applications, *Jour. of Machine vision and application*, Springer int., Nov. 1988, pp.223-261.
- [3] E.D. DICKMANS, 4-D Dynamic vision for intelligent motion control, *Int. Jour. for Engineering Applications of A.I.*, spec. issue on Autonomous Intelligent Vehicles, C. Harris (ed.), 1991.
- [4] E.D. DICKMANS, Expectation-based dynamic scene understanding, *Active Vision*, MIT Press, 1992, pp. 303-335.
- [5] E.D. DICKMANS, B. MYSLIWETZ, and T. CHRISTIANS, An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles, *IEEE Transactions on Systems, Man and Cybernetics*, 20(6), Nov. 1990, pp. 1273-1284.
- [6] M. HEBERT, 3-D Landmark Recognition from Range Images, *IEEE Conference on Computer Vision and Pattern Recognition*, Champaign, June 1992, pp. 360-365.
- [7] I.S. KWEON, T. KANADE. High Resolution Terrain Map from Multiple Data, *IEEE International Workshop on Intelligent Robots and Systems*, Tsuchiura, Jul. 1990.
- [8] I. S. KWEON, Modeling Rugged Terrain by Mobile Robots with Multiple Sensors, *Ph.D. thesis, Robotics Institute, Carnegie Mellon University*, Feb. 1991.
- [9] M. DAILY, J. HARRIS, D. KEIRSEY, K. OLIN, D. PAYTON, K. REISER, J. ROSENBLATT, D. TSENG, V. WONG, Autonomous Cross-Country Navigation with the ALV, *IEEE Int. Conf. on Robotics and Automation*, Philadelphia, pp.718-726, Apr.1988.
- [10] R.A. BROOKS, R. GREINER, T.O. BINFORD, The ACRONYM model-based vision system, *6<sup>th</sup> International Joint Conference on Artificial Intelligence*, Tokyo, Aug. 1979, pp. 105-113.
- [11] R.A. BROOKS, Symbolic Reasoning among 3D Models and 2D Images, *Artificial Intelligence*, vol. 17, pp. 285-348, 1981.
- [12] T.O. BINFORD, Visual perception by a computer, *IEEE conf. on Systems and Control*, Miami, Dec 1971.
- [13] R.A. BROOKS, T.O. BINFORD, Representing and reasoning about specified scenes, *Proc. DARPA IU workshop*, Apr. 1980, pp. 95-103.
- [14] G. GIRALT, R. CHATILA, R. ALAMI, Remote Intervention, Robot Autonomy, and Teleprogramming: Generic Concepts and Real-World Application Cases, *IEEE Int. Workshop on Intelligent Robots and Systems*, Yokohama (Japan), Juillet 1993, pp. 314-320.
- [15] G. GIRALT, L. BOISSIER, The French Planetary Rover Vap: Concept and Current Developments, *IEEE International Workshop on Intelligent Robots and Systems*, Raleigh, Jul. 1992, pp. 1391-1398.
- [16] R. CHATILA, L'expérience du robot mobile HILAIRE, *Tech. Rep. 3188*, Laboratoire d'Automatique et d'Analyse des Systèmes, Toulouse, 1984.
- [17] P. FILLATREAU, M. DEVY, Localization of an Autonomous Mobile Robot from 3D Depth Images using heterogeneous Feature, *IEEE International Workshop on Intelligent Robots and Systems*, Yokohama (Japan), Jul. 1993.
- [18] P. GRANDJEAN, A. ROBERT DE SAINT-VINCENT, 3-D modeling of indoor scenes by fusion of noisy range and stereo data, *IEEE International Conference on Robotics and Automation*, Scottsdale, May 1989, pp. 681-687.
- [19] M. DEVY, J. COLLY, P. GRANDJEAN, T. BARON, Environment Modelling from a Laser/Camera Multisensor System, *IARP 2<sup>nd</sup> Workshop on Multi-Sensor Fusion and Environment Modelling*, Oxford (U.K.), Sept. 1991.
- [20] F. NASHASHIBI, Ph. FILLATREAU, B. DACRE-WRIGHT, T. SIMEON. 3D Autonomous Navigation in a Natural Environment, *IEEE International Conference on Robotics and Automation*, San Diego, May 1994.
- [21] P. MOUTARLIER, P. GRANDJEAN, R. CHATILA, Multisensory Data Fusion for Mobile Robot Location and 3D Modeling, *IARP 1<sup>st</sup> Workshop on Multi-Sensor Fusion and Environment Modeling*, Toulouse (France), Oct. 1989.
- [22] S. LACROIX, P. FILLATREAU, F. NASHASHIBI, R. CHATILA, M. DEVY, Perception for Autonomous Navigation in a Natural Environment, *Workshop on Computer Vision for Space Applications*, Antibes, Sept. 1993.
- [23] B. ZAVIDOVIQUE, First Steps of Robotic Perception: The Turning Point of the 1990s, *Proceedings of the IEEE.*, vol 90, n°7, 2002.
- [24] J. ALOIMONOS, I. WEISS, A. BANDOPADHAY, Active vision, *International Journal of Computer Vision*, 1(4) pp. 333-356, 1987.
- [25] R. BAJCSY. Active perception vs. passive perception, *Proceedings of the IEEE*, 1985, pp. 55-59.
- [26] V. ANANTHARAM, P. VARAIYA, Optimal strategy for a conflict resolution problem, *System and Control Letters*, 1986.
- [27] B. ZAVIDOVIQUE, A. LANUSSE, P. GARDA, Robot perception systems: some design issues, *NATO Adv. Res. Work. MARATEA*, A.K. Jain Ed. Springer Verlag, Aug. 87.
- [28] X. MERLO, Techniques probabilistes d'intégration et de contrôle de la perception en vue de son exploitation par le système de décision d'un robot, *PhD. thesis*, Institut National Polytechnique de Lorraine, France, 1988.

- [29] Y. ALOIMONOS, Visual shape computation, *Proc. of the IEEE*, 76(8), jan. 1988, pp. 899-916.
- [30] C. FERMÜLLER, Y. ALOIMONOS, Vision and action, *Image and vision computing*, 13(10), 1993, pp.725-744.
- [31] P. BOUKIR, Reconstruction 3D d'un environnement statique par vision active, *PhD. thesis*, Oct. 1993.
- [32] F. CHAUMETTE, S. BOUKIR, P. BOUTHEMY, D. JUVIN, Structure from controlled motion, *IEEE Transactions on PAMI*, 4(11), Feb. 1996, pp.372-389.
- [33] X. MERLO, A. LANUSSE, B. ZAVIDOVIQUE, Optimal control of a robot perception system, *IATED Int'l Symp. GENEVE*, 1987.
- [34] B. ZAVIDOVIQUE, Perception for decision or Decision for perception? *Human and Machine Perception*, V. Cantoni Ed. Plenum Press, 1997, pp. 155-178.
- [35] R. BAJCZY, Active perception, *Proc. of the IEEE*, vol 8(76), aug. 1988, pp. 996-1005.
- [36] J. MAVER and R. BAJCSY, Occlusion as a guide for planning the next view, *IEEE Trans. on Pattern Analysis and Machine intelligence*, 15(5), May 1993, pp 417-433.
- [37] R. PITO, A sensor based solution to the next best view problem, *IAPR Int. Conf on Pattern Recognition*, Aug. 1996, pp. 941-945.
- [38] J. KOSEKA, H. CHRISTENSEN, and R. BAJCSY, Discrete event modeling of visually guided behaviors, *International Journal of Computer Vision*, 14(2), Mai 1995, pp.179-191.
- [39] P. WHAITE and F. FERRIE, Autonomous exploration: Driven by uncertainty, *IEEE International conference on Computer Vision and Pattern Recognition, CVPR '94*, Seattle, pp. 339-346, June 1994.
- [40] D.H. BALLARD, Animate vision, *Artificial Intelligence*, 48(1), Août 1991, pp.57-86.
- [41] D. NOTON, L. STARK, Eye movement and visual perception, *Scientific American*, 224(6), Juin 1971, pp. 34-43.
- [42] A.L. YARBUS, Eye movements and vision, *Plenum Press*, 1967.
- [43] E. MILIOS, M. JENKIN, and J. TSOTSOS, Design and performance of trish, a binocular robot head with torsional eye movements, *International Journal of Pattern Recognition and Artificial Intelligence*, 7(1), Février 1993, pp. 51-68.
- [44] D. MURRAY, K. BRADSHAW, P. MCLAUCHLAN, P. SHARKEY, Driving saccade to pursuit using image motion, *International Journal of Computer Vision*, 16(3), Mars 1995, pp. 205-228.
- [45] J. TSOSOS, A complexity level analysis of vision, *IEEE Int. Conf on Computer Vision, ICCV'87*, London, June 1987.
- [46] K. BRUNNSTROM, J.O. EKLUNDH, and T. UHLIN, Active fixation for scene exploration, *International Journal of Computer Vision*, 17(2), Feb. 1996, pp.137-162.
- [47] Y. ALOIMONOS, Purposive and qualitative active vision, *IAPR Conf. on Pattern Recognition ICPR'90*, Atlantic City, pp.346-360.
- [48] M.I. TARR, M.J. BLACK, A computational and evolutionary perspective on the role of representation in vision, *Computer Vision, Graphics, and Image Processing: Image Understanding*, jul. 1994, pp. 65-73.
- [49] C. BROWN, Towards general vision, *Computer Vision, Graphics, and Image Processing: Image Understanding*, 60(1), 1994, pp.89-91.
- [50] S. SANDINI, E. GROSSO, Why purposive vision? *Computer Vision Graphics, and Image Processing: Image Understanding*, 60(1),1994, pp. 109-112.
- [51] A. L. ABBOTT and N. AHUJA, Active surface reconstruction by integrating focus, vergence, stereo and camera calibration, *Third International Conference on Computer Vision*, 1990, pp. 489-492.
- [52] B. B. BEDERSON, R. S. WALLACE, E. L. SCHWARTZ, Two miniature pan-tilt devices, *International Conference on Robotics and Automation*, 1992, pp. 658-663.
- [53] C. BROWN, Gaze controls cooperating through prediction, *Image and vision computing*, 8(1), 1990, pp. 10-17.
- [54] C.M. BROWN, Kinematic and 3D motion prediction for gaze control, *Workshop on Interpretation of 3D Scenes*, 1989, pp.145-151.
- [55] W.S. CHING, P.S. TOH, K.L. CHAN, M.H. ER, Robust vergence with concurrent detection of occlusion and specular highlights, *Fourth Int. Conference on Computer Vision*, 1993, pp. 384-394.
- [56] J. J. CLARK, N. J. FERRIER, Modal control of an attentive vision system, *2<sup>nd</sup> Int. Conf. on Computer Vision*, 1988, pp. 514-523.
- [57] D. COOMBS, C. BROWN, Real-time smooth pursuit tracking for a moving binocular robot, *Computer Vision and Pattern Recognition*, 1992, pp. 23-28.
- [58] J. L. CROWLEY, P. BOBET, M. MESRABI, Gaze control for a binocular camera head, *2<sup>nd</sup> European Conf. on Computer Vision*, 1992, pp. 588-596.
- [59] J. C. FIALA, R. LUMIA, K. J. ROBERTS, A. J. WAVERING, Triclops: a tool for studying active vision, *International Journal of Computer Vision*, 12(2, 3), 1994, pp. 231-250.
- [60] E. GROSSO, D. H. BALLARD, Head-centred orientation strategies in animate vision, *4<sup>th</sup> Int. Conf. on Computer Vision*, pp. 395-402.
- [61] I. HORSWILL and M. YAMAMOTO, A \$ 1000 active stereo vision system, *CVPR*, 94.
- [62] E. KROTKOV, F. FUMA, J. SUMMERS, An agile stereo camera system for flexible image acquisition, *IEEE Journal on Robotics and Automation*, 4(1), 1988, pp.108-113.
- [63] B. MARSH, C. BROWN, T. LEBLANC, M. SCOTT, T. BECKER, P. DAS, J. KARLSSON, C. QUIROZ, Operating system support for animate vision, *Jour. of Parallel and Distributed Computing*, 15, 1992, pp.103-117.
- [64] D.W. MURRAY, P.F. MCLAUCHLAN, I.D. REID, and P.M. SHARKEY, Reactions to peripheral image motion using a head/eye platform, *4<sup>th</sup> International Conference on Computer Vision*, 1993, pp. 403-411.
- [65] K. PAHLAVAN, J. O. EKLUNDH, A head-eye system-analysis and design, *Computer Vision, Graphics, and Image Processing: Image Understanding*, 56(1), Jul. 1992, pp.41-56.
- [66] K. PAHLAVAN, T. UHLIN, and J.O. EKLUNDH, Integrating primary ocular processes, *2<sup>nd</sup> European Conference on Computer Vision*, May 1992, pp 526-541.
- [67] K. PAHLAVAN, T. UHLIN, and J.O. EKLUNDH, Dynamic fixation, *4<sup>th</sup> Int. Conf. on Computer Vision*, May 1993, pp. 412-419.
- [68] K.N. KUTULAKOS, C.R. DYER, Recovering shape by purposive view-point adjustment, *Computer Vision and Pattern Recognition*, June 1992, pp. 16-22.
- [69] K.N. KUTULAKOS, C.R. DYER, Recovering shape by purposive viewpoint adjustment, *International Journal of Computer Vision*, 12(2/3), 1992, pp.113-136.
- [70] R. PISSARD-GIBOLLET, P. RIVES, Asservissement visuel appliqué à un robot mobile : État de l'art et modélisation cinématique, *Research report RRI1577*, INRIA-Sophia Antipolis, France, Dec. 1991.
- [71] L. E. WIXSON, Exploiting world structure to efficiently search for objects, *Technical Report 434*, University of Rochester, C. S. Department, Rochester, New York, Jul. 1992.
- [72] L. E. WIXSON, Gaze selection for visual search, *PhD thesis, University of Rochester*, Rochester, New York, 1994.
- [73] L. BIRNBAUM, M. BRAND, and P. COOPER, Looking for trouble: using causal semantics to direct focus of attention, *4<sup>th</sup> International Conference on Computer Vision*, May 1993, pp. 49-56.
- [74] R. D. RIMEY P. A. VON KAENEL, C. M. BROWN, Goal-oriented dynamic vision, *Technical report, University of Rochester*, New York, Aug. 1993.
- [75] R. D. RIMEY, Control of selective perception using Bayes nets and decision theory, *PhD thesis, University of Rochester*, New York, Dec 1993.
- [76] M. ISARD, A. BLAKE, Contour tracking by stochastic propagation of conditional density, *European Conf. on Computer Vision*, Cambridge 1996, pp. 343-356.
- [77] M. ISARD, A. BLAKE, Condensation: unifying low-level and high-level tracking in a stochastic framework, *5<sup>th</sup> European conference on Computer Vision*, 1998.
- [78] O. DESSOUDE, Contrôle Perceptif en milieu hostile: allocation de ressources automatique pour un système multicapteur, *PhD Thesis*, University Paris-Sud, 1993.
- [79] C. OLIVIER, Stratégies d'acquisition, de traitement et de prise en compte d'informations pour contrôle de robot en environnement non structuré, *PhD. thesis*, University Paris-Sud, 1993.

- [80] A. STEINBERG, C. BOWMAN, F. WHITE, Revision to the JDL Data fusion Model, *Proc. of AeroSense Conference, SPIE* vol. 3719, pp. 430,441, 1999.
- [81] B. DASARATHY, «Optimal Features-In Feature-Out (FEIFEO) Fusion for Decisions in Multisensor Environments», *Proc. SPIE 3376, Sensor Fusion: Architectures, Algorithms and Applications II*, 1998.
- [82] M. BEDWORTH, J. O'BRIEN, The Omnibus Model: A New Model for Data Fusion, *Proc. of FUSION 99*, USA, 1999.
- [83] M. ENDSLEY, «Towards a Theory of Situation Awareness in Dynamic Systems», *Human Factors Journal*, vol. 37, pp. 32-64, 1995.
- [84] J. SALERNO, «Information Fusion: a High-Level Architecture Overview», *Proc. of FUSION'00*, Paris, 2000.
- [85] C.B. FRANKEL, M. BEDWORTH, «Control, Estimation and Abstraction in Fusion Architectures: Lessons from Human Information Processing», *Proc. of FUSION'00*, Paris, 2000.
- [86] J. GAINNEY, E. BLASCH, Development of Emergent Processing Loops as a System of Systems Concept, *Proc. of AeroSense Conference, SPIE* vol 3179, pp. 186-195, 1999.
- [87] J. LLINAS, C. BOWMAN, G. ROGOVA, A. STEINBERG, E. WALTZ, F. WHITE, «Revisions and Extensions to the JDL Data Fusion Model II», *Proc. of FUSION'04*, Stockholm, 2004.
- [88] P.K. VARSHNEY, Distributed detection and data fusion, *Springer ed*, NewYork, 1997.
- [89] R. REYNAUD, S. BOUAZIZ, «Architecture de systèmes multi-capteurs», *Revue Traitement du Signal, Méthodologie de la gestion intelligente des senseurs*, Vol 22 numéro 4 , pp 393-405, 2005.



Bertrand Zavidovique

Bertrand Y. Zavidovique est Professeur à Paris XI, responsable de l'opération Vision, Architecture, Contrôle de Circuits et Systèmes à l'Institut d'Electronique Fondamentale. Fondateur et directeur de l'Ecole Doctorale STIS de 1992 à 2000, il a été Conseiller Scientifique au Ministère de la Défense (DRET-ETCA) et directeur du laboratoire Système de Perception de 1981 à 1996. Sa Recherche intéresse la perception des robots, leur organisation interne vs. leur efficacité, les implantations temps-réel dont les architectures et méthodes de programmation de circuits intégrés. Auteur de 300 articles de revue ou congrès en Traitement d'Image, Architecture de Systèmes Numériques, Fusion Multi Capteurs et Contrôle Intelligent, B.Z. a reçu 3 prix scientifiques.



Roger Reynaud

Roger Reynaud est Professeur à Paris XI, IUT d'Orsay. Il a été responsable du département Architecture, Contrôle, Communication, Images, Systèmes jusqu'en 2007 à l'Institut d'Electronique Fondamentale. Il s'intéresse à l'adéquation entre algorithmes et architectures, aux architectures des systèmes de fusion et aux mécanismes de fusion de plus bas niveau. Dans le cadre de la nouvelle organisation de la recherche sur les systèmes complexes au sein du RTRA DIGITEO, il défend une voie Systèmes Autonomes incluant capteurs, communication, algorithmes rapides, architectures associées, fusion multi capteurs, marquage temporel, degré d'autonomie. Il est auteur de 80 articles de revue ou congrès en Architecture de Systèmes Numériques, Fusion Multi Capteurs et Contrôle Intelligent.