



UNIVERSITE DE LA ROCHELLE

Laboratoire Informatique Image Interaction

HABILITATION A DIRIGER DES RECHERCHES

présentée par :

VINCENT COURBOULAY

Date : 6 Octobre 2014

Discipline : Informatique

Recherche de l'information dans les images :
de la saillance à la prégnance

JURY :

Anne Guérin-Dugué
Matthieu Cord
Antoine Tabbone

Professeur Université Joseph Fourier de Grenoble, *Rapporteur*
Professeur, Université Pierre et Marie Curie, *Rapporteur*
Professeur, Université de Lorraine, *Rapporteur*

Jean Philippe Domenger
Christine Fernandez Maloigne
Patrick Le Callet

Professeur, Université de Bordeaux
Professeur, Université de Poitiers
Professeur, Université de Nantes

Arnaud Revel

Professeur, Université de La Rochelle

Résumé

Mots clés : Recherche de l'information, modélisation attentionnelle, saillance, prégnance, valorisation de contenus numériques.

Problématique de recherche

La problématique de recherche abordée dans ce manuscrit est la suivante :

Comment chercher l'information la plus pertinente lors de la réalisation d'une activité centrée autour de l'image ?

En d'autres termes, comment sont reliées images et information ? Comment extraire l'information contenue dans une image ? Comment acquérir de l'information grâce à l'image ?

Si la problématique était déjà importante à la fin des années 90, elle a acquis un caractère fondamental depuis. En quinze ans, notre monde a vu l'explosion d'Internet, l'apparition des réseaux sociaux, du cloud, de la dématérialisation massive de notre patrimoine, l'hyper connexion de notre monde.

Une des voies possibles, que j'ai choisie, pour mieux appréhender ce problème est l'étude et la modélisation du système attentionnel.

Une des questions à laquelle je me suis donc attachée concerne la modélisation des mécanismes dynamiques sous-jacents au calcul de l'orientation attentionnelle. Les paramètres de cette modélisation sont nombreux, variés et, pour certains, encore à découvrir. En proposer une classification sommaire m'a permis de définir plus complètement ma problématique de recherche. J'ai ainsi déterminé deux classes complémentaires et cohérentes permettant une approche incrémentale et sectorisée de l'attention visuelle : la connaissance de l'observateur ; les mécanismes attentionnels de l'observateur.

Pour chacun de ces points, je présenterai les problématiques et je détaillerai mes contributions qui sont résumées ci-dessous.

Modélisation de la dynamique attentionnelle

L'attention, en tant qu'outil de gestion de l'information, nous permet de construire une perception adaptée à nos capacités et nos besoins. L'efficacité de ce mécanisme passe par une sélection contextualisée des données les plus pertinentes. Cette sélection est d'autant plus importante dans le cas de la vision biologique, que la quantité de données à traiter est massive, de 10^8 à 10^9 bits par seconde et par œil. Sans l'attention, notre système visuel ne pourrait pas traiter en un temps raisonnable un problème si complexe. La complexité du phénomène attentionnel a ouvert la porte à une multiplicité d'interprétations théoriques. Dans cette optique de modélisation attentionnelle, j'ai apporté une contribution pour modéliser la dynamique attentionnelle grâce à un système d'EDP de type proies/prédateurs. Cette équation traduit la dynamique optimale, au sens d'un flot d'information, du processus d'observation, *i.e.* du transfert de l'information visuelle des indices bas niveau (intensité, couleur, orientation, ...) vers l'observateur sans *a priori*. Cette équation est similaire à l'équation de Volterra-Lotka. De nombreuses expériences ont montré sa pertinence et validité dans le cadre de la thèse de M. Perreira Da Silva. J'ai travaillé à son implémentation temps réel ainsi qu'à une série d'applications comme la mesure de complexité, le recadrage dynamique et l'extension à la vidéo ou aux images de profondeur.

Vers la recherche attentionnelle d'information

Dans le cadre de mes contributions autour de la problématique de la recherche attentionnelle d'information, j'ai proposé une recherche plus proactive de l'information en se focalisant sur les zones d'intérêt pour y déployer des stratégies classiques de recherche d'information. En effet, mon objectif était, et est encore, de travailler sur des systèmes qui permettent une analyse plus efficace et plus bio-inspirée de l'observation. Il s'agit de se focaliser sur des zones d'intérêts de l'image pour permettre une analyse contextualisée en fonction d'une représentation de connaissances que le système intégrerait. Dans le cadre de la thèse de D. Awad, j'ai proposé un filtrage perceptuel de points d'intérêts, puis une chaîne perceptuelle CBIR complète ainsi que son évaluation sur la base VOC2007 et sur une base de planches de bandes dessinées. La conclusion principale réside dans la nécessité de faire des choix cohérents entre détecteurs descripteurs CBIR. J'ai également travaillé sur deux applications reposant sur une analyse attentionnelle : une annotation automatique d'images, une reconnaissance de geste par un robot.

Prégnance & Concept

J'ai également réalisé et encadré des travaux concernant ce que j'appelle la recherche prégnante d'information. La recherche prégnante d'information s'appuie

sur la nécessité de ne pas prendre en compte que les objectifs de l'utilisateur (ce que l'on aurait pu appeler recherche contextuelle d'information), mais également sa vie intérieure, ses émotions, ses connaissances. On voit l'extrême difficulté, voire l'impossibilité, à prévoir ses éléments, et tout l'intérêt d'accompagner l'utilisateur pour, petit à petit, co-construire sa démarche de recherche avec lui. Il est à noter qu'il est très difficile d'avoir une approche holistique de ce domaine compte tenu de la complexité de l'approche. J'ai cependant travaillé sur l'utilisation d'une approche plus formelle afin de résoudre un problème de recherche d'informations dans des documents anciens. J'ai ainsi proposé l'intégration d'informations de plus haut niveau sémantique à différents niveaux : enrichissement sémantique de documents, travaux relatifs à l'émotion et connaissance extraite d'une interaction avec l'utilisateur.

De la valorisation à l'avalorisation

Après avoir montré le positionnement et l'orientation de mes travaux dans une approche cohérente, me permettant d'avoir une vision large spectre de la recherche de l'information, il m'a paru essentiel de repenser ma recherche et lui donner un nouvel élan. Mon projet de recherche trouve son origine dans l'apparition progressive et irréversible des données massives ou big data.

Je propose de positionner mon projet dans une vision court, moyen et long terme.

Un des projets les plus importants que je porte concerne la mise en place, en partenariat avec la chaire d'excellence « Gestion du conflit et de l'après-conflit », d'un Centre international de documentation pour la collecte et analyse de documents et d'archives provenant, entre autres, de tribunaux pénaux internationaux et de la Cour pénale internationale.

Ce projet vise à la numérisation des archives, leur indexation, leur conservation et leur mise à disposition des chercheurs et du public. J'ai déjà récupéré de nombreuses archives (Tribunal pour le Cambodge – Album des Balkans – le procès d'Oradour) et commencé à regrouper autour de moi un groupe de chercheurs transdisciplinaires (SHS, juristes, historiens et informaticiens).

Ce manuscrit ainsi présente le positionnement de mes travaux et la manière dont j'ai dirigé ma recherche pour avoir une vision large spectre de la recherche de l'information.

Table des matières

Résumé	i
Table des matières	v
Table des figures	ix
Liste des tableaux	xiii
Avant propos	1
0.1 Mes travaux de thèse	1
0.2 Contexte de recherche	3
0.3 Organisation de ce mémoire	4
Synthèse des Activités de Recherche	7
1 La recherche de l'information : de la saillance vers la prégnance	9
1.1 La recherche de l'information plutôt que celle d'information?	10
1.2 L'attention comme mesure du potentiel informationnel	13
1.2.1 La connaissance de l'observateur	14
1.2.2 Les mécanismes attentionnels de l'observateur	15
1.3 Démarche scientifique : la <i>diagonale informationnelle</i>	16
1.3.1 Exemple de positionnement de travaux	16
1.4 Conclusion	20
Points clés	24
2 Modélisation de la dynamique attentionnelle	27
2.1 L'attention dans tous ses états	28
2.1.1 Attention ouverte ou couverte	28
2.1.2 Les approches exogène et endogène	31

2.1.3	Attention orientée espace ou objet	32
2.1.4	Attention centralisée ou distribuée	32
2.1.5	Attention et système dynamique	32
2.2	Information physique extrême et observation	34
2.2.1	Information physique extrême	34
2.2.2	Dérivation d'une dynamique attentionnelle	36
2.2.2.1	Hypothèses	36
2.2.2.2	Loi d'évolution de la dynamique attentionnelle	37
2.3	Description du modèle et implémentation proies/prédateurs	38
2.3.1	Simulation de l'évolution du focus d'attention	40
2.3.2	Validation du modèle	41
2.3.3	Implémentation	43
2.4	Applications	43
2.4.1	Complexité des images	43
2.4.2	Recadrage dynamique	47
2.5	Extensions du modèle	47
2.5.1	Extension à la vidéo	47
2.5.2	Extention aux <i>nouvelles</i> images	50
2.6	Conclusion	55
	Points clés	58
3	Vers la recherche attentionnelle d'information	61
3.1	Positionnement de mes travaux	61
3.2	Recherche attentionnelle de l'information	62
3.2.1	Filtrage attentionnel	63
3.2.2	« Attentive CBIR »	66
3.3	Applications en recherche d'information	69
3.3.1	Annotation de documents basée saillance	69
3.3.2	Interprétation visuelle de gestes pour l'interaction homme machine	73
3.4	Conclusion	79
	Points clés	80
4	Prégnance & Concept	81
4.1	Positionnement de mes travaux	81
4.2	Vers la prise en compte d'informations de plus haut niveau...	82
4.2.1	issues du web	82
4.2.2	issues d'indices subjectifs	84
4.2.3	issues d'indices objectifs	90
4.3	Une tentative d'approche formelle concernant l'intégration de connaissances <i>expert</i>	92

4.4 Conclusion	96
Points clés	97
Projet de Recherche	99
5 De la valorisation à l'avalorisation - du big data au big dat@rt	101
5.1 Problématiques sociétales	101
5.1.1 De la valorisation	101
5.1.2 ...à l'avalorisation des données	102
5.2 Problématique et démarche scientifique	105
5.2.1 Les différents niveaux d'émergence	106
5.2.2 Les différents niveaux de complexité	106
5.2.3 Démarche scientifique	109
5.3 Intégration des travaux en cours	109
5.4 Exemples de projets	110
5.4.1 Projet <i>BigDatArt</i>	110
5.4.2 Projet de valorisation des archives pénales internationales . .	112
5.4.3 Institut de Valorisation des Contenus Numériques	114
5.5 Mise en perspective	114
5.5.1 Vision à court terme	114
5.5.2 Vision à moyen terme	115
5.5.3 Vision à long terme	118
5.5.4 Transfert de technologie	119
Points clés	121
Conclusion	123
Bibliographie	127
CV détaillé	137

Table des figures

1.1.1	Description du principe général de la recherche d'images par le contenu	11
1.3.1	Arc dramatique d' <i>Aristote</i>	20
1.3.2	Caractérisation des états mentaux en fonction des compétences impliquées et des défis à relever.	21
1.3.3	AutomatonEditor : Editeur graphique de création d'automates (possibilité de visualiser / sauvegarder les compositions)	21
1.3.4	Noorad : Editeur graphique du modèle de narration interactive. Création des éléments Structurels (abstrait) , des éléments concrets (instanciés à l'histoire) et de l'arborescence des scènes créées.	22
1.3.5	StoryTeller : Représentation visuelle de l'exécution de la narration interactive	23
2.1.1	Evolution du nombre de documents traitant de <i>saillance</i> (haut) et d' <i>attention</i> (bas) depuis 1970 (source Scopus).	29
2.1.2	Les mouvements des yeux dépendent de la tâche confiée à l'observateur.	30
2.1.3	« Où est Charlie? ».	31
2.1.4	Modèle attentionnel de Lesser et Murray	33
2.3.1	Architecture du système proies/prédateurs	39
2.3.2	Interface de notre logiciel de modélisation de la dynamique attentionnelle	44
2.3.3	Cartes accessibles depuis notre démonstrateur	45
2.4.1	Exemples de résultat d'analyse de complexité des images basée sur la dynamique attentionnelle	48
2.5.1	Schéma général du système d'attention visuelle	49
2.5.2	Exemple de séquence « Boule + Grille ».	49
2.5.3	Quelques trames de la vidéo « TOP 20 Tennis Master Points ».	51
2.5.4	Quelques trames de la publicité « Levis - Mr Oizo ».	52
2.5.5	Quelques trames de la bande-annonce du film « Hancock ».	53

TABLE DES FIGURES

2.5.6	Proposition d'extention de notre modèle attentionnel aux <i>nouvelles</i> images	54
2.5.7	Notre système composé d'un véhicule Lego Mindstorm et un Kinect (R2D2 voulait absolument être sur la photo)	55
2.5.8	Exemples de résultats	56
3.2.1	Architecture de l'algorithme Zhang	64
3.2.2	Architecture de notre modèle	65
3.2.3	Illustration de l'utilisation du filtrage par la matrice de transformation de Hartley	67
3.2.4	Impact du couplage détecteurs/descripteurs sur le résultat sur la tâche de reconnaissance sur la base VOC2007	68
3.3.1	Schéma global de l'approche	71
3.3.2	Exemple de descripteurs	72
3.3.3	Résultats de la classification de documents administratifs	73
3.3.4	Exemple de logos non-identifiés	74
3.3.5	Exemple d'annotations sur des images naturelles	74
3.3.6	Exemple de captures présentant notre logiciel d'annotation	75
3.3.7	Intégration d'un robot capable d'interaction dans un musée (MICA)	76
3.3.8	Exemples d'un jeu de postures utilisé pour le détection de gestes.	77
3.3.9	Résultats de détection de postures	78
4.2.1	Logo de la fédération MIREs	85
4.2.2	Exemple d'images de la base SENSE.	86
4.2.3	Interface d'annotation émotionnelle.	87
4.2.4	Présentation web des résultats de l'évaluation	88
4.2.5	(a)-(c) Exemples d'images évaluées lors d'une seconde campagne avec le pourcentage d'image présentée, avec en (d)-(f) les images complètes.	89
4.2.6	Évaluation similaire entre les images réduites et complètes	89
4.2.7	Pourcentage d'images désormais classées positive ou négative en fonction du pourcentage d'images visualisées	90
4.2.8	Implémentation de l'approche information émotionnelle	91
4.3.1	Exemples de lettrine et sa décomposition	93
4.3.2	Illustration du schéma de caractérisation des éléments constitutifs des lettrines	95
4.3.3	Exemple de décomposition de lettrine	95
4.3.4	Exemples de lettres extraites automatiquement à partir des lettrines	96
4.3.5	Exemple de segmentation d'une texture dans une lettrine	96
5.1.1	Quelques images illustrant le Big Data	103

5.1.2	Les échelles du Big Data (source HP)	104
5.2.1	Jean-Louis Le Moigne (à droite) avec ... Edgar Morin	108
5.2.2	Matrice support de mon projet de recherche	109
5.3.1	Positionnement des travaux en cours et des perspectives identifiées.	110
5.4.1	Proposition d'interface du projet <i>BigDatArt</i>	111
5.4.2	Google Glass et moi à l'Université de Bratislava (Juin 2014). Attention l'abus de Google Glass nuit sérieusement à votre image.	113
5.5.1	Mise en perspective temporelle et territoriale de mon projet de recherche.	115
5.5.2	Principaux verrous numériques des défis sociétaux relevés par l'ate- lier 7 (A : critique, B : essentiel)	116
5.5.3	Ambition « Valorisation Big Data » Commission <i>Innovation 2030</i> .	119
5.5.4	La pyramide des besoins de Maslow (1954)	124
5.5.5	La pyramide des besoins de Maslow 2.0	124

Liste des tableaux

1.1	Sous-familles proposées pour segmenter les méthodes d'analyse de l'attention en se basant sur la connaissance de l'observateur.	14
1.2	Sous-familles proposées pour segmenter les méthodes d'analyse de l'attention en se basant sur les mécanismes attentionnels de l'observateur.	16
1.3	Mon système de recherche : <i>la diagonale informationnelle</i>	17
1.4	<i>Détection des indices caractéristiques dans le projet FT-AutiSTIC</i>	19
2.1	Benchmark de différents modèles de saillance. Le numéro à l'intérieur des parenthèses est le rang de notre méthode.	42
5.1	Modèles de Boulding	107
5.2	Hierarchie des systèmes de von Bertalanffy (source Wikipedia)	108
5.3	Modèle de Le Moigne	109
5.4	Adéquation verrous / recherche.	117

Avant propos

Avant de commencer le développement scientifique de ce mémoire, il convient de replacer mes travaux dans leur contexte.

0.1 Mes travaux de thèse

Après l'obtention d'un diplôme d'ingénieur obtenu en 1998 à l'école supérieure de procédés électroniques et optiques (ESPEO, devenu Polytech'Orléans), je suis arrivé en 1999 à l'Université de La Rochelle pour une thèse en imagerie médicale en partenariat avec *General Electric Medical System*.

Mes travaux portaient sur l'extraction de l'information bas niveau au sens d'une optimisation d'un compromis incertitude/imprécision.

Ma principale contribution fût d'aborder le problème du traitement d'images dans un cadre physique prenant en compte l'interaction de l'observateur et du système. Ce cadre est celui de l'Information Physique Extrême (EPI).

Son utilisation m'a permis d'unifier un grand nombre d'approches de traitement d'images. J'ai montré que l'équation de Klein-Gordon avec champs, et sa limite relativiste, l'équation de Schrödinger avec champs, est l'EDP optimale de restauration d'images en terme de compromis incertitude-imprécision. Pour ce faire, j'ai intégré un observateur sans *a priori* et sans connaissance préalable au système d'acquisition d'images. Sous ces hypothèses, cet observateur est similaire à un ordinateur faisant l'acquisition d'images sans contextualisation particulière. J'ai ensuite optimisé le transfert de l'information entre l'image et cet observateur.

La solution du principe EPI nous donne l'évolution $\psi(\mathbf{x}, t)$ permettant de gérer le compromis incertitude-imprécision sous la forme de l'EDP suivante :

$$\frac{\partial\psi}{\partial t} = \frac{1}{2}\Delta\psi \tag{0.1.1}$$

On retrouve l'équation de la chaleur. On peut dès lors identifier le paramètre t gérant le compromis incertitude-imprécision à un paramètre d'échelle.

Grâce à cette équation, j'ai retrouvé la plupart des EDP classiquement utilisées en traitement d'images : l'équation de la chaleur, les opérateurs de voisinages

génériques, la diffusion isotrope et anisotrope, les structures multi-échelles et les contours actifs géodésiques.

J'ai soutenu ma thèse en 2002, puis je suis parti en poste d'ATER à l'Université d'Angers. Ce départ m'a permis de collaborer avec le *Laboratoire d'Ingénierie des Systèmes Automatisés* (LISA), afin d'étendre mes travaux à la prise en compte d'*a priori* bas niveau.

J'ai enrichi mon modèle afin de dériver du principe EPI une équation de diffusion plus générale, anisotrope et paramétrable. Pour ce faire, j'ai proposé le changement de variable suivant :

$$\nabla \rightarrow \nabla - \mathbf{A} \quad (0.1.2)$$

$$\frac{\partial}{\partial t} \rightarrow \frac{\partial}{\partial t} + \phi \quad (0.1.3)$$

où \mathbf{A} est un potentiel vectoriel et ϕ un potentiel scalaire.

La solution au principe EPI devenant alors :

$$\frac{\partial \psi}{\partial t} = (\nabla - \mathbf{A}).(\nabla - \mathbf{A})\psi + \phi\psi. \quad (0.1.4)$$

Les potentiels \mathbf{A} et ϕ permettent ainsi d'intégrer des contraintes sur l'évolution de l'image au cours du processus de diffusion. \mathbf{A} agit comme un potentiel s'appliquant le long de l'axe des échelles et contraint l'image à n'être lissée que dans des directions particulières. Le potentiel ϕ peut être, quant à lui, interprété comme un terme d'attache aux données initiales permettant de contraindre l'évolution de la diffusion vers une solution cohérente par rapport à l'image initiale. Des choix particuliers pour \mathbf{A} et ϕ permettant alors, à partir de l'équation précédente, de retrouver des résultats classiques de diffusion.

Mes travaux sont, par exemple, à l'origine de la définition d'une équation de restauration appliquée à restauration des IRM cardiaques marquées¹.

L'objectif de l'étude était de rendre robuste le suivi automatique de la grille de marquage permettant l'analyse de la déformation d'un ventricule cardiaque. La méthode développée pour le suivi de cette grille utilise un modèle de contour actif dont l'énergie image se construit grâce à une diffusion sélective permettant la sauvegarde de l'information utile. Cette approche permet l'obtention de résultats satisfaisants à la fois en terme de précision, de robustesse, de reproductibilité et de rapidité.

Je montrerai que cette approche informationnelle est restée très importante pour moi, car j'ai pu l'étendre à la modélisation de la saillance. En 2004, suite à deux postes d'ATER, j'ai choisi le poste proposé par le *L3I* à l'Université de La Rochelle.

1. <http://www.sciencedirect.com/science/article/pii/S0167865509001883>

0.2 Contexte de recherche

Jusqu'en 2002, le L3I signifiait *Laboratoire d'Informatique et d'Imagerie Industrielle*, ma thèse et la collaboration engagée avec *General Electric* s'inscrivait pleinement dans ses thématiques.

En 2002, le L3I devenait *Laboratoire Informatique, Image, Interaction* et structurait sa recherche selon deux axes.

- *IDDC* (Images, Documents, Données Complexes) : L'objectif de cet axe est depuis de fournir des systèmes et des méthodes permettant l'aide à l'expertise dans des systèmes manipulant des données de niveau de structuration variable, c'est-à-dire utiliser de manière complémentaire des méthodes d'extraction de connaissances à partir de sources hétérogènes, de données complexes et des méthodes de modélisation de la sémantique.
- *S2I* (Systèmes interactifs et image) : Cet axe s'intéresse au développement de systèmes adaptatifs envisagés du point de vue de deux types de boucles d'interactions : l'interaction concepteur/système auteur d'une part (boucle lente) ; l'interaction système/utilisateur, d'autre part (boucle rapide). Les équipes S2I cherchent à appréhender et orienter l'interaction pour permettre une évolution adaptative du déroulement de l'application : en fonction du comportement de l'utilisateur ; en respectant les consignes du concepteur ; en tenant compte de l'état des ressources.

Bien que scientifiquement proche de l'axe IDDC, j'ai souhaité au début de ma carrière intégrer l'axe S2I afin de décloisonner les recherches de ces deux axes. J'ai développé et appliqué ma recherche dans un large spectre d'applications et participé aux montages de projets originaux puisque transdisciplinaires (projet FT-AutiSTIC ou RobAutiSTIC).

Plus récemment, j'ai coordonné, entre autres, une proposition d'ANR entre le L3I, le LISIS, le LIRIS, le GePaSud et ITESOFT, été le leader d'une tâche dans une ANR piloté par le LIP6 (VISIIR : VISual Seek for Interactive Image Retrieval) et co-organisé une journée transverse GDR ISIS et GDR VISION sur la saillance le 19 Juin 2014². Je suis également porteur d'un projet Tournesol 2015 autour de l'art et des grandes masses de données en partenariat avec l'UMons (nom de code : *Big DatArt*).

Je suis également à l'initiative de l'implication du laboratoire comme membre fondateur de la chaire d'excellence « Gestion du Conflit et de l'Après Conflit », chaire qui m'a permis de regrouper autour de moi une équipe de recherche transdisciplinaire autour de la valorisation des archives pénales³. J'ai d'ailleurs été contacté pour présenter cette activité lors de la *Semaine du Numérique et de l'In-*

2. <http://www.gdr-isis.fr/index.php?page=reunion&idreunion=251>

3. <http://fondation.unilim.fr/?Chaire-Gestion-du-conflit-et-de-l-71>

novation en 2014⁴. Je souhaite faire de cette activité un axe fort du futur *institut de valorisation des contenus numériques* en cours de création sur l'Université de La Rochelle. A ce titre, j'ai obtenu un financement de thèse en co-direction pour l'année 2014 autour du sujet : *Analyse haut niveau de données hétérogènes massives : de l'enrichissement vers la conceptualisation. Applications à la valorisation d'archives pénales.*

Ces années m'ont aussi permis de lire des auteurs qui se sont avérés êtres des sources d'inspiration inépuisables, entre autres, Edgar Morin et sa Méthode, Francisco Varela, René Thom, Alain Berthoz, Henri Atlan, Gerald Edelman et bien d'autres. Ces auteurs m'ont permis de voir tout l'intérêt de réfléchir à une approche bio-inspirée de la recherche de l'information dans les images.

(Dans un autre style, je remercie Terry Pratchett, Arnaldur Indriðason, Manu Larcenet et Pierre Rabhi ;o)

0.3 Organisation de ce mémoire

Dans ce manuscrit, je présente pourquoi dès 2004, j'ai choisi d'orienter mes recherches autour de la recherche de l'information dans les images et les séquences d'images sous un aspect attentionnel et bio-inspiré. Les projets auxquels j'ai choisi de participer ou dont je suis à l'origine m'ont permis de définir et conforter rapidement ma ligne directrice (ou devrais-je dire ma *diagonale directrice*?) et ainsi construire ma propre stratégie de recherche.

L'objectif de ce mémoire est de présenter, d'une part, une synthèse des travaux réalisés en soulignant la stratégie de recherche que j'ai mise en place, et d'autre part de présenter mes perspectives de recherche.

Dans le chapitre 1, j'expose ma problématique de recherche depuis ses origines jusqu'au déploiement du système de recherche que j'ai développé. J'y présente deux projets structurants et la démarche scientifique guidant mes travaux et la structure de ce manuscrit. Dans le chapitre 2, j'introduis l'apport théorique majeur de ce manuscrit : une approche informationnelle de modélisation de la dynamique attentionnelle. Je présente ensuite des applications découlant du modèle, son implémentation ainsi que certaines extensions.

Dans le chapitre 3, je décris la problématique de recherche attentionnelle d'information. Je présente les travaux proposant un cadre de travail performant permettant d'appréhender la recherche attentionnelle d'information comme la conception d'un système d'attention visuelle pour l'analyse contextualisée de la scène. Je propose une recherche plus proactive de l'information, se focalisant sur les zones d'intérêt et étant capable d'y déployer des stratégies de recherche d'information.

4. <http://fr.valconum.sni.beta.aksrv.net/>

Je propose également deux applications relatives à cette *conception d'un système d'attention visuelle pour l'analyse contextualisée*. Le premier présente l'analyse de *l'observé*, au travers de l'extension des travaux précédemment introduits. Le second s'intéresse à l'observateur et propose une analyse contextualisée de ses gestes dans le cadre d'une application interactive.

Le chapitre 4 présente les travaux autour de la recherche prégnante d'information. J'y apporte des contributions en présentant tout d'abord mes travaux concernant la prise en compte d'informations de haut niveau issues du web, d'indices subjectifs et d'indices objectifs.

Enfin, dans le chapitre 5, j'expose mon projet de recherche qui oriente mes choix depuis maintenant deux ans. Je le positionne dans un contexte sociétal et scientifique et en propose une double grille de lecture. Je propose d'envisager mon projet selon une vision court, moyen et long terme, ainsi que selon une échelle locale, transrégionale, nationale et internationale.

Entre ce que je pense, ce que je veux dire, ce que je crois dire, ce que je dis, ce que vous avez envie d'entendre, ce que vous entendez, ce que vous comprenez... il y a dix possibilités qu'on ait des difficultés à communiquer. Mais essayons quand même...

Bernard Werber.

Syntèse des Activités de Recherche

« Il faudrait essayer d'être
heureux, ne serait-ce que pour
donner l'exemple. »"

(Jacques Prévert)

Chapitre 1

La recherche de l'information : de la saillance vers la prégnance

« Dans une recherche active, le chercheur ne choisit pas les problèmes à résoudre (l'histoire les lui impose) mais il crée sa problématique, c'est-à-dire que pour résoudre un problème donné il choisit un certain nombre de critères et élabore à partir de ceux-ci son système de recherche »

Dumazedier & Ripert, *Loisir et culture*, 1966.

La problématique de recherche que j'ai abordée durant presque 15 ans (depuis le début de ma thèse en 1999) est restée quasi constante.

Comment chercher l'information la plus pertinente lors de la réalisation d'une activité centrée autour de l'image ?

En d'autres termes, comment sont reliées images et information ? Comment extraire l'information contenue dans une image ? Comment acquérir de l'information grâce à l'image ?

Si la problématique était déjà importante à la fin des années 90, elle a acquis un caractère fondamental¹ depuis. En quinze ans, notre monde a vu l'explosion d'Internet, l'apparition des réseaux sociaux, du cloud, de la dématérialisation massive de notre patrimoine, de l'hyper connexion de notre monde et même des Humanités Digitales². Aujourd'hui, même les frigos envoient des spams³. On compte par exemple :

- plus de 5 millions de photos postées chaque jour sur Instagram (Janvier 2013)
- plus de 16 milliards de photos ont déjà été uploadées sur Instagram (Janvier 2013)

1. *Qu'y a-t-il au dessus de fondamental ?*

2. <http://cdh.epfl.ch/digital>

3. <http://www.tomshardware.fr/articles/internet-objet-frigo-spam,1-46695.html>

- plus de 4500 photos sont envoyées chaque seconde sur les serveurs de Facebook (2013)
- 600 millions de photos, 200 millions de messages vocaux et 100 millions de messages vidéo sont postés sur WhatsApp ... chaque jour (2014)

Ces chiffres donnent le tournis et sont la source principale de l'émergence du *Big Data* comme technologie clé des années futures.

Je partage ainsi le constat de J.P. Rangaswami⁴, un journaliste et blogueur indien :

I believe that we do live in an age of information overload, and that we have to find ways of simplifying our access to the information ; of assessing the quality of the information ; of having better tools to visualise the information, to enrich and improve it, of passing the information on.

Le point commun à tous ces éléments repose sur la recherche de l'information, et pour éviter d'être un *Monsieur Jourdain* de l'information, une précision s'impose.

1.1 La recherche de l'information plutôt que celle d'information ?

L'association des professionnels de l'information et de la documentation distingue dans Le *Vocabulaire de la Documentation*⁵, la recherche de l'information de la recherche d'information :

- **recherche d'information** : « Ensemble des méthodes, procédures et techniques permettant, en fonction de critères de recherche propres à l'utilisateur, de sélectionner l'information dans un ou plusieurs fonds de documents plus ou moins structurés ».
- **recherche de l'information** : « Ensemble des méthodes, procédures et techniques ayant pour objet d'extraire d'un document ou d'un ensemble de documents les informations pertinentes ».

C'est dans cette seconde problématique que je positionne mes travaux. Si la différence peut paraître minime, les recherches associées n'y sont en effet pas identiques.

La recherche d'information

La recherche d'information (RI) est le domaine qui étudie la manière de retrouver des informations dans un corpus. La RI inclut deux aspects, l'indexation et

4. <http://confusedofcalcutta.com/>

5. http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm?RH=OUTILS_VOC&RF=OUTILS_VOC

l'interrogation [Amini 13].

Initialement et de manière sommaire, la recherche d'information dans les images fut basée sur des mots clés ou tags. Il n'y a pas si longtemps finalement, *Google images*⁶ proposait de retrouver des images en recherchant les relations entre la requête et le texte accompagnant les images. Cette recherche s'oppose à la recherche d'image par le contenu ou CBIR (Content Based Image Retrieval). Ce domaine, le plus actif de la RI, repose sur des techniques permettant de rechercher des images à partir de ses caractéristiques visuelles, c'est-à-dire induite de leurs pixels. Les images sont alors classiquement décrites à partir de leur texture, couleur ou forme. Durant sept ans, le challenge Pascal VOC⁷ a permis de fédérer une partie de la recherche mondiale dans ce domaine et de fournir à la communauté un précieux benchmark (petite pensée pour le regretté Mark Everingham [Everingham 10]). La recherche a tellement progressé dans ce domaine que désormais, Google base également ses réponses sur le contenu intrinsèque des images.

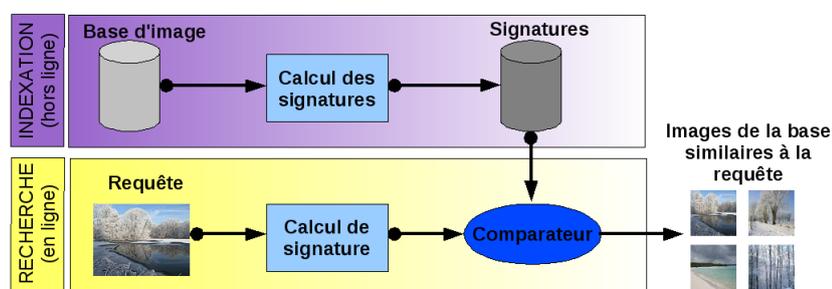


FIGURE 1.1.1 : Description du principe général de la recherche d'images par le contenu

De plus en plus, il est apparu nécessaire de compléter la simple requête par des informations supplémentaires en provenance de l'utilisateur. Le *retour de pertinence* est une approche qui peut, par exemple, prendre en compte un jugement de satisfaction sur les documents présentés à l'utilisateur à l'issue de sa recherche. L'objectif est donc pour le système de savoir quels sont, parmi les documents présentés, ceux qui répondent vraiment au besoin de l'utilisateur [Picard 08], [Buttcher 10].

Ainsi, si l'on considère un domaine d'application précis, et un certain nombre d'hypothèses concernant le contexte d'exécution du système, il est alors possible de concevoir des algorithmes égalant voire dépassant la performance humaine. Cependant, dès que l'on souhaite traiter des tâches plus générales, dans un environnement complexe, de nombreux verrous doivent être levés : ambiguïté dans l'interprétation des données, explosion de la combinatoire liée à l'exploration des données, etc.

6. <https://www.google.fr/about/company/history/>

7. <http://pascalvin.ecs.soton.ac.uk/challenges/VOC/>

Une des voies possibles pour tenter de résoudre ce problème est de chercher des mécanismes efficaces de sélection d'informations *pertinentes* et donc de chercher avant tout à extraire de l'information.

La recherche de l'information

Mais qu'est ce que l'information ?

Si dans un document il manquait une lettre, voire un mot complet, vous seriez à même de retrouver le sens de la phrase, parfois sans même y penser. Vous corrigeriez d'éventuelles coquilles par simple réflexe. Vous êtes entraîné à lire, comme vous êtes entraîné à voir. Taciau tai parasyta lietuviskai, ir jus turite, kaip vaikas, skaityti raide po raides, idant sudetumete zodi, po to sakini⁸. Nous traitons les images comme nous traitons les mots : on extrait des informations clés que l'on assemble pour former une scène porteuse du maximum d'information possible.[Courboulay 02]

Au sens étymologique, l'information est ce qui donne une forme à l'esprit. Elle vient du verbe latin *informare*, qui signifie « donner forme à » ou « se former une idée de ». **Hors contexte**, elle représente le véhicule des données comme dans la théorie de l'information et **hors support** elle représente un facteur d'organisation. E.Morin approfondit la notion d'*information* dans le tome 1 de *la Méthode, La Nature de la Nature* [Morin 81].

Pertinente, utile, saillante, d'intérêt, facteur d'organisation, voilà toute une série de vocabulaire qui s'est présentée lorsqu'il a fallu caractériser l'information.

Une des voies possibles et que j'ai choisie pour mieux appréhender ce problème est de chercher des mécanismes efficaces de sélection des informations pertinentes. On trouve des exemples de ce type de mécanisme dans le règne animal, pour lequel l'évolution a favorisé l'émergence du **système attentionnel**. Chez l'homme, il est étudié intensivement depuis la fin des années 1950 (d'abord par des psychologues, puis des neuropsychologues). Plus récemment, grâce aux neurosciences computationnelles, des modèles informatiques de l'attention (visuelle) sont apparus. Ils permettent de la modéliser et de tenter d'en comprendre le fonctionnement.

L'attention est donc le point d'entrée que j'ai privilégié pour rechercher de l'information.

8. Traduction : "Mais là, c'est écrit en Lituanien et vous devez lire, comme un enfant, une syllabe après l'autre, pour former un mot, puis une phrase". (Merci à Arunas Stockus pour la traduction).

1.2 L'attention comme mesure du potentiel informationnel

Je ne peux préciser mes contributions dans le domaine de l'attention sans en donner une définition. L'une des plus anciennes et reconnues est donnée par le psychologue William James :

« Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others. »

[James 50].

Mais ce que nous percevons est-il vraiment le reflet de ce que nous voyons ? Notre vision du monde semble précise, continue et cohérente. Pourtant, l'étude des différents composants de notre système visuel laisse apparaître une situation bien différente. Un flux constant d'environ 10^8 à 10^9 bits pénètre nos yeux chaque seconde [Koch 06], [Itti 00]. Le traitement de ces données en temps réel est une tâche extrêmement ardue sans l'aide de mécanismes ingénieux qui réduisent la quantité de données visuelles. Nous pouvons encore citer W. James :

Nous ne percevons que la minime partie des impressions dont nous assiège constamment toute notre périphérie sensorielle. Jamais leur somme ne pénètre intégralement dans notre expérience, j'entends dans notre expérience consciente, qui se creuse un lit à travers cette multitude comme ferait un petit ruisseau à travers une large prairie émaillée de fleurs.

Les processus cognitifs complexes tels que la reconnaissance d'objets ou l'interprétation d'une scène s'appuient sur des données qui ont été transformées de manière à pouvoir être manipulables facilement. Ce mécanisme est appelé *attention visuelle*, il fait écho à la recherche d'information vue précédemment. Chez l'homme, l'attention est facilitée par une rétine qui a évolué vers une fovéa centrale haute résolution et une périphérie basse résolution. L'attention visuelle oriente cette structure anatomique vers les parties *importantes* de la scène afin de recueillir des informations plus détaillées.

Une des questions à laquelle je me suis attaché concerne la modélisation des mécanismes dynamiques sous-jacents au calcul de cette orientation.

Les paramètres de cette modélisation sont nombreux, variés et, pour certains, encore à découvrir. Proposer une classification sommaire m'a permis de définir plus complètement ma problématique de recherche. J'ai ainsi déterminé deux classes

complémentaires et cohérentes permettant une approche incrémentale et sectorisée de l'attention visuelle. Les deux classes que je propose sont :

- la connaissance propre à l'observateur ;
- les mécanismes attentionnels qu'il met en jeu.

1.2.1 La connaissance de l'observateur

L'observateur ou plus généralement l'entité qui observe (cela peut être en effet un robot, un ordinateur ou pourquoi pas un animal), dispose d'une connaissance interne fortement variable.

Depuis l'acquisition d'une image par un ordinateur, à l'observation d'un cliché médical par un radiologue, les connaissances impliquées dans l'observation sont excessivement variables. Dans un cas, l'ordinateur « voit » une image sans *a priori*, alors que le praticien « regarde » et cherche l'origine d'une pathologie.

Entre ces deux extrêmes, on trouve l'observation avec *a priori*, celle qui « paramètre » l'observation pour que l'on ne s'intéresse qu'à une forme, région, couleur... Un des exemples classiques est le process de vision industrielle qui recherche un défaut, un niveau, une forme. On trouve enfin l'observation avec bouclage ; j'entends pas là l'observation qui, disposant d'une connaissance donnée, va pouvoir rétroagir sur elle-même afin de s'adapter. Prenons le cas de la vision active qui peut manipuler l'acquisition afin de mieux découvrir son environnement et d'obtenir de meilleures informations concernant ce dernier.

Au final, cette classe comprend les sous-familles présentées dans le tableau 1.1.

Sous-famille	Description	Exemple
sans <i>a priori</i>	l'entité qui observe ne dispose d'aucune connaissance sur ce qu'elle voit	acquisition photographique
avec <i>a priori</i>	l'entité dispose de consignes précises et limitées sur ce qu'elle doit observer	contrôle qualité
avec bouclage	l'entité peut agir sur sa propre observation	découverte d'un environnement
connaissance complexe	l'entité dispose d'une base de connaissance large et complexe afin de regarder, d'analyser et de découvrir à partir d'observations	vision humaine

TABLE 1.1 : Sous-familles proposées pour segmenter les méthodes d'analyse de l'attention en se basant sur la connaissance de l'observateur.

1.2.2 Les mécanismes attentionnels de l'observateur

L'autre classe proposée concerne les mécanismes attentionnels mis en jeu dans l'observation. C'est la lecture de *Prédire n'est pas expliquer* [Thom 09] et d'*Esquisse d'une sémiophysique : Physique aristotélicienne et théorie des catastrophes* [Thom 88] de René Thom qui m'a éclairé sur ces différentes façons de « voir » l'attention. Thom évoque l'aspect dual de la *saillance* et de la *prégnance*. Je reproduis ici quelques extraits de [Thom 88] :

J'appellerai forme saillante toute forme vécue qui se sépare nettement du fond continu sur lequel elle se détache. Si l'on passe du temps à l'espace, alors une forme saillante se dira de tout objet visuellement perçu qui se distingue nettement par contraste par rapport à son fond, l'espace "substrat" dans lequel habite la forme. En général une forme saillante vue aura un intérieur dans le champ visuel ; elle présentera par suite une frontière : son contour apparent.[...] Les prégnances sont des entités non localisées, émises et reçues par les formes saillantes. Lorsqu'une forme saillante capture une prégnance, elle est investie par cette prégnance ; elle subit de ce fait des transformations de son état interne qui peuvent produire des manifestations extérieures dans sa forme : ce sont les effets figuratifs. [...] La perception d'une forme prégnante crée un "effet tunnel" qui précipite le point représentatif dans un bassin périphérique, avec libération d'énergie.[...]

Certaines formes ont (chez l'animal) une signification biologique ; telles sont les formes des proies pour le prédateur (affamé), ou le prédateur pour la proie, le partenaire sexuel en période appropriée... La reconnaissance de ces formes suscite une réaction de grande ampleur chez le sujet : libération d'hormones, excitation émotionnelle, comportement d'attraction ou de répulsion à l'égard de la forme inductrice. J'appellerai prégnantes ces formes, et prégnance ce caractère spécifique. Bien entendu, toute forme prégnante est de ce fait saillante.

Il existerait donc une attention exogène ou bottom-up, *i.e.* une attention *réactive* causée par des stimulus extérieur et une attention endogène ou top-down qui résulte souvent de processus tellement complexes que l'on ne peut les modéliser. Cette attention endogène est basée sur les intentions de l'observateur.

Ainsi, l'attention peut relier la saillance, perçue comme de l'information bas-niveau (voire moyen si l'on parle d'objet), à la prégnance perçue comme un niveau d'intention, un niveau sémantique. On peut passer de l'information *pixel* à l'information *objet* et à l'information *induite par interaction*.

Entre ces deux extrêmes, on trouve les mécanismes permettant une adaptation de l'acquisition bas-niveau ou bottom-up (modification de la focalisation, de couleurs ou de formes recherchées), et également les mécanismes modifiant les régions

d'intérêts, top-down. Au final, cette famille comprend les sous-familles suivantes :

Sous-famille	Description	Exemple
vision passive	acquisition en boucle ouverte, information intrinsèque à l'image	caméra
vision réactive (bottom-up)	acquisition avec rétroaction sur les paramètres bas-niveau	vision industrielle
vision réactive (top-down)	acquisition avec rétroaction sur les paramètres haut-niveau	CBIR avec bouclage de pertinence
vision prégnante	observation et réflexions contextualisées	vision humaine

TABLE 1.2 : Sous-familles proposées pour segmenter les méthodes d'analyse de l'attention en se basant sur les mécanismes attentionnels de l'observateur.

1.3 Démarche scientifique : la *diagonale informationnelle*

Ma problématique scientifique étant précisée :

Comment chercher l'information la plus pertinente lors de la réalisation d'une activité centrée autour de l'image ?

Il me restait donc élaborer mon propre système de recherche. La classification présentée précédemment m'a permis de le définir facilement. Ce système est apparu assez naturellement dès 2007, et a pris la forme de ce que j'appelle une **diagonale informationnelle**. Cette diagonale est celle d'un tableau à double entrée qui sont :

- la connaissance de l'observateur ;
- les mécanismes attentionnels de l'observateur.

J'ai positionné les travaux engagés dans ce tableau et diriger ma recherche dans une démarche cohérente me permettant d'avoir une vision large spectre de la recherche de l'information (voir Figure 1.3).

1.3.1 Exemple de positionnement de travaux

Dans ce chapitre, je présente deux exemples de projets, leur origine, objectifs et leur positionnement dans mon système de recherche.

	Vision passive	Vision réactive bottom-up	Vision active top-down	Vision pregnante
Sans <i>a priori</i>	Thèse (99-02)			
Avec <i>a priori</i>	Collaboration LISA (02-04)	FT-AutiSTIC (05-08)		
Avec bouclage				
Connaissance complexe				Le spect-acteur (07-12)

TABLE 1.3 : Mon système de recherche : *la diagonale informationnelle*

La vision réactive avec *a priori* : le projet FT-AutiSTIC

L'équipe *ImagIN* du L3I s'intéressait aux modèles, architectures et outils permettant de mettre en œuvre un environnement paramétrable, personnalisable et reconfigurable pour la conception et l'exécution adaptative d'activités interactives à exécution contrôlée et adaptative.

Ce domaine de recherche n'est pas au cœur de mes compétences, cependant ayant toujours été convaincu de l'intérêt de la transdisciplinarité (j'y reviendrai dans mon projet de recherche), je me suis donc intégré à l'équipe du professeur Pascal Estrailier qui pilote cet axe.

Au début des années 2000, nous étions en relation avec l'hôpital pédopsychiatrique de La Rochelle. Nous développions une application ludique pour les enfants présentant des troubles autistiques [Sehaba 06].

Les nouvelles technologies étaient et sont encore considérées comme pertinentes pour les personnes autistes dans la mesure où elles permettent de faire intervenir de manière contrôlée les comportements communicatifs et émotionnels dans des activités interactives [Fabri 08, Grynszpan 07, Dautenhahn 02, Breazeal 04]. L'orientation générale visait à mettre à disposition des praticiens un environnement informatique paramétrable, personnalisable et reconfigurable pour la conception et le suivi d'activités interactives. L'environnement accompagnait un enfant autiste lors de sa participation à l'activité interactive définie par le praticien. Il l'assistait, en temps-réel, en le maintenant dans une trame scénaristique prédéfinie, tout en tenant compte de manière personnalisée de son comportement (explicite ou implicite), de ses compétences, de ses préférences et d'un historique.

On exploitait, en particulier, l'observation du suivi du regard et de gestes répertoriés pour analyser le comportement de l'enfant.

En 2005, j'ai répondu à un appel à projet de la fondation France Télécom désormais fondation Orange⁹. Ce projet nommé **FT-AutiSTIC** consistait en ce qui me concerne à :

- **Évaluer les capacités d'attention** et comprendre les stratégies adaptatives de l'enfant en réponse aux stimuli présentés (voir si l'action produite est liée à une compréhension entre la cause et l'effet),
- **Aider à l'interaction** et permettre de modifier les croyances de l'enfant en proposant des images virtuelles qui interagissent avec l'enfant, mais en tenant compte de ses spécificités autistiques, par exemple en ralentissant le mouvement pour que l'enfant puisse en extraire une information générale, utile pour réduire ses troubles émotionnels, langagiers, perceptifs cognitifs. Ainsi, par opposition au besoin de « pareil » sécurisant des enfants autistes, nous devions introduire des procédures d'introduction du « pas-pareil » (aléatoire-non prévisible) afin d'éviter l'enfermement dans l'immuabilité de la répétition,
- **Établir une relation partagée** permettant de donner du sens aux traces et aux mouvements de l'enfant.

J'ai proposé un système logiciel et matériel permettant d'adapter l'exécution des jeux aux besoins et profil de chaque enfant en tenant compte des consignes de l'expert (cf Figure 1.4). J'étais responsable de la tâche *Détection des indices caractéristiques permettant de reconnaître les comportements inventoriés, association indices-comportements*.

Ce projet m'a permis de recruter un ingénieur, Matthieu Perreira Da Silva (futur doctorant) qui a développé un système de suivi de visage par webcam [Perreira Da Silva 09, Perreira Da Silva 08], un logiciel d'annotation de séquences (*L3IAnnote*). Il m'a permis de commencer à encadrer un doctorant A.Ould Mohammed qui n'a pas pu finir sa thèse pour des raisons de santé [Mohamed 06, Ould Mohamed 07a].

Au travers de ce projet, j'ai pu développé une application d'étude de l'attention. Cette dernière tentait d'analyser en temps réel l'orientation du regard d'un apprenant ou d'un enfant autiste, et de déterminer s'il regardait l'écran ou pas. Le suivi du regard était un sujet d'étude à part entière. Il se définissait comme la mesure de la direction spatiale du regard et du comportement oculaire (saccade, clignement, fixation...), mais aussi de ce qui était observé. On souhaitait enrichir cette information par l'analyse comportementale du sujet (fatigue, attention, énervement...) ainsi que par **les éléments saillants de la scène observée** (visages, éléments mobiles...). Ce projet a naturellement trouvé sa place comme intersection de la vision réactive bottom-up avec *a priori* et a donné lieu à un certain nombre

9. <http://www.fondationorange.com/>

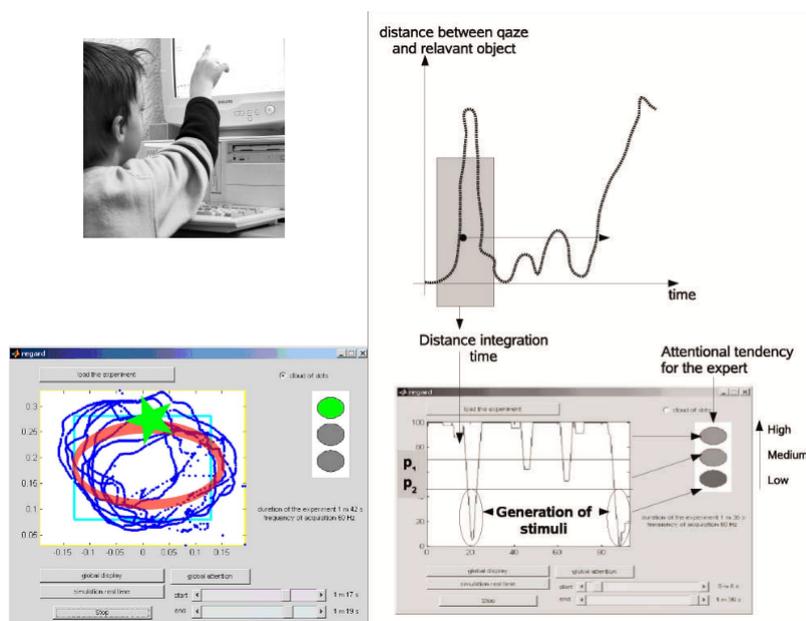


TABLE 1.4 : Détection des indices caractéristiques dans le projet FT-AutiSTIC

de publications [Mohamed 06, Ould Mohamed 07b, Pereira Da Silva 07].

La vision prégnante et complexe : le spect-acteur attentif

En 2007, J'ai proposé de co-encadrer la thèse de Nicolas Rempulski dont le sujet portait sur *la synthèse dynamique de superviseur pour l'exécution adaptative d'applications interactives*. Selon Aristote, l'évolution de la narration doit se faire selon une courbe précise, comme le représente la figure 1.3.1, afin de capter l'**attention** du public et de l'impliquer plus fortement dans l'histoire. Nous souhaitons nous intéresser à l'entité observatrice plutôt qu'à la scène observée, nous voulions mieux comprendre la *prégnance* qui s'incarnait dans un spectateur.

La narration interactive, au centre de la thèse de Nicolas Rempulski, était avant tout une narration, un processus menant à la création d'un récit. La notion d'interactivité caractérisant l'intervention du public dans ce processus. Les parties pris du récit n'étant plus du seul fait de l'auteur. Le public s'approprie un espace d'expression par lequel il va influencer sur le récit. Il ne s'agit pas là d'une influence fictive, au travers de sa perception ou de son imagination, mais bien d'une contribution concrète au récit produit [Rempulski 09, Rempulski 13].

Cette notion clairement issue de la tension dramatique d'Aristote, est généralisée dans l'ouvrage de Raphael Baroni pour l'ensemble des récits et des supports

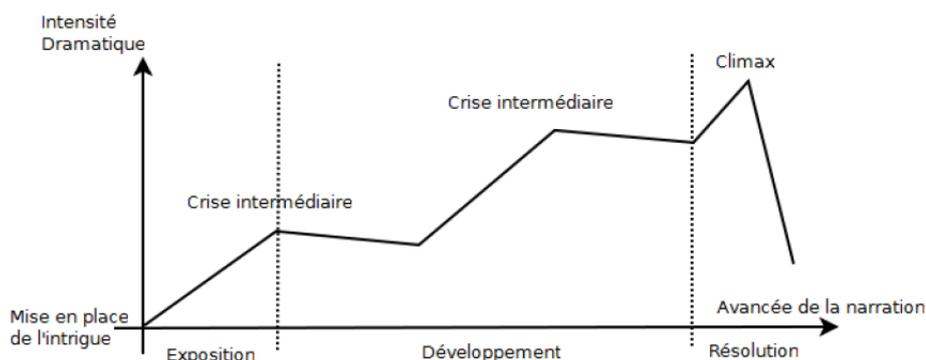


FIGURE 1.3.1 : Arc dramatique d'*Aristote*

de leur fixation. On y trouve la notion de niveau de tension à respecter **pour maintenir l'attention du public** [Baroni 07]. Ce paradigme est exactement celui développé par Mihály Csíkszentmihályi dans son ouvrage [Csíkszentmih 97], qui théorise l'état d'esprit particulier répondant à la définition ci dessus qu'il dénomme **Flow** et qu'il définit comme l'état où l'attention du destinataire est totalement tournée vers le contenu qui lui est proposé. Ainsi la qualité du contenu est directement liée à la qualité narrative perçue par l'utilisateur. Comme dans une œuvre dramatique, les éléments de narration stimuleront l'intérêt et focaliseront l'attention du joueur : surprises, retournements de situation, ou encore rencontres inattendues. Ce travail très théorique a donné lieu à la réalisation d'outils de création et de manipulation d'automates et trouve naturellement sa place dans la vision prégnante avec connaissance complexe. Il a aussi permis le développement d'*AutomatonEditor*, un éditeur graphique de création d'automates (cf. Figure 1.3.3) et de *Noorad*, un éditeur graphique du modèle de narration interactive (cf. Figure 1.3.4). Enfin, nous avons développé *StoryTeller*, un outils de représentation visuelle de l'exécution de la narration interactive (cf. Figure 1.3.5).

1.4 Conclusion

Dans ce chapitre j'ai souhaité présenter ma problématique de recherche, son origine et la structure de recherche que j'ai mis en place pour apporter des contributions à sa résolution.

Les chapitres suivants détailleront de manière plus précise des contributions correspondants à des éléments de la diagonale informationnelle :

- théorie de l'information et modélisation de l'attention bottom-up
- de la recherche de l'information à la recherche d'information
- de l'enrichissement vers la conceptualisation de l'information

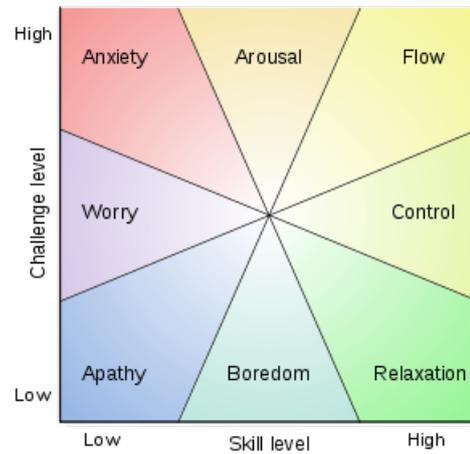


FIGURE 1.3.2 : Caractérisation des états mentaux en fonction des compétences impliquées et des défis à relever.

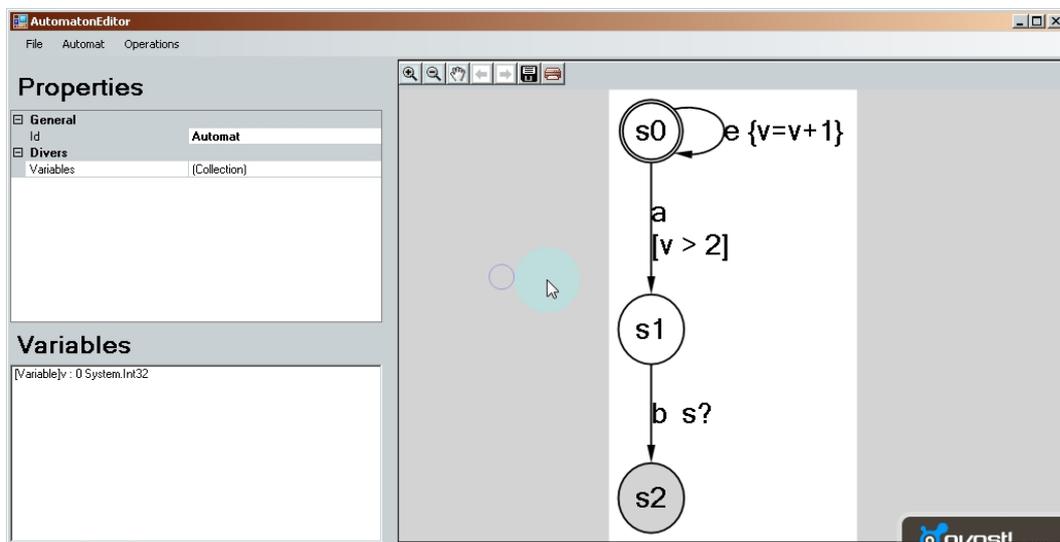


FIGURE 1.3.3 : AutomatonEditor : Editeur graphique de création d'automates (possibilité de visualiser / sauvegarder les compositions)

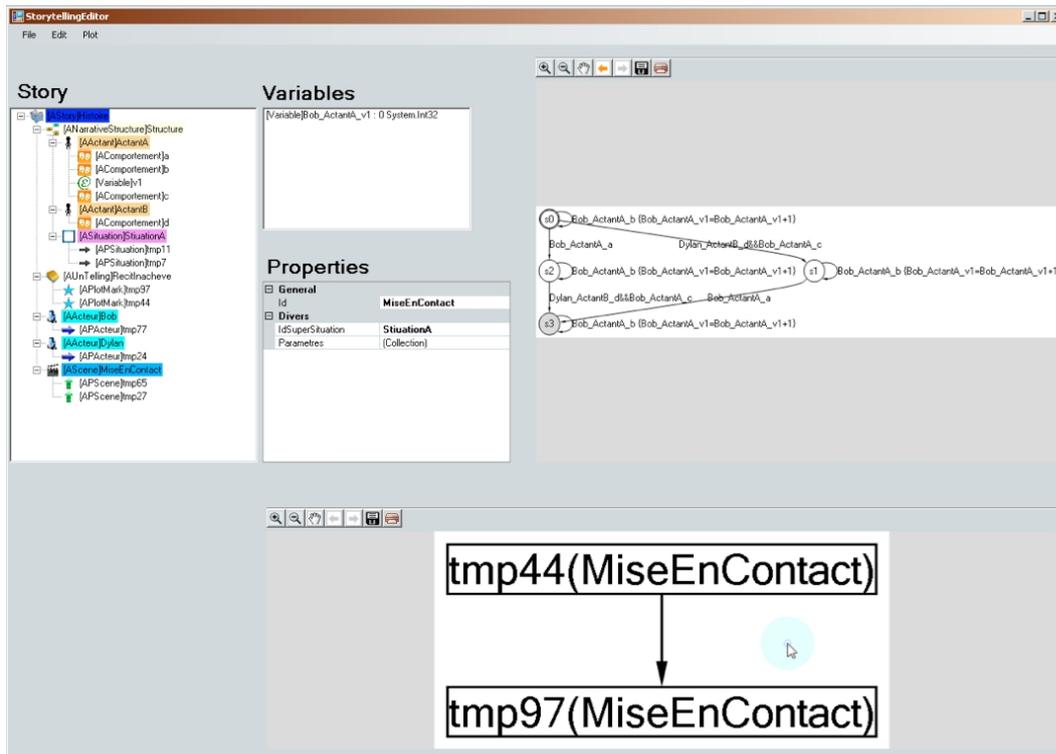


FIGURE 1.3.4 : Noorad : Editeur graphique du modèle de narration interactive. Création des éléments Structuraux (abstrait) , des éléments concrets (instanciés à l'histoire) et de l'arborescence des scènes créées.

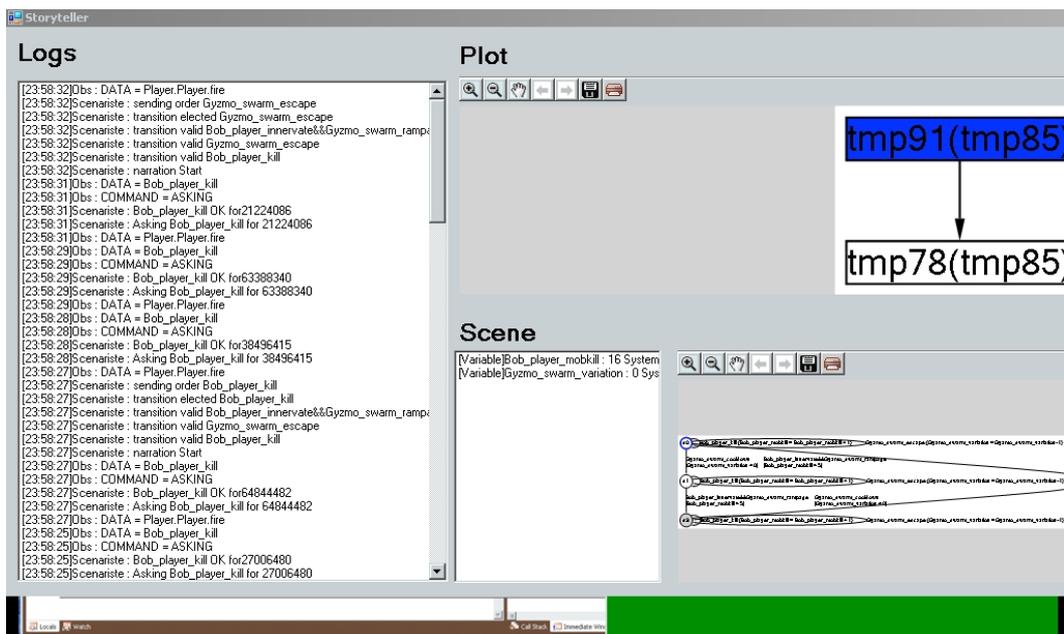


FIGURE 1.3.5 : StoryTeller : Représentation visuelle de l'exécution de la narration interactive

- Projet de recherche : *analyse haut niveau de données hétérogènes massives* (valorisation des archives pénales)

Points clés

Positionnement

- ❑ Définition de la problématique
- ❑ L'attention comme mesure du potentiel informationnel
- ❑ Intégration de la problématique dans une vision matricielle

Contributions

- ❑ Vision intégrée multi-niveaux de la recherche de l'information
- ❑ Rôle clé de l'observateur
- ❑ Intégration de la *pregance* comme concept complémentaire à la saillance

Publications

- Perreira Da Silva M., Courboulay V., Prigent A., Estraillier P. Fast, low resource, head detection and tracking for interactive applications. *PsychNology Journal* 7, 3 (2009) 243-264
- N. Rempulski, A. Prigent, P. Estraillier, V. Courboulay and M. Perreira Da Silva, Adaptive Storytelling based on model-checking approaches, *IJIGS - International Journal of Intelligent Games and Simulation*, 2009, Volume 5, Number 2, 33-42.
- K. Sehaba, V. Courboulay, and P. Estraillier. Interactive system by observation and analysis of behavior for children with autism. *Technology and Disability Journal*, 2006 Volume 18, Number 4 pp 181 – 188.
- Ould Mohamed A., Courboulay V., Ménard M. Analyse de l'attention chez les enfants autistes lors de l'utilisation de logiciels pédo éducatifs In *MAJESTIC 2006 - MAJESTIC 2006*, France (2006)
- Ould Mohamed A., Courboulay V., Sehaba K., Ménard M. Attention Analysis in Interactive Software for Children with Autism In *ACM-SIGACCESS - The Eighth International ACM SIGACCESS Conference on Computers & Accessibility*, United States (2006)
- Perreira Da Silva M., Courboulay V., Prigent A. Gameplay experience based on a gaze tracking system Dans "Gaze-based Creativity, Interacting with Games and On-line Communities" *INPROCEEDINGS in proceedings of COGAIN 2007 (Communication by Gaze Interaction IST FP6 European Project)* - "Gaze-based Creativity, Interacting with Games and On-line Communities" in proceedings of *COGAIN 2007 (Communication by Gaze Interaction IST FP6 European Project)*, Royaume-Uni (2007)

- Perreira Da Silva M., Courboulay V., Prigent A., Estraillier P. Real-Time Face Tracking for Attention Aware Adaptive Games In Computer Vision Systems - 6th International Conference on Computer Vision Systems, Vision for Cognitive Systems, Greece (2008)

Chapitre 2

Modélisation de la dynamique attentionnelle

Dans ce chapitre, je présente les travaux que j'ai réalisés et encadrés depuis une dizaine d'années dans le domaine de la modélisation de la dynamique attentionnelle. L'objectif étant de travailler sur la problématique *vision réactive avec a priori* de la diagonale informationnelle.

	Vision passive	Vision réactive	Vision active	Interaction, <u>pregance</u>
Sans <i>a priori</i>	Restauration d'images isotrope & EPI [2002]			
Avec <i>a priori</i>	Restauration anisotrope d'images et EPI [2004]	MODÉLISATION DE LA DYNAMIQUE ATTENTIONNELLE [2006-...]		
Avec bouclage				
Intégration de l'utilisateur				

J'expose tout d'abord l'approche informationnelle qui m'a permis de proposer un système dynamique basé sur un jeu d'équations de type proies/prédateurs puis je présente le modèle d'attention et des travaux s'y rapportant. Mais dans un premier temps, je présente brièvement différentes façons d'appréhender l'attention dont la dernière est à l'origine d'un des principaux apports théoriques de ce manuscrit.

2.1 L'attention dans tous ses états

Nous avons vu précédemment que l'attention, en tant qu'outil de gestion de l'information, nous permet de construire une perception adaptée à nos capacités et nos besoins. L'efficacité¹ de ce mécanisme passe par une sélection contextualisée des données les plus pertinentes. Cette sélection est d'autant plus importante dans le cas de la vision biologique, que la quantité de données à traiter est massive, de 10^8 à 10^9 bits par seconde et par œil [Koch 06]. Sans l'attention, notre système visuel ne pourrait pas traiter en un temps raisonnable un problème si complexe².

La complexité du phénomène attentionnel a ouvert la porte à une multiplicité d'interprétations théoriques, mais c'est à la fin des années 90 que la puissance de calcul des ordinateurs est devenue suffisante pour envisager une mise en œuvre computationnelle de ces multiples modèles. De nouvelles théories continuent aussi à voir le jour, et sont presque systématiquement accompagnées d'une implémentation sur ordinateur permettant la simulation et donc une nouvelle forme de validation (voir Figure 2.1.1).

Durant les travaux d'Abdallahi Ould Mohamed [Mohamed 06] et la thèse de Matthieu Pereira Da Silva [Pereira da Silva 10a], nous avons recensé les principales approches de l'attention et proposé un état de l'art complet s'y rapportant [Pereira Da Silva 12a].

2.1.1 Attention ouverte ou couverte

L'attention ouverte (*overt attention*) est la plus simple à observer : lorsque nous portons notre attention sur un objet, nos yeux se déplacent afin de fixer cet objet. C'est à partir de cette constatation, somme toute triviale, que les premières ébauches de la notion d'attention ont été définies³. L'un des pionniers de l'étude de l'attention visuelle ouverte est un psychologue russe Iarbovs ou Yarbus [Yarbus 67]. Il a beaucoup étudié le lien entre attention visuelle et mouvement des yeux. Il a notamment montré que la trajectoire de notre regard lors de l'exploration d'une scène dépendait de la tâche demandée, montrant ainsi que l'allocation de notre attention (ouverte) n'est pas un processus uniquement *bottom-up*, mais également *top-down* (voir 2.1.2).

L'attention couverte (*covert attention*) correspond à notre faculté à focaliser notre attention sur une cible (objet ou position) sans déplacer nos yeux. Puisqu'elle n'implique aucun mouvement oculaire, son observation est particulièrement

1. Capacité de produire le maximum de résultats avec le minimum d'effort.

2. Une autre théorie veut que c'est nos limitations corporelles qui oblige à filtrer les informations entrantes afin de nous permettre d'agir.

3. On pourra consulter le chapitre introductif de [?] pour un historique des recherches concernant l'attention visuelle.

2.1 L'attention dans tous ses états

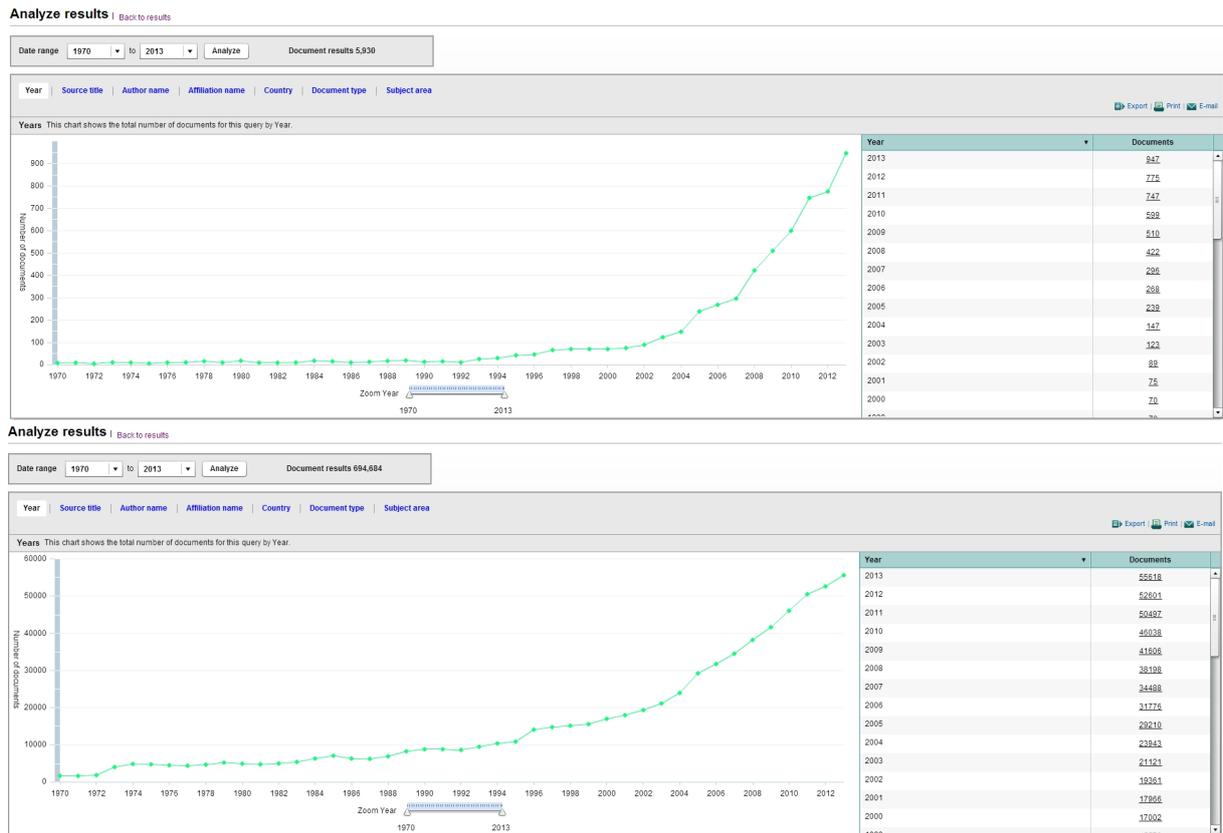


FIGURE 2.1.1 : Evolution du nombre de documents traitant de *saillance* (haut) et d'*attention* (bas) depuis 1970 (source Scopus).

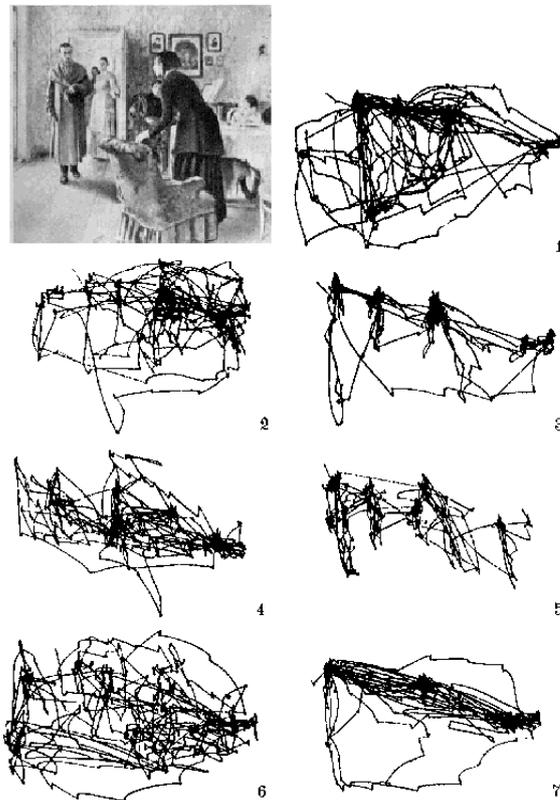


FIGURE 2.1.2 : Les mouvements des yeux sont dépendants de la tâche confiée à l'observateur [Yarbus 67]. (1) observation libre. Avant l'expérience on a demandé aux sujets (2) déterminez la situation matérielle de la famille, (3) donnez l'âge des personnes, (4) essayez de deviner ce que la famille faisait avant l'arrivée du « visiteur surprise », (5) mémorisez les vêtements portés par les personnes, (6) mémorisez la position des personnes et objets dans la pièce (7) estimez depuis combien de temps le « visiteur surprise » n'a pas vu la famille.

difficile. Cependant, dès 1896, on trouve des traces d'expériences d'Hermann Von Helmholtz rapportant ce phénomène. Sa prise en compte dans des modèles d'attention visuelle est bien plus tardive. L'attention couverte est allouée avant l'attention ouverte afin de faciliter le déplacement des yeux vers des *stimuli* « intéressants ».

2.1.2 Attention exogène et endogène

L'attention *exogène* (également appelée ascendante ou *bottom-up*) représente l'ensemble des processus automatiques, déclenchés par les *stimuli* externes captés par notre système visuel. C'est par ces mécanismes que, par exemple, nous tournons la tête si nous percevons un mouvement brusque à la périphérie de notre champ de vision.

L'attention *endogène* (également appelée descendante ou *top-down*) est volontaire et dépend, par exemple, de nos objectifs, elle se rapproche de ce que nous avons appelé la *prégnance*. C'est typiquement le type d'attention que nous mettons en œuvre lorsque nous jouons au fameux jeu « où est Charlie ? » (figure 2.1.3).

Ne dépendant pas de facteurs extérieurs, l'attention *bottom-up* est plus simple à modéliser que son homologue *top-down*. Ainsi, les premiers modèles computationnels d'attention visuelle étaient basés exclusivement sur celle-ci [Koch 85, Itti 98]. Il est cependant difficile chez l'homme de séparer les deux types de traitement (*bottom-up* et *top-down*) [Itti 01].

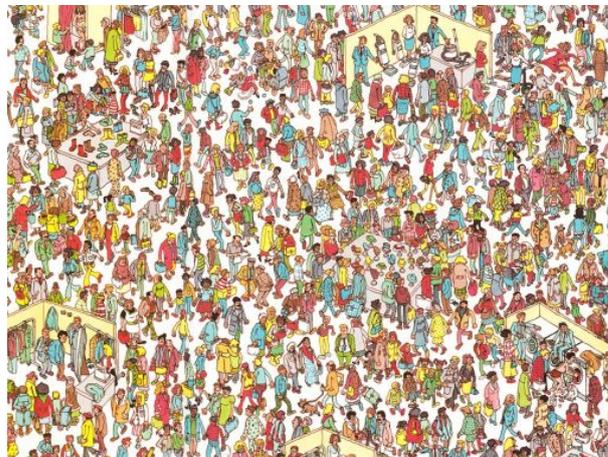


FIGURE 2.1.3 : « Où est Charlie ? », application type de l'usage de notre attention exogène. L'objectif est de trouver un personnage grand et maigre, au pull rayé rouge, avec un bonnet, une canne et des lunettes.

2.1.3 Attention orientée espace ou objet

Les études de psychologie comportementale ont permis de révéler deux modes d'allocation de l'attention visuelle. Posner [Posner 78], montre par exemple que l'attention peut être dirigée dans des zones plus ou moins grandes de la scène visuelle. Ce type de conclusion a permis l'établissement de modèles où l'attention est vue comme une sorte de faisceau [Treisman 80], gradient ou zoom.

D'autres études suggèrent que l'allocation de l'attention pourrait être dirigée vers des objets entiers, qui seraient le résultat d'un regroupement (*grouping*) effectué pré-attentivement. La théorie de l'attention orienté objet permettrait de faire le lien entre attention et théorie de la forme. Élaborée par les psychologues gestaltistes [Guillaume 37], cette théorie présuppose l'existence de différentes règles (continuité, proximité, similitude, etc.) utilisées par notre cerveau pour structurer la scène visuelle. Puisque les deux phénomènes sont observés, il semblerait que les deux mécanismes soient impliqués dans l'allocation de l'attention. Il est cependant difficile de savoir pour l'instant s'il s'agit de deux processus indépendants, ou deux manifestations différentes d'un même processus.

2.1.4 Attention centralisée ou distribuée

L'idée d'une carte centrale de représentation spatiale de l'attention est partagée par de nombreux modèles théoriques. Pour Treisman [Treisman 80], à l'origine de ce concept, elle est appelée *carte générale de localisation* (master map of locations), pour d'autres on évoque une carte de saillance, une carte d'activation ou une carte de priorité. Selon les modèles développés, les caractéristiques et le rôle de cette carte centralisée sont variables, mais l'esprit d'une carte centralisée reste commun. Ce principe est séduisant car il permet de représenter dans une même carte le potentiel attentionnel de toute la scène visuelle. Puisqu'il n'est pas avéré qu'il existe une représentation centralisée de la saillance dans notre système attentionnel, une alternative est de considérer que cette représentation n'existe pas. Dans ce cas, l'attention est distribuée et devient une **propriété émergente de la compétition entre les différents *stimuli* pour obtenir le focus attentionnel.**

2.1.5 Attention et système dynamique

Enfin, dans un article de 1998, Mike Lesser et Dinah Murray [Lesser 98] avancent l'idée que l'esprit *modèle* son environnement ; ils décrivent le lien entre l'expérience et l'activité par une équation définit à partir d'études écologiques menées par Allen et Lesser [Allen 89]. Dans leur approche, l'attention est une ressource fondamentale mais limitée. Leur modèle est basé sur la notion d'*intérêt* ; l'intérêt couvrant chez eux un large spectre de l'attention, du désir à l'émerveillement.

Leur système comprend deux équations différentielles, spatialement discrétisées, similaires à un modèle de réaction-diffusion chimique type *Volterra-Lotka* [Volterra 28]. Les auteurs écrivent en parlant de l'espace d'état généré :

Its metric is cognitive map space with the addition of perception and imagination. Emotion is reduced to a single aversive/attractive value which, in combination with the state of the matrix provides the value of $X(i, j)$. The differential equations are fluctuated to simulate an unknown environment. The brain is modeled as a noisy far from equilibrium network.

Ils sont allés plus loin en simulant l'espace des possibles et sont parvenus à démontrer que le système avait des propriétés émergentes. Là encore, je cite les auteurs :

Which is to say that the model would not only produce interests that were present in its initial state. Furthermore the model also produces interests that had no immediate connection with any other interests. This can be viewed as analogous to the creative function in a human mind.

Dans la section suivante, je propose une nouvelle approche pour modéliser la dynamique attentionnelle.

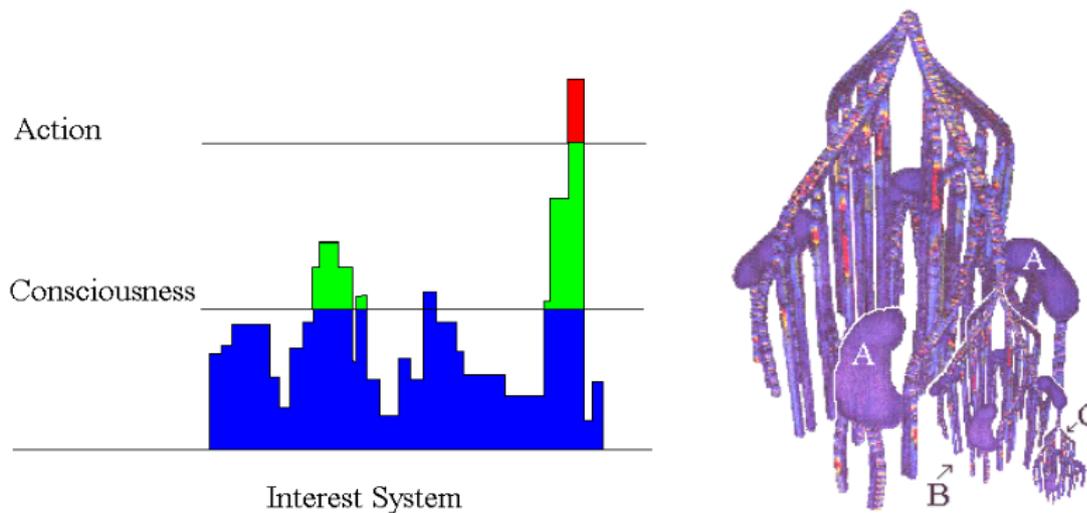


FIGURE 2.1.4 : A gauche : représentation du niveaux de sous-conscience, conscience et activation. A droite : Simulation de l'espace d'état du modèle attentionnel, l'intérêt suscité par un événement soudain est représenté en A (extraits de [Murray 05])

2.2 Information physique extrême et observation

The observer creates his or her local reality.

Cette phrase de Roy B. Frieden est extraite de *Physics from Fisher Information* publié en 1998 [Frieden 98]. Il y définit un principe unificateur de la physique, celui de l'information physique extrême ou EPI. La principale particularité de cette information est qu'elle introduit comme élément à part entière de la mesure, l'observateur effectuant cette mesure.

Lors de mes travaux de thèse, j'ai utilisé cette approche afin de modéliser l'extraction d'information sans *a priori* dans un système fermé [Courboulay 02]. Avant de présenter son application à l'attention, je présente brièvement son principe.

2.2.1 Information physique extrême

Roy B. Frieden a développé une théorie puissante et complète qui propose une approche unifiée de la physique et de l'analyse de données à partir de l'optimisation d'un transfert d'information de Fisher.

L'information de Fisher représente la mesure de la qualité d'estimation d'un paramètre, mais elle mesure aussi l'état de désordre d'un système ou d'un phénomène et peut donc, à ce titre, être considérée comme une pierre angulaire de la théorie physique.

L'objectif principal de l'approche de Frieden est de développer je cite :

a theory of measurement that incorporates the observer into the phenomenon under measurement.

De manière générale, il définit $\mathbf{p}(\mathbf{x})$ comme la fonction densité de probabilité d'un bruit de mesure \mathbf{x} intrinsèque à une acquisition d'un vecteur de dimension n et \mathbf{q}^2 l'amplitude de probabilité réelle définie comme $\mathbf{p} = \mathbf{q}^2$ (voir [Frieden 98] pour une présentation plus complète), l'information de Fisher I peut alors s'écrire comme :

$$I[\mathbf{q}] = 4 \sum_i \int dx_i \sum_v \left(\frac{\partial q_i}{\partial x_{iv}} \right)^2 \quad (2.2.1)$$

, où $q_i \equiv q_i(x_i)$ est la i^{th} amplitude de probabilité de la fluctuation associée à la mesure $x_i = (x_{i1}, \dots, x_{ip})$.

Frieden propose alors d'injecter une information physique concernant le scénario de mesure sachant que la forme de \mathbf{I} est définie. Son interrogation principale est "*how to do this from a physical (and not ad hoc) standpoint?*" Quelle contrainte est imposée par la nature physique de la mesure? Frieden propose de considérer que la mesure consiste en un transfert, un flux d'information depuis l'entité d'intérêt vers l'observateur :

$$J \rightarrow I \quad (2.2.2)$$

, où J représente l'ensemble des contraintes imposées par le phénomène mesuré.

Il est intéressant de noter que la traduction anglaise de ce flux d'information est **flow**, le même nom que l'état où l'attention du destinataire est totalement tournée vers le contenu qui lui est proposé dans la théorie de Mihály Csíkszentmihályi [Csíkszentmih 97].

J est appelée *information de Fisher intrinsèque*, sous-entendu, au phénomène mesuré. Autant l'information I a une forme définie, autant J est spécifique au processus étudié.

Le rôle de l'observateur se caractérise par une mise en relation de la mesure et du processus. Il perturbe le phénomène qu'il souhaite mesurer. Il effectue un prélèvement d'information du processus vers l'espace des mesures.

Ce postulat est à rapprocher du démon de Maxwell. Cette expérience de pensée a été imaginée par James Clerk Maxwell en 1871, pour suggérer que la seconde loi de la thermodynamique n'était vraie que de manière statistique. La levée de ce paradoxe a été effectuée par le physicien Léon Brillouin (grand artisan de la théorie de l'information). Il stipule que pour prendre des décisions, le démon est obligé d'observer des particules et donc d'utiliser l'information dont il dispose. La quantité d'information que cela représente est minime, mais si on passe au niveau microscopique, avec 10^{23} fois plus de molécules, l'information ainsi utilisée par le démon de Maxwell, que l'on suppose non disponible par l'observateur macroscopique, est importante.

Mathématiquement, l'EPI est basée sur l'extrémisation de l'expression suivante :

$$\partial(I[\mathbf{q}(\mathbf{x})] - J[\mathbf{q}(\mathbf{x})]) = 0 \quad \text{ou} \quad I - J = \text{Extrem} \quad (2.2.3)$$

Frieden définit aussi un coefficient κ , tel que :

$$I - \kappa J = 0 \quad 0 \leq \kappa \leq 1 \quad (2.2.4)$$

Dans le cas d'un système fermé, le flux d'information est total et κ vaut 1 ; dans le cas d'un système ouvert, Frieden statue que κ vaut $\frac{1}{2}$.

Frieden définit enfin une nouvelle quantité K qu'il appelle *information physique*. K représente une information et plus précisément un manque ou une perte d'information. Déterminer $\mathbf{q}(\mathbf{x})$ revient donc à déterminer $K[\mathbf{q}(\mathbf{x})]$ en tant qu'extremum.

$$I - J \equiv K = \text{extrem} \quad (2.2.5)$$

Frieden identifie ce principe variationnel comme traduisant l'action de la mesure sur le processus physique considéré. Les solutions obtenues (\mathbf{q} ou \mathbf{p}) sont alors interprétées comme des fonctions qui permettent *la meilleure mesure de l'information* [Frieden 98, Frieden 04].

2.2.2 Dérivation d'une dynamique attentionnelle

Lorsque je me suis confronté à l'analyse de l'attention des enfants autistes dans le cadre du projet FT-AutiSTIC, je ne pouvais bien évidemment plus utiliser cette approche telle quelle, le système n'étant pas fermé au sens physique du terme. Dans les systèmes fermés, il n'y a pas moyen de gagner, de perdre ou d'intégrer des éléments *a priori*.

Ce monde clos n'est évidemment pas adaptée à l'observation visuelle, dans lequel l'environnement et la *prégnance* de l'observateur influencent le processus. L'observation, au sens commun du terme, appartient clairement à un système ouvert, d'échange entre mondes intérieurs et monde extérieur, c'est une tâche de traitement de haut niveau correspondant à un flux d'informations constant depuis notre rétine et ses détecteurs (couleur, l'orientation et l'intensité) [Le Meur 06b] vers une représentation complexe dans l'obscurité de notre cerveau.

Je me suis donc penché sur les extensions possibles à des systèmes ouverts et dans [Frieden 04], il expose :

*EPI used in an **open** system will permit us to conclude that EPI applies hollistically to the observer, i.e., both to his external world and to his internal world :*

(a) *externally*, defining the microscopic and macroscopic phenomena he purposely observes, including the ecology surrounding him ; and

(b) *internally*, defining the growth of the competing populations (functioning, cells, bacteria, possible cancer cells, etc.) that exist within his body. These govern his ability to lie and, ultimately, to **observe**.

Cette citation a été le point clé de mes travaux concernant la modélisation de la dynamique attentionnelle.

2.2.2.1 Hypothèses

Dans cette partie je présente les hypothèses que j'ai utilisées pour appliquer le principe EPI à l'observation.

- Premièrement, l'observation visuelle conduit à une représentation globale et optimale construite à partir de caractéristiques bien connues fournies par notre rétine (couleur, intensité et orientation [Le Meur 06b]).
- Deuxièmement, j'adopte l'hypothèse de Lesser [Lesser 98] qui propose que chaque caractéristique bas niveau extraite, se comporte comme une population d'individus. Chaque population se *nourrissant* d'information lui étant associée dans la scène observée. Je propose ainsi que puisque la capacité de notre rétine est limitée, la population de ces individus est, elle aussi, limitée et constante

égale à N . Soient $m_n = 1, \dots, N$ le niveau de chacune de ces populations. Le niveau global de ces populations sur la rétine est alors de :

$$M \equiv \sum_n m_n$$

- Enfin, je propose que l'intérêt soit considéré comme une population à part entière, consommant ces populations d'informations bas niveau, et que la dynamique attentionnelle soit reliée à l'évolution du maximum de densité de cette population.

Une fois ces hypothèses posées, je propose de réutiliser les travaux de Frieden concernant les processus de transport et de croissance [Frieden 04]. Ces processus sont discrets et dynamiques et imposent aux populations d'interagir, grandir et mourir. Ainsi m_n varie en fonction du temps. Pour retomber dans le cadre de l'EPI, je définis les populations relatives comme :

$$m_n/M \equiv p_n,$$

et $\mathbf{p} = [p_1, \dots, p_N]$.

Je propose d'utiliser $N = 4$ avec, dans un premier temps, quatre types de population. Trois familles de populations bas niveau : couleur, intensité et orientation, ainsi qu'une population représentant l'intérêt.

L'intérêt de notre approche est qu'elle peut s'étendre à toutes sortes de population.

J'ai proposé l'intégration d'une population représentant le mouvement [Perreira Da Silva 11b] pour les vidéos (cf. section suivante) ou une population représentant la profondeur et l'isolement pour l'attention 3D ([Courboulay 14]).

Une fois ces hypothèses faites, il suffit d'appliquer le principe de l'information physique extrême pour trouver les fonctions traduisant la *meilleure* façon d'extraire de l'information lors d'une observation, *meilleure* par rapport à la quantité d'information transmise lors d'une mesure sans *a priori* dans un système ouvert.

2.2.2.2 Loi d'évolution de la dynamique attentionnelle

En se référant par exemple à [Fath 04, Fath 03, Frieden 04], l'information de Fisher discrétisée s'écrit :

$$I(t) = \sum_n \frac{z_n^2}{p_n}, \quad p_n \equiv p(n|t), \quad z_n \equiv \frac{dp_n}{dt} \equiv z(n|t) \quad n = 1, \dots, 4 \quad (2.2.6)$$

Nous avons vu dans la section précédente que le **flow** d'information pouvait d'écrire :

$$I - \kappa J = 0 \quad 0 \leq \kappa \leq 1 \quad (2.2.7)$$

et que dans un système ouvert $\kappa = \frac{1}{2}$.

Frieden a montré que dans un cas de figure similaire, J pouvait s'écrire :

$$J \equiv \sum_n J_n(\mathbf{g}, \mathbf{p}, \mathbf{z}, t), \quad \mathbf{g} \equiv (g_n(\mathbf{p}, t), d_n(\mathbf{p}, t), n = 1, \dots, 4) \quad (2.2.8)$$

Ainsi K est donné simplement par :

$$K \equiv \sum_m \left[\frac{z_m^2}{p_m} - J_m(\mathbf{g}, \mathbf{p}, \mathbf{z}, t) \right] = \text{extrem} \quad (2.2.9)$$

Extrémiser l'équation précédente revient à différencier par $\partial/\partial z_n$, ce qui donne :

$$2 \frac{z_n}{p_n} - \sum_m \frac{\partial J_m}{\partial z_n} = 0, \quad n = 0, \dots, 4. \quad (2.2.10)$$

Après quelques calculs décrits complètement dans [Frieden 04], on obtient :

$$z_n \equiv \frac{1}{2} f_n(\mathbf{g}, \mathbf{p}, t) p_n \quad (2.2.11)$$

où :

$$f_n(g, p, t) = 2 [g_n(\mathbf{p}, t) + d_n(\mathbf{p}, t)], \quad (2.2.12)$$

Finalement j'obtiens :

$$\frac{dp}{dt} = [g_n(\mathbf{p}, t) + d_n(\mathbf{p}, t)] p_n, \quad n = 1, \dots, 4. \quad (2.2.13)$$

Cette équation traduit la dynamique optimale au sens d'un **flow** d'information, du processus d'observation, *i.e.* du transfert de l'information visuelle des indices bas niveau (intensité, couleur, orientation, ...) vers l'observateur sans *a priori*. Cette équation est similaire à l'équation de Voltera-Lotka confirmant les intuitions de Lesser et Murray [Murray 03].

Dans la thèse de Matthieu Perreira Da Silva, nous avons proposé une implémentation temps réel et une validation de ce modèle que nous avons appelé *compétitif*, compte tenu du fait que sa dynamique suit une équation de type proies/prédateur. Je présente rapidement cette implémentation dans la section suivante.

2.3 Description du modèle et implémentation proies/prédateurs

L'attention visuelle peut donc être vue comme une compétition entre différentes sources d'information et l'attention, comme l'illustre l'analogie *champêtre* présentée Figure 2.3.1

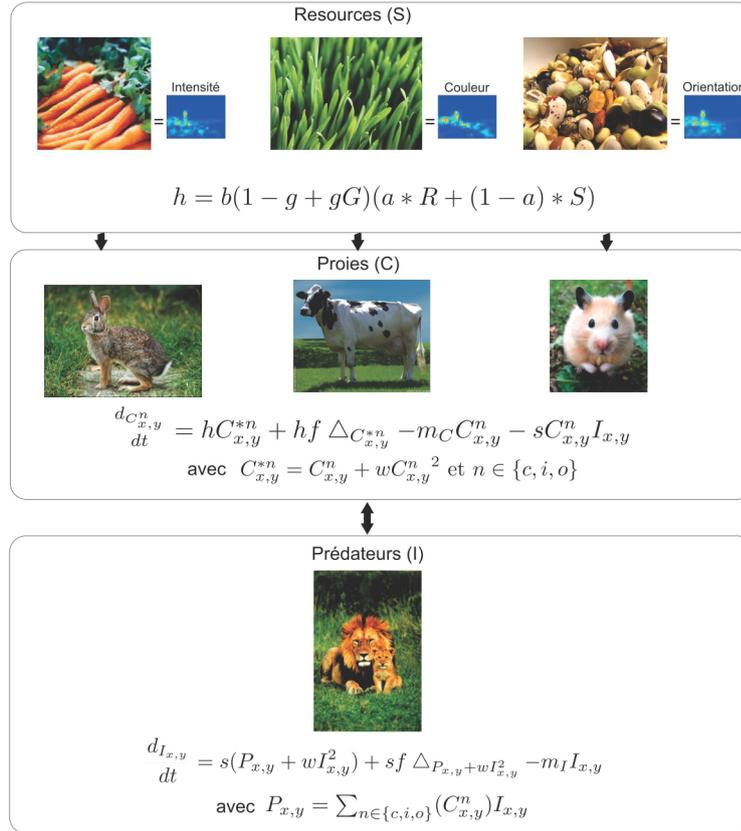


FIGURE 2.3.1 : Architecture du système proies/prédateurs de fusion de cartes de singularité [Perreira Da Silva 12a]

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

Pour chacune des quatre cartes de caractéristiques (couleur, intensité, orientation et mouvement) calculées, l'équation de la matrice des proies C est régie par l'équation (2.3.1), directement dérivée de l'équation (2.2.13) :

$$\frac{dC_{x,y}^n}{dt} = hC_{x,y}^{*n} + hf \Delta_{C_{x,y}^{*n}} - m_C C_{x,y}^n - sC_{x,y}^n I_{x,y} \quad (2.3.1)$$

avec $C_{x,y}^{*n} = C_{x,y}^n + wC_{x,y}^n{}^2$ et $n \in \{c, i, o, m\}$, ce qui signifie que cette équation est valable pour les quatre matrices C^c, C^i, C^o et C^m représentant respectivement la couleur, l'intensité, l'orientation et le mouvement. C représente la curiosité générée à partir de la singularité intrinsèque de l'image. Elle est créée à partir d'une combinaison h de quatre facteurs :

$$h = b(1 - g + gG)(a * R + (1 - a) * SM_n)(1 - e) \quad (2.3.2)$$

- la singularité SM_n de l'image (avec $n \in \{c, i, o, m\}$), dont la contribution est inversement proportionnelle au facteur a ;
- une source R de bruit aléatoire, simulant le haut niveau de bruit de l'activité de notre cerveau [Fox 07] et dont a définit l'intensité par rapport à S . Les équations différentielles modélisant l'évolution de notre système proies/prédateurs deviennent alors des équations différentielles stochastiques. On pourra, en faisant varier a , donner un peu de liberté au système attentionnel et lui faire explorer des zones moins saillantes de l'image ou au contraire, contraindre le système à ne visiter que les zones de forte singularité ;
- une carte gaussienne G permettant de simuler la préférence centrale observée lors des expérimentations psycho-visuelles. L'importance de cette carte est modulée par le facteur g ;
- l'entropie e de la carte de singularité (couleur, intensité ou orientation) normalisée entre 0 et 1. La modulation par $(1 - e)$ permet de favoriser les cartes possédant un nombre limité de maximum locaux. Traduit en termes de proies/prédateurs, on favorise la croissance des populations de proies les plus organisées (regroupées en un petit nombre de sites).

L'évolution de la matrice I des prédateurs consommant ces 3 types de proies est régie par l'équation (2.3.3) :

$$\frac{dI_{x,y}}{dt} = s(P_{x,y} + wI_{x,y}^2) + sf \Delta_{P_{x,y} + wI_{x,y}^2} - m_I I_{x,y} \quad (2.3.3)$$

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

Le terme quadratique modulé par le facteur w permet de renforcer la dynamique du système et facilite l'émergence d'un comportement chaotique en favorisant la

saturation de certaines valeurs des matrices. Enfin, la curiosité C est consommée par l'intérêt I et le point de fixation à un instant t est le maximum de la carte d'intérêt⁴. Pour permettre un changement moins fréquent de la position du focus d'attention, j'ai proposé d'ajouter un mécanisme optionnel d'hystérésis permettant de ne changer le focus d'attention que si le nouveau maximum de la carte des prédateurs dépasse l'ancien de plus d'un certain seuil.

2.3.2 Validation du modèle

Dans [Perreira Da Silva 10b], nous avons présenté une évaluation subjective de notre modèle. Dans [Perreira Da Silva 11b], nous l'avons confirmé évaluation plus classique et objective. Nous avons utilisé la cross-correlation, la divergence de Kullback-Leibler et le normalized scanpath saliency (NSS) dans une étude comparant notre modèle à 6 autres modèles ainsi qu'à une vérité terrain. Nous avons ainsi montré que notre modèle est hautement :

- stable
- reproductible
- exploratoire, i.e. qu'il permet de définir une stratégie d'exploration de la scène (curieux, focalisé...)
- dynamique
- plausible
- rapide
- hautement configurable

Toutes les mesures présentées dans cette section ont été effectuées sur deux bases d'images. La première est proposée par [Bruce 09]. Elle est composée de 120 images couleur présentant des rues, des jardins, des véhicules ou des bâtiments. La seconde, proposée par [Le Meur 06b], contient 26 images couleur. Elles représentent des scènes de sport, animaux, bâtiments, scènes d'intérieur ou paysages.

Compte tenu des nombreux modèles qui existent dans la littérature, il est cependant difficile de confronter et d'évaluer notre modèle totalement générique. Pour pallier ce problème, Tilke Judd du *MIT* a proposé un jeu de données de référence contenant 300 images naturelles avec vérité terrain issue de 39 observateurs. Comme elle l'écrit, *this is the largest data set with so many viewers per image*. Elle calcule la performance de modèles à prédire les fixations de la vérité terrain à l'aide de trois mesures différentes : une courbe ROC, une mesure de similarité, et une

4. Ce paramètre de curiosité permet une simulation de comportements de type autistique que sont le polytropisme et son contraire, le monotropisme.

Le monotropisme est une caractéristique typique de l'autisme. Cela correspond à une aire de répartition limitée des intérêts, la personne atteinte de monotropisme ne peut prêter attention qu'à ce qui est dans son tunnel d'attention. À l'inverse le polytropisme pourrait s'apparenter à une hyper activité du focus attentionnel.

distance. Nous avons téléchargé la base de données, avons appliqué notre modèle pour créer des cartes de saillance de chaque image et les lui avons renvoyées. Nous présentons les résultats dans le tableau 2.1.

Model Name	Area under ROC curve (higher is better)	Similarity (higher is better)	Earth mover's distance (lower is better)
Chance	0.503	0.327	6.352
Achanta	0.523	0.297	6.854
Itti&Koch	0.562	0.284	5.067
SUN saliency	0.672	0.34	5.088
Hou & Zhang	0.682	0.319	53368
Torralba	0.684	0.343	7.715
Context-Aware saliency	0.742	0.39	4.9
Preys/predators model	0.7496 (7)	0.4147 (8)	4.7735 (11)
Itti&Koch 2	0.75	0.405	4.56
Bruce and Tsotsos AIM	0.751	0.39	4.236
Random Center Surround Saliency (Narayan model)	0.753	0.42	3.465
RARE2012	0.7719	0.4363	4.1019
Saliency for Image Manipulation	0.774	0.439	4.137
Center	0.783	0.451	3.719
CovSal	0.7999	0.4869	3.4224
Graph Based Visual Saliency (GBVS)	0.801	0.472	3.574
Judd et al.	0.811	0.506	3.13
Humans	0.922	1	0

TABLE 2.1 : Benchmark de différents modèles de saillance. Le numéro à l'intérieur des parenthèses est le rang de notre méthode.

Sans aucun *a priori*, dans la seule démarche de la recherche de l'information pertinente, notre modèle obtient un rang très intéressant, arrivant derrière des modèles non dynamiques ou adaptés à la recherche d'objet.

2.3.3 Implémentation

Durant sa thèse, Matthieu Perreira Da Silva a développé un logiciel de démonstration temps réel basé sur ce modèle. Ce logiciel temps réel permet une simulation de la dynamique attentionnelle sur des images et des vidéos, ainsi qu'un contrôle sur les paramètres du modèle.

La figure 2.3.2 présente l'interface de notre logiciel.

Il permet un travail en batch mais propose également une interface intuitive permettant de visualiser la dynamique attentionnelle. Il permet également la visualisation temps réel de toutes les images intermédiaires (Figure 2.3.3). Ce logiciel est toujours utilisé dans de nouveaux projets de recherche comme l'ANR VISIIR⁵, ou l'analyse des émotions relatives aux images [Gbehounou 13].

Courant 2014, nous avons développé une version *SAAS* dans le cadre d'un stage de Master 2 de Axel Jean-Caurant, co-encadré avec Benjamin Duthil ingénieur recruté dans le cadre d'un partenariat avec la société ITESOFT (mais nous en reparlerons lors de l'enrichissement sémantique de documents).

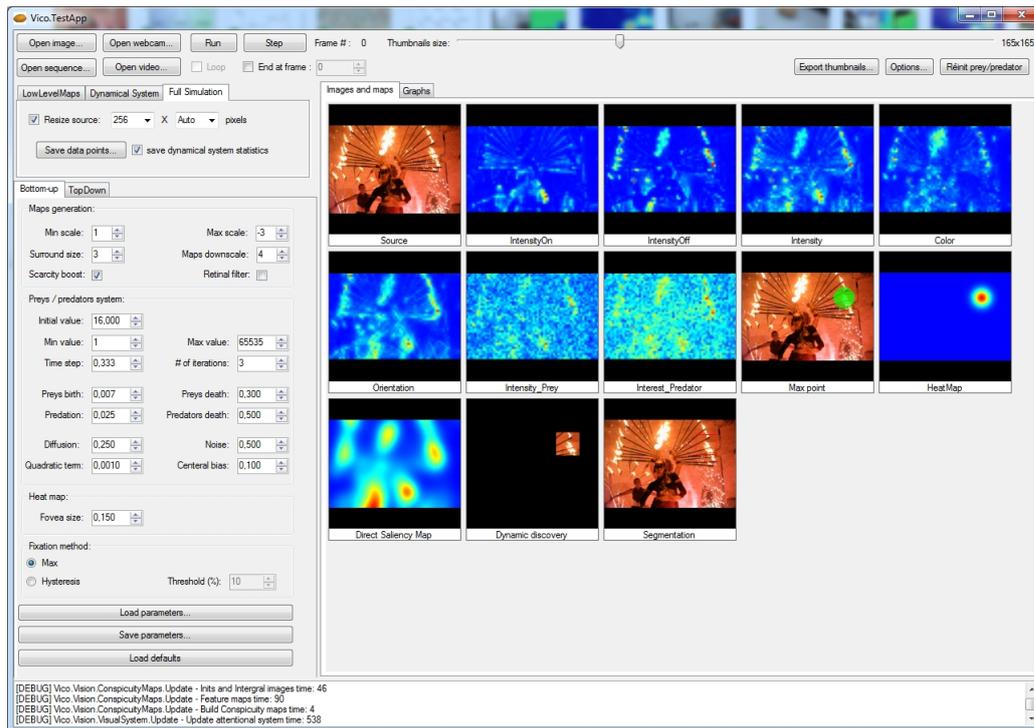
2.4 Applications

Dans cette section, je présente deux applications dérivées de notre modèle [Perreira Da Silva 11a, Perreira Da Silva 12b].

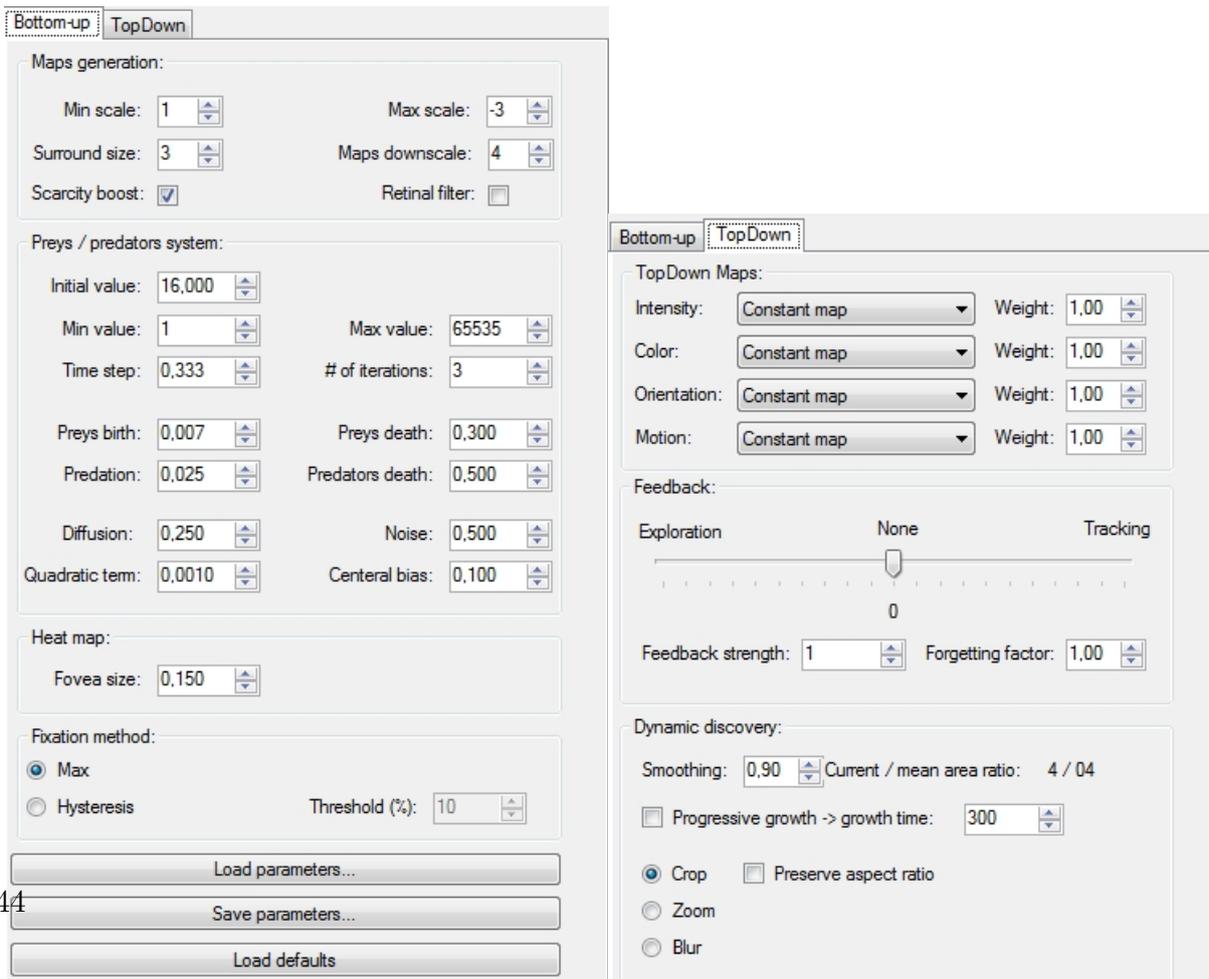
2.4.1 Complexité des images

L'estimation de la complexité d'une image est intéressante, tant d'un point de vue de l'étude psychologique des phénomènes de perception, d'analyse et de mémorisation de l'information visuelle, que d'une exploitation plus informatique. Connaître la complexité d'une image permet par exemple de mesurer qualitativement les performances relatives d'un algorithme de vision ou de traitement d'images en fonction de la complexité des images qui lui sont présentées. Celle-ci peut aussi être une caractéristique intéressante pour un système de recherche

5. Visual Seek for Interactive Image Retrieval : Le projet VISIIR explore de nouvelles méthodes pour l'annotation sémantique d'images. Cette thématique a été largement étudiée lors de la dernière décennie, du fait du très grand nombre de d'applications qui en découlent, dans des domaines aussi variés que la recherche d'information, la vision artificielle, le traitement d'images ou encore l'intelligence artificielle. L'annotation sémantique d'images vise à prédire un concept sémantique à partir du contenu visuel d'une image. Comblé le fossé sémantique entre les données visuelles brutes et les concepts est le but principal poursuivi par les chercheurs du domaine. Les méthodes d'apprentissage supervisé nécessitent un grand nombre de données annotées pour être efficaces. En recherche d'information interactive (CBIR), l'utilisateur introduit une requête image, et la recherche s'effectue avec des boucles d'interaction pour affiner le concept sous-jacent recherché.



(a) Interface générale de notre logiciel de modélisation de la dynamique attentionnelle



(b) Contrôles associés à l'attention bottom-up et top-down

FIGURE 2.3.2 : Interface de notre logiciel de modélisation de la dynamique attentionnelle

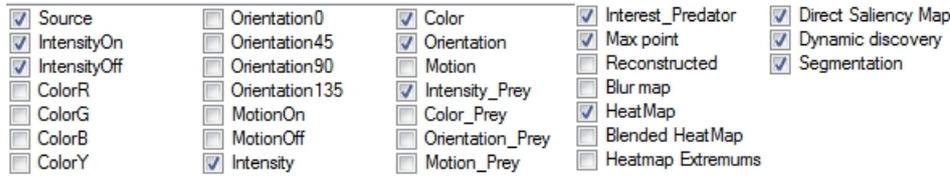


FIGURE 2.3.3 : Cartes accessibles depuis notre démonstrateur

d'images par le contenu : on pourra alors rechercher des images possédant les mêmes attributs de formes, couleur, etc. et de même complexité que l'image requête. La complexité peut également servir à estimer la capacité de watermarking d'une image : plus celle-ci sera complexe, plus il sera facile d'y insérer de l'information sans en dégrader la qualité.

Cependant, son calcul se heurte à un problème majeur : la définition même de la complexité. En effet, celle-ci est relativement floue. Le *Petit Robert* nous donne par exemple la définition suivante :

« Complexe : qui se compose d'éléments différents, combinés d'une manière qui n'est pas immédiatement saisissable. »

Les mesures de complexité découlant d'une telle définition sont alors nombreuses. Lloyd [Lloyd 01], en propose une liste non exhaustive, classée en trois catégories :

- les mesures de difficulté de description,
- les mesures de difficulté de création,
- les mesures de difficulté d'organisation.

Dans le domaine de la complexité des images, les propositions sont moins nombreuses, mais tout aussi variées. Les méthodes utilisées peuvent par exemple être fractales, floues ou, plus classiquement, liées à la théorie de l'information. Concernant ce dernier point, des études récentes montrent que les méthodes liées à la complexité de Kolmogorov (mesure du taux de compression GIF ou JPEG des images) permettent d'obtenir une estimation automatique assez fidèle de l'évaluation de la complexité d'images faites par des humains.

J'ai proposé de construire une mesure basée non plus sur la complexité *directe* du contenu des images, mais sur leur complexité de perception, au travers du filtre attentionnel. Nous avons mesuré la complexité des trajectoires de focalisation de notre système attentionnel afin de déterminer si une image est complexe à parcourir. Cette mesure n'avait pas pour objectif de concurrencer ou remplacer les mesures traditionnelles, mais plutôt de fournir un complément autour de la complexité attentionnelle des images.

Notre mesure de la complexité de la trajectoire du focus d'attention est basée sur les principes de la théorie de l'information et plus précisément la complexité de Kolmogorov. Cependant, celle-ci n'étant pas directement calculable, nous en effectuons une estimation *via* deux méthodes. L'une utilise la compression (principe

du minimum description length), l'autre l'entropie spectrale de la longueur des saccades.

- Taux de compression des coordonnées de la trajectoire : cette méthode exploite le principe du *minimum description length* afin d'estimer la complexité de la trajectoire du focus d'attention. D'après ce principe, plus la trajectoire sera complexe, plus elle sera difficile à compresser efficacement (puisqu'*a priori* non prévisible). Nous mesurons, pour une liste de coordonnées (x_i, y_i) avec $i \in [0, \dots, nbPoints]$, le ratio entre la taille $T_{RawCoord} = 2 \times 4 \times nbPoints$ des coordonnées des focalisations lorsqu'elles sont stockées sous forme d'entiers en mémoire, et la taille de ces mêmes données, compressées avec la méthode de compression deflate⁶. En comparant les taux de compression obtenus pour différentes images avec un même temps d'observation, on va aussi classer celles-ci selon leur complexité.
- Entropie spectrale de la longueur des saccades : soit une série de données s_k avec $k \in [0, \dots, nbPoints - 2]$, formée par les distances entre les différentes focalisations successives. On aura alors $\forall k \in [0, \dots, nbPoints - 2]$:

$$s_k = \sqrt{(x_k - x_{k+1})^2 + (y_k - y_{k-1})^2}$$

Bien que moins riche spatialement, cette représentation permet une analyse plus fine de la dynamique. La comparaison avec la première méthode permet d'évaluer si cette perte d'information spatiale a une influence importante sur l'évaluation de la complexité.

Nous calculons la transformée de Fourier $S = \mathcal{F}(s)$ de la série de données ainsi générée puis transformons celle-ci en une distribution statistique en divisant chaque coefficient par la somme des coefficients : $p(S_k) = \frac{S_k}{\sum_k S_k}$

L'entropie spectrale est l'entropie E de cette distribution, définie par : $E = -\sum_k p(S_k) \log(p(S_k))$

On peut alors comparer l'entropie obtenue pour différentes images, avec un même temps d'observation.

Nous avons mesuré l'entropie spectrale et le taux de compression des coordonnées des trajectoires des focalisations obtenues grâce à notre modèle attentionnel pour les 147 images des bases « Bruce » et « Le Meur ». Le classement obtenu est cohérent avec le critère de complexité mesuré. Les images comportant un objet très saillant sont considérées comme simples. La trajectoire des focalisations nécessaires pour parcourir ces images n'est pas très complexe, puisque le système reste principalement focalisé sur l'objet le plus saillant. A l'opposé, les images ne comportant pas d'objet particulièrement saillant sont considérées comme complexes car le système attentionnel génère des focalisations dans des endroits variés.

Nous avons présenté nos travaux lors de la conférence ICIP 2011 [Perreira Da Silva 11a].

6. <http://fr.wikipedia.org/wiki/Deflate>

La figure 2.4.1 présente un exemple de classification

2.4.2 Recadrage dynamique

Pour cette application du modèle, je suis parti du constat qu'il existait de nombreux périphériques multimédia mobiles, permettant de visualiser des images et vidéos. Ces périphériques sont généralement de petite taille. L'affichage des photos et vidéos sur ce type d'appareil peut alors être problématique, la faible résolution de l'écran rendant le sujet principal de la photo / vidéo difficile à discerner. Pour faire face à ce problème, Le Meur [Le Meur 06a] a proposé d'utiliser les informations fournies par une carte d'intégration des trajectoires oculaires (obtenues via un oculomètre) ou une carte de saillance (obtenue grâce à son modèle d'attention) afin de recadrer les images en n'en gardant que les éléments les plus saillants. Son approche est efficace, mais souffre de quelques limitations :

- dans sa version originale, elle est limitée aux images ;
- le recadrage des images est totalement statique.

Nous avons donc proposé d'étendre les travaux originaux de Le Meur et présentons une méthode exploitant la dynamique de notre système attentionnel afin d'appliquer un recadrage variable dans le temps, sur des images et vidéos. Ce travail a fait l'objet d'une collaboration avec l'IRCCyN de Nantes et une publication en 2012 dans l'IEEE International Conference on Multimedia and Expo [Perreira Da Silva 12b].

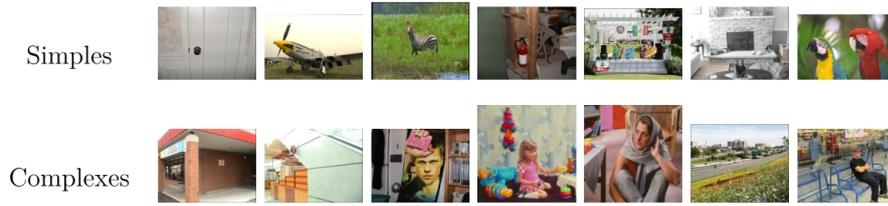
Nous proposons une méthode basée sur le centroïde et la variance de la heatmap. Cette méthode permet de s'affranchir de seuil en prenant en compte tous les pixels de la heatmap, modulant leur influence en fonction de leur intensité. Pour une heatmap H , le centroïde (cx, cy) et les différentes variances vg, vd, vh, vb (représentant respectivement la variance des pixels situés à gauche, à droite, au-dessus et en-dessous du centroïde). Nous avons apporté plusieurs autres améliorations comme l'utilisation d'un facteur d'oubli lors de la construction itérative d'une heatmap qui permet d'obtenir une représentation temporellement limitée de la saillance de l'image ou un recadrage dynamique ajustable en temps réel.

2.5 Extensions du modèle

2.5.1 Extension à la vidéo

L'évolution naturelle de ce modèle appliqué à l'image est bien évidemment l'application à la vidéo.

La conception même du système fait que l'ajout d'une information ne perturbe en rien le modèle de dynamique ; cela consiste à ajouter une nouvelle population



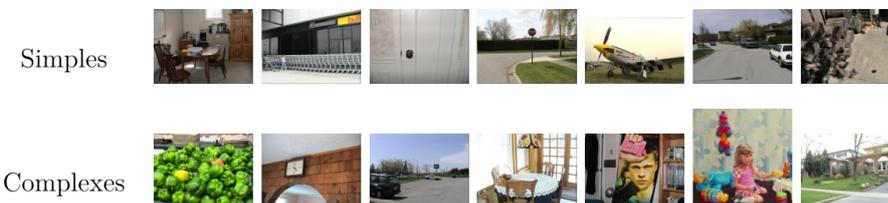
Deflate compression ratio, t=50



Deflate compression ratio, t=150



(a) Classement des 7 images les plus simples et les plus complexes en fonction du temps de simulation, pour la méthode « Deflate compression ratio »



Saccade length fourier entropy, t=50



Saccade length fourier entropy, t=150



Saccade length fourier entropy, t=300

(b) Classement des 7 images les plus simples et les plus complexes en fonction du temps de simulation, pour la méthode « Saccade length fourier entropy »

FIGURE 2.4.1 : Exemples de résultat d'analyse de complexité des images basée sur la dynamique attentionnelle

de proie, en l'occurrence celle reflétant l'évolution de l'image.

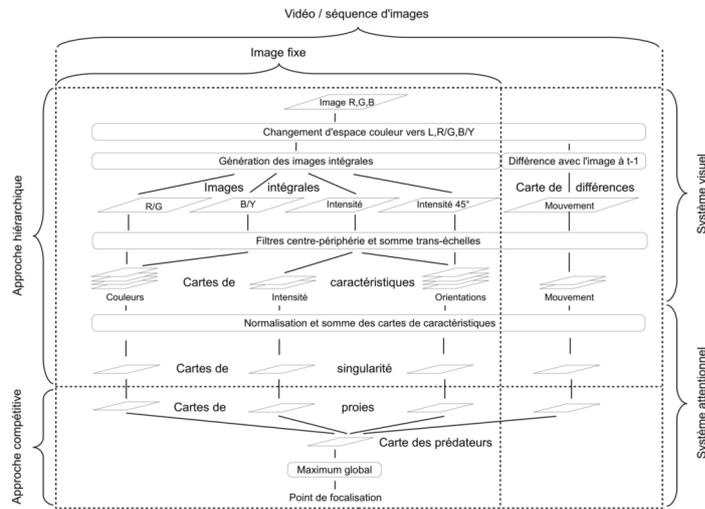


FIGURE 2.5.1 : Schéma général du système d'attention visuelle

Notre système est parfaitement robuste à toute une série de distracteurs. La séquence ci-dessous « Boule + Grille » est illustrative de la robustesse de notre système attention. Une boule rouge se déplace de droite à gauche de haut en bas puis de bas en haut pendant qu'une grille de points noirs se déplace de gauche à droite. Le focus d'attention suit la boule sans perturbation particulière.

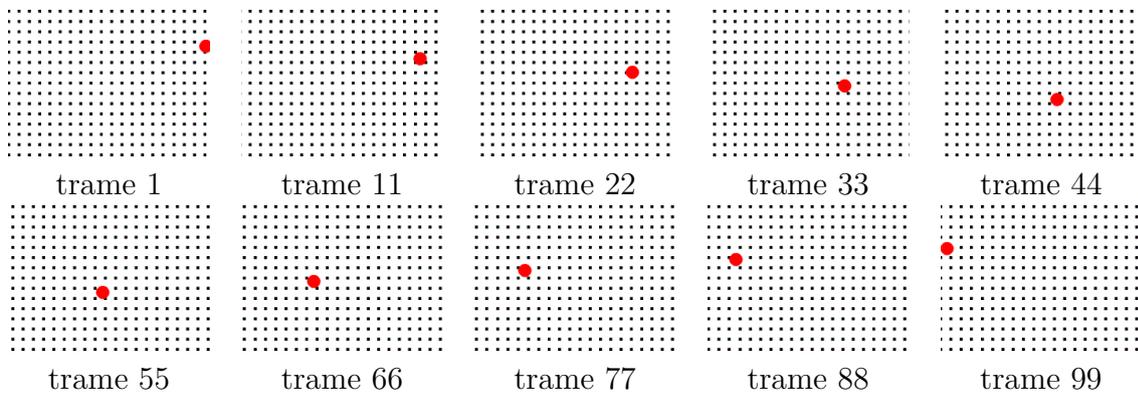


FIGURE 2.5.2 : Exemple de séquence « Boule + Grille ».

Je présente des exemples d'application de notre algorithme de recadrage dynamique à des séquences vidéo de différents types :

- Une vidéo de sport : « TOP 20 Tennis Master Points » (figure 2.5.3). Comme son nom l’indique cette séquence est une compilation de différents extraits de matchs de tennis.
- Une publicité : « Levis - Mr Oizo » (figure 2.5.4).
- Une bande annonce du film « Hancock » (figure 2.5.5).

La vidéo de tennis est consultable sur youtube aux adresses suivantes : <http://www.youtube.com/watch?v=gpQSYG-X10> et <http://www.youtube.com/watch?v=MZQ1WOHIwkg>.

Les deux autres vidéos sont consultables aux adresses suivantes : <http://www.youtube.com/watch?v=-UHRneYLDE> et <http://www.youtube.com/watch?v=CtHdWJUGqSM> (« Levis - Mr Oizo »), <http://www.youtube.com/watch?v=0QgCLHtLlpw> et <http://www.youtube.com/watch?v=0CCjmZAW584> (« Hancock »).

On peut constater que globalement le système propose un recadrage cohérent : les éléments principaux restent dans le cadre, et la vidéo recadrée est toujours compréhensible.

2.5.2 Extention aux *nouvelles images*

En 2012, j’ai proposé à G.Lacombe un stage M2 autour de l’extension du modèle à de *nouvelles images*. Depuis quelques années, en effet émergent des dispositifs grand public délivrant simultanément des images couleur et de profondeur, tels que la Kinect ou les caméras Lightfield (Lytro, Raytrix). Ces capteurs permettent des interactions entre utilisateurs et consoles.

Dans le contexte d’émergence d’application robotique du laboratoire, j’ai voulu voir la faisabilité d’implantation de système attentionnel chez un robot.

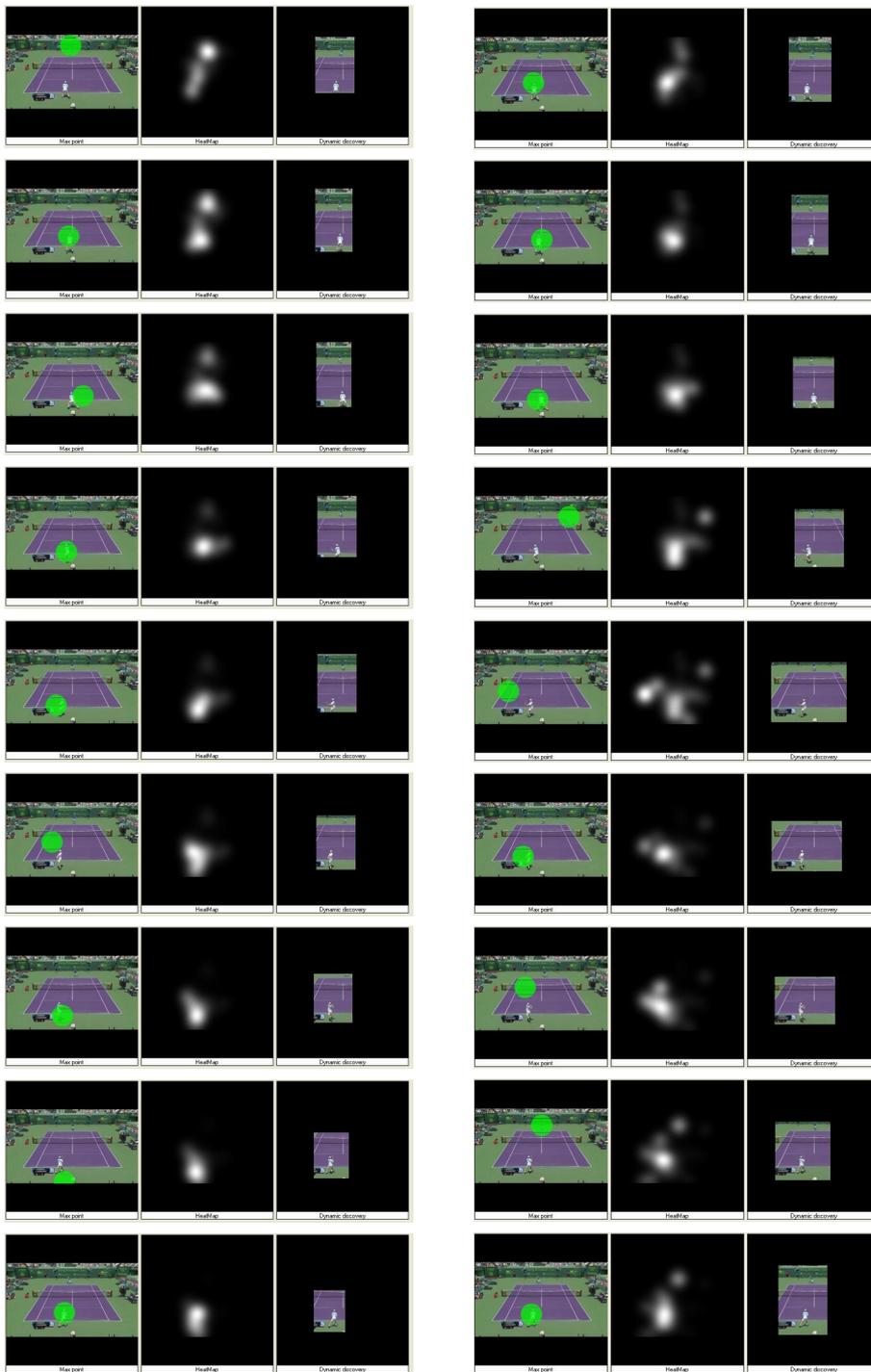
J’ai donc proposé, un modèle computationnel d’attention *implémentable* sur un système robotique. J’ai ajouté deux nouvelles cartes de singularité :

- carte de profondeur : contribue à promouvoir les éléments les plus proches
- carte d’isolement : fait ressortir un élément, même banale ou diffus, mais clairement séparé du reste de son environnement

Ce choix est purement arbitraire et ne s’appuie que sur une volonté d’expérimentation.

La carte de profondeur

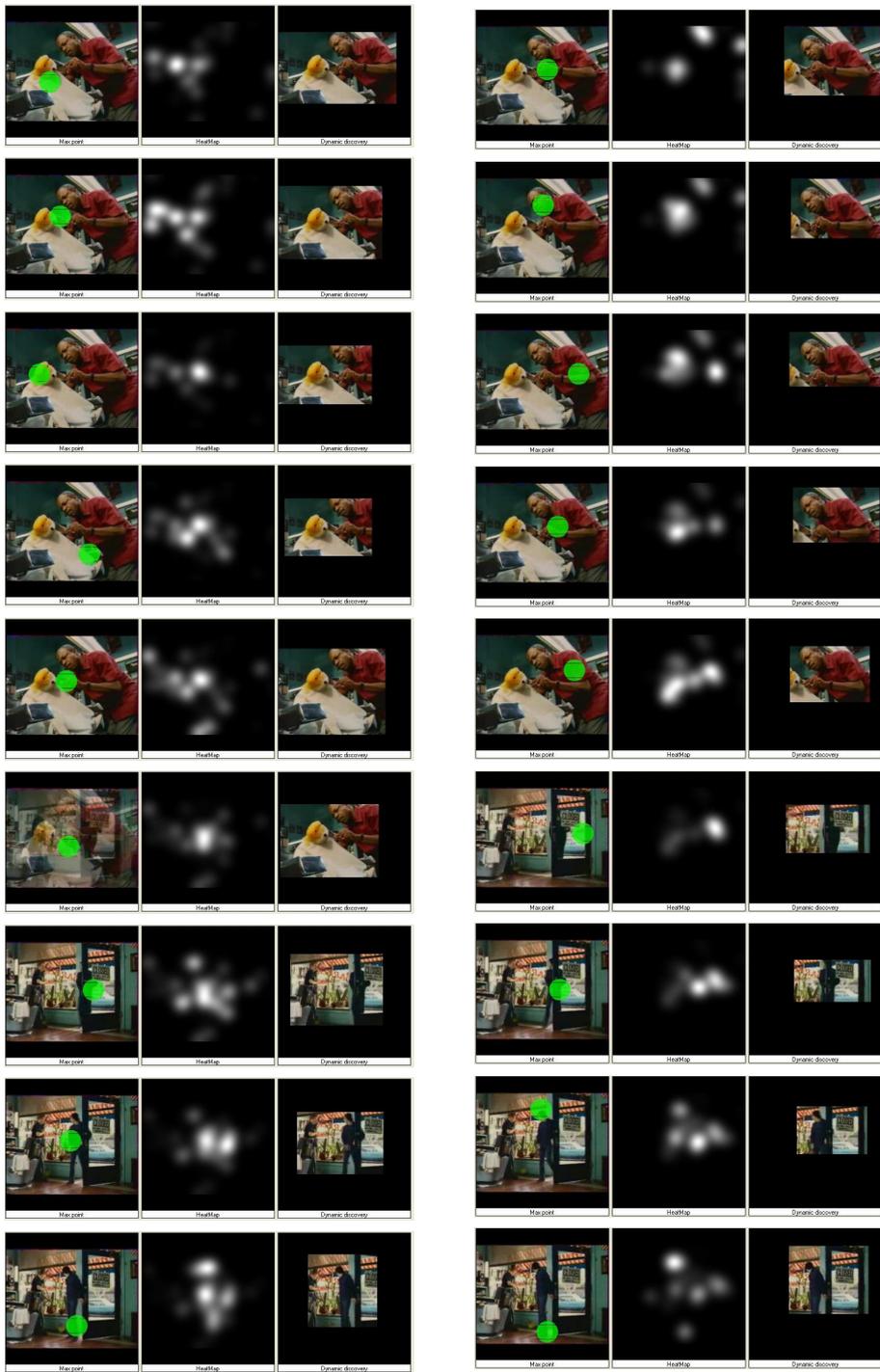
Cette carte représente l’inverse de la profondeur de la scène devant le robot. Nous avons utilisé un système Kinect de Microsoft. Le SDK Kinect fournit aux développeurs l’accès aux flux image et profondeur. Le capteur de profondeur est constitué d’un projecteur infra-rouge associé à un capteur CMOS monochrome, qui



(a) Paramétrage réalisé pour un suivi du joueur, le taux moyen de surface affichée est de 38%.

(b) Paramétrage réalisé pour un suivi de l'échange, le taux moyen de surface affichée est de 46%.

FIGURE 2.5.3 : Quelques trames de la vidéo « TOP 20 Tennis Master Points ».



(a) $Feedback (curiosité) = 0$, le taux moyen de surface affiché est de 69%.
 (b) $Feedback (curiosité) = 0.25$, le taux moyen de surface affiché est de 51%.

FIGURE 2.5.4 : Quelques trames de la publicité « Levis - Mr Oizo ».



(a) *Feedback* (curiosité)= 0, le taux moyen de surface affiché est de 70%. (b) *Feedback* (curiosité)= 0.25, le taux moyen de surface affiché est de 49%.

FIGURE 2.5.5 : Quelques trames de la bande-annonce du film « Hancock ».

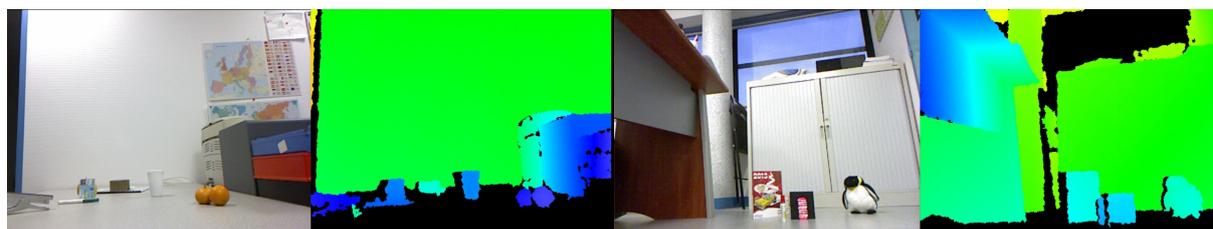
capture des données vidéo en 3D. Chaque pixel de l'image de profondeur correspond à la distance entre la Kinect et chaque objet de la scène. Afin de promouvoir les objets proches plutôt que d'un lointain, nous définissons la carte de profondeur comme l'inverse de l'image de profondeur :

$$C_d(i, j) = \frac{dyn_{C_d}}{dyn_{I_d} * I_d(i, j)} + \alpha \quad (2.5.1)$$

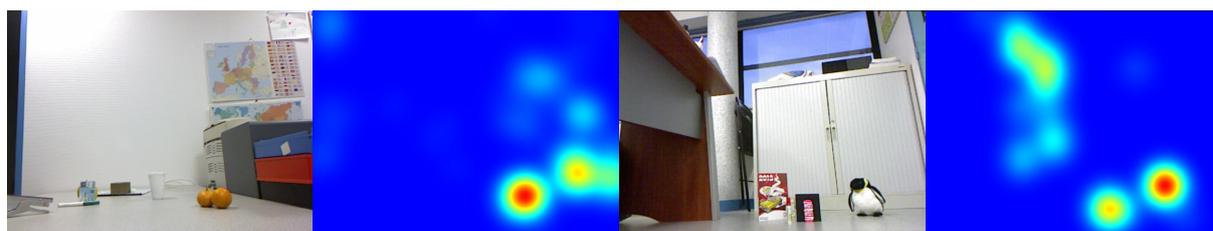
, où dyn_X représente la dynamique de l'image X et α un coefficient pour contraindre la carte à être positive. Chaque valeur nulle de la carte Kinect reste nulle dans la carte de profondeur attentionnelle.

La carte d'isolement

L'objectif de cette carte est de mettre le focus sur un élément isolé. Un élément isolé est caractérisé par une distance au capteur différente de son voisinage. Nous avons décidé d'utiliser la même approche que celle utilisée pour détecter les informations intensité, orientation et couleur. L'entrée devenant la carte de profondeur fournie par la Kinect. Nous calculons, comme dans l'approche d'Itti, les différences *centre-surround* sur différentes échelles pour déterminer le contraste d'une image de profondeur et obtenir ainsi une carte informant sur l'isolement des objets.



(a) Exemples de cartes de profondeur



(b) Exemples de carte d'isolement

FIGURE 2.5.6 : Proposition d'extention de notre modèle attentionnel aux *nouvelles* images

Pour notre expérimentation, j'ai proposé l'utilisation d'un système composée par un système Lego Mindstorm et un Kinect. Le Mindstorm gérant le mouvement,

alors que Kinect gérait l'acquisition vidéo. Les Lego Mindstorms comprennent une intelligence programmable qui contrôle le système, un ensemble modulaire de capteurs et de moteurs. Pour notre test, j'ai relié le Lego Mindstorm et le Kinect à un ordinateur grâce à une liaison USB plutôt que Bluetooth (Figure 2.5.7).

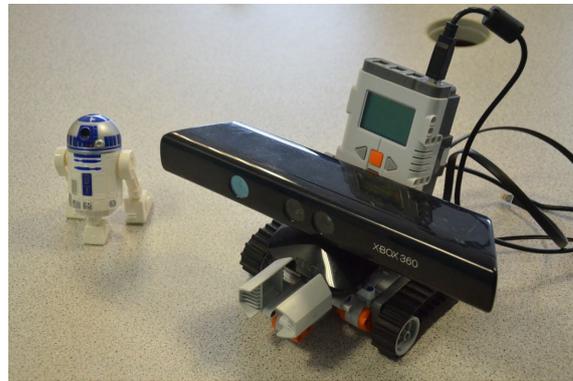


FIGURE 2.5.7 : Notre système composé d'un véhicule Lego Mindstorm et un Kinect (R2D2 voulait absolument être sur la photo)

La tâche confiée au robot était simple, il devait *aller photographier l'objet proche le plus intéressant pour lui*.

Nous présentons une série de résultat dans la figure 2.5.8. Ces travaux ont été présentés lors d'une journée GDR ISIS *Analyse conjointe des images RGB-Depth* [Courboulay 14] et font l'objet d'une soumission commune avec un chercheur de Polytech Mons à une revue internationale.

2.6 Conclusion

Dans ce chapitre, j'ai présenté mes contributions autour de la problématique de *la modélisation de la dynamique attentionnelle* et de *la vision réactive avec a priori* de la diagonale informationnelle.

J'ai tout d'abord présenté une rapide taxonomie de méthodes de modélisation d'attention. J'ai ensuite présenté une contribution théorique permettant de modéliser la dynamique attentionnelle grâce à un système d'EDP de type proies/prédateurs et sa validation.

J'ai exposé son implémentation temps réel ainsi qu'une série d'applications comme la mesure de complexité, le recadrage dynamique et l'extension à la vidéo ou aux images de profondeur.

Les perspectives de ces travaux sont nombreuses. Par exemple, je souhaite travailler sur l'implémentation de mon modèle sur le robot Nao ou son grand frère Roméo (<http://projetromeo.com/>).

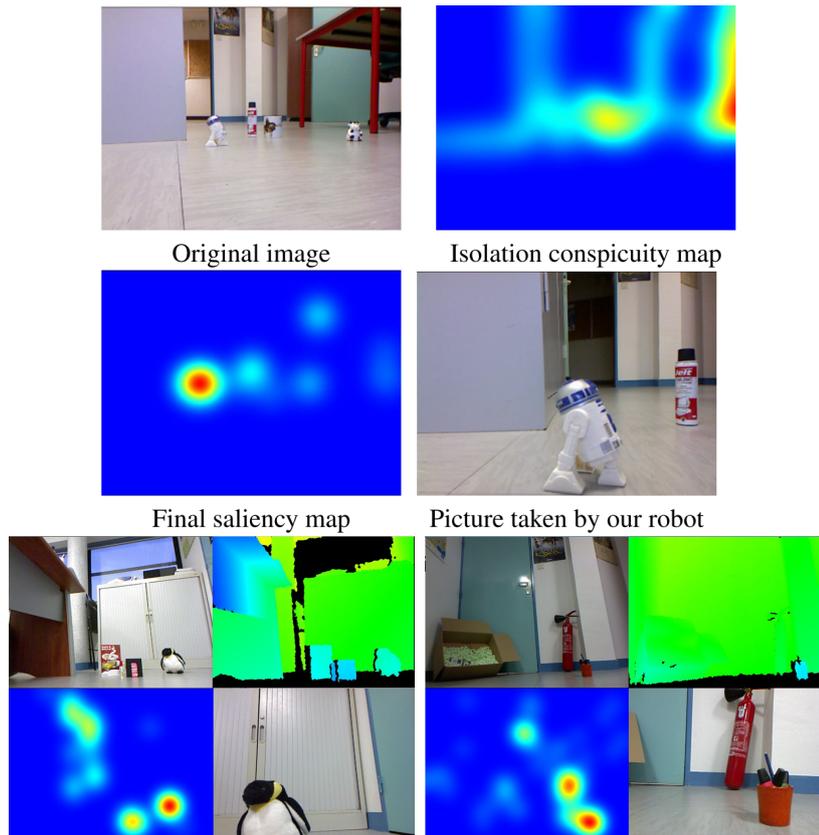


FIGURE 2.5.8 : Exemples de résultats

Suite à un séjour de recherche à la Faculté Polytechnique de Mons au TCTS Lab dans l'équipe de Matei Mancas, nous avons décidé de travailler sur la validation des algorithmes de saillance, non pas sous l'aspect des métriques classiques (NSS, ROC, etc...) mais d'une *application-driven validation* en considérant la tâche de CBIR, là encore ce travail fait l'objet d'une soumission à une revue internationale.

J'envisage aussi, comme autre suite à ce travail, une étude sur la similarité des dynamiques modélisées par rapport à celles enregistrées par eye-tracking et surtout l'étude du bouclage du système pour coller à un comportement oculaire particulier plutôt que moyen pour mieux le comprendre et l'analyser (pathologie, achat sur internet...). Ses travaux sont menés avec Nathalie Guyader de l'équipe Vision and Brain Signal Processing du Gipsa-Lab de Grenoble. Nous envisageons à ce titre le dépôt d'un projet PEPS pour l'année 2015.

Points clés

Positionnement

- ❑ Recherche concernant la deuxième case de la diagonale informationnelle
- ❑ Taxonomie des modèles de la dynamique attentionnelle
- ❑ Développement d'applications et de démonstrateur

Contributions

- ❑ Une approche informationnelle originale de la modélisation attentionnelle
- ❑ Un modèle stable, reproductible, exploratoire, dynamique, plausible, rapide, hautement configurable
- ❑ Une implémentation temps réel et une évaluation complète
- ❑ Des applications variées (complexité, recadrage, robotique attentionnelle)

Publications

- V.Courboulay, M.Mancas, CuriousMind photographer : inciting robots to find out remarkable scenes, EAI Endorsed Transactions on Creative Technologies, to be published.
- Ferreira Da Silva M., Courboulay V. Implémentation et évaluation d'un modèle d'attention pour la vision adaptative, Traitement du Signal vol 28, 6 (2012) p611-641
- M. Ferreira Da Silva, V. Courboulay "Implementation and evaluation of a computational model of attention for computer vision" In Developing and Applying Biologically-Inspired Vision Systems : Interdisciplinary Concepts. IGI Global (2012) pp 273-306
- Ferreira Da Silva, M., Courboulay, V., & Le Callet, P. (2012). Real Time Dynamic Image Re-targeting Based on a Dynamic Visual Attention Model. In 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (pp. pp.653–658).
- Courboulay, V., Ferreira Da Silva, M., & Courboulay, V. (2012). Real-time computational attention model for dynamic scenes analysis : from implementation to evaluation. In SPIE (Ed.), SPIE Optics, Photonics and Digital Technologies for Multimedia Applications - Visual Attention
- M. Ferreira Da Silva, V. Courboulay, P. Estrailier, «Image complexity mesure based on visual attention », ICIP 2011.
- M. Ferreira Da Silva, V. Courboulay, « Une nouvelle mesure de complexité pour les images basée sur l'attention visuelle », GRETSI 2011.

- M. Perreira Da Silva, V. Courboulay, P. Estrailier, « Objective validation of a dynamical and plausible computational model of attention », EUVIP 2011.
- M. Perreira Da Silva, V. Courboulay, A. Prigent, P. Estrailier, « Evaluation of preys / predators systems for visual attention simulation », VISAPP 2010.

Chapitre 3

Vers la recherche attentionnelle d'information

3.1 Positionnement de mes travaux

Dans ce chapitre, je présente les travaux que j'ai réalisés et encadré concernant la recherche attentionnelle d'information.

	Vision passive	Vision réactive	Vision active	Interaction <u>pregance</u>
Sans <i>a priori</i>	Restauration d'images isotrope & EPI [2002]			
Avec <i>a priori</i>	Restauration anisotrope d'images et EPI [2004]	Modélisation de la dynamique attentionnelle [2006-...]	VERS LA RECHERCHE D'INFORMATION [2009-...]	
Avec bouclage				
Intégration de l'utilisateur				

Jusqu'en 2009, mes travaux concernaient *la recherche de l'information*¹. J'ai depuis orienté mes recherches vers la *recherche attentionnelle d'information*, autrement dit *la conception d'un système d'attention visuelle pour l'analyse contextualisée de la scène*.

1. Ensemble des méthodes ayant pour objet d'extraire d'une image les informations pertinentes.

Mon objectif était et est encore de travailler sur des systèmes qui permettent une analyse plus *efficente* et plus *bio-inspirée* de l'observation. Il s'agit de se focaliser sur des zones d'intérêt de l'image pour permettre une analyse contextualisée en fonction d'une représentation de connaissances que le système intégrerait. En effet, l'analyse d'images en vue de la reconnaissance s'appuie pour l'heure, principalement sur l'analyse exhaustive de son contenu au détriment d'une compréhension de plus haut niveau. Si l'on est capable de traiter activement des données, l'ordinateur doit pouvoir se focaliser sur les informations pertinentes quelles que soient leur échelle ou leur importance dans l'image.

J'ai proposé une recherche plus proactive de l'information en se focalisant sur les zones d'intérêt et d'y déployer, dans un premier temps, des stratégies classiques de recherche d'information². La présentation de ces travaux constitue la première partie de ce chapitre.

L'objectif de la section 3.3 est de présenter deux exemples relatifs à cette *conception d'un système d'attention visuelle pour l'analyse contextualisée*. Le premier présente l'analyse de *l'observé*, au travers de l'extension des travaux précédemment introduits. Le second s'intéresse quant à lui à *l'observateur* et propose une analyse contextualisée de ses gestes dans le cadre d'une application interactive.

3.2 Recherche attentionnelle de l'information

J'ai indiqué en introduction de ce manuscrit la nécessité de considérer les grandes masses de données émergeant du web comme une des causes principales du changement de paradigme nécessaire en traitement de l'information.

Pour rappel, selon la revue *Time*, plus de 130 000 photos sont téléchargés chaque minute sur Facebook. La recherche d'information dans un si grand nombre d'images reste encore un vrai challenge. En contrepoint de ces *big data* modernes, Descartes s'imposait trois préceptes dans son célèbre *Discours de la Méthode* [Descartes 00], en 1637 :

1. Diviser chacune des difficultés afin de mieux les examiner et les résoudre.
2. Établir un ordre de pensées, en commençant par les objets les plus simples jusqu'aux plus complexes et divers, et ainsi de les retenir toutes et en ordre.
3. Passer toutes les choses en revue afin de ne rien omettre.

Malgré les presque 400 ans d'écart entre ces deux idées, j'ai été frappé par la similarité de ces deux visions. Les préceptes de l'une sont facilement transposables à l'autre.

2. Ensemble des méthodes permettant, en fonction de critères de recherche propres à l'utilisateur, de sélectionner l'information dans un ou plusieurs fonds structurés, voir chapitre 1.

Le challenge consiste à travailler sur toute l'image (3) en proposant une segmentation facilitant la résolution de la recherche de l'information (1) tout en établissant un ordonnancement basé sur la saillance des régions extraites (2).

De nombreux travaux ont été réalisés dans le domaine de catégorisation d'images [Thériault 13]. La plupart de ceux-ci se base sur une extraction de descripteurs bas niveau sur toute l'image. Cependant, il n'y a pas de relation simple entre ces descripteurs et une représentation plus sémantique. L'absence de ce type de relation est nommée «*fossé sémantique*».

Ces détecteurs de points d'intérêt, même s'ils ont prouvé leur efficacité, ne permettent pas forcément de sélectionner les régions les plus informatives reliées à la perception humaine [Dave 12]. De plus, ces détecteurs extraient des milliers de points par image, dès lors en plus de la grande masse des données, se rajoutent des problèmes de mémoire et de temps de calculs.

Je présente dans la suite, les travaux relatifs à ces problématiques que nous avons menés durant la thèse de Dounia Awad «*Vers un système perceptuel de reconnaissance d'objets*» et le stage de fin d'étude de Jeudi Hao «*Développement d'un outil de classification de documents administratifs par sacs de mots visuels*» autour de la recherche attentionnelle d'information.

Je présente d'abord une utilisation de système d'attention visuelle comme filtrage pour réduire le nombre des points d'intérêts dans un cadre CBIR classique, puis une proposition de *CBIR attentionnel*.

3.2.1 Filtrage attentionnel

De nombreux challenges ont été proposé pour tester l'efficacité et la robustesse des méthodes de reconnaissance d'objet. Le challenge le plus connu est le challenge VOC pour *Visual Object Classes Challenge*. VOC a été proposé pour la première fois en 2005 avec pour seul objectif : «*reconnaitre des objets à partir d'un certain nombre des classes d'objets dans des images et des scènes naturelles*». En 2005, douze équipes ont participé à l'évaluation de leurs systèmes sur les bases de challenge³.

L'algorithme vainqueur de Zhang consistait à extraire des représentations invariantes à partir des images et les utiliser comme base pour classifier l'image avec un classifieur SVM non linéaire. L'algorithme de Zhang peut être divisé en 3 parties :

1. Représentation de l'image : cette partie consiste à extraire un ensemble des descripteurs SIFT $K_{Zhang}(I)$ à partir d'une image $I(x, y)$ utilisé comme entrée du système. Elle est formée des 2 étapes :
 - Détecteur des points d'intérêt : Zhang utilise deux détecteurs invariants à l'échelle complémentaires pour extraire les structures d'image les plus

3. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2005/>

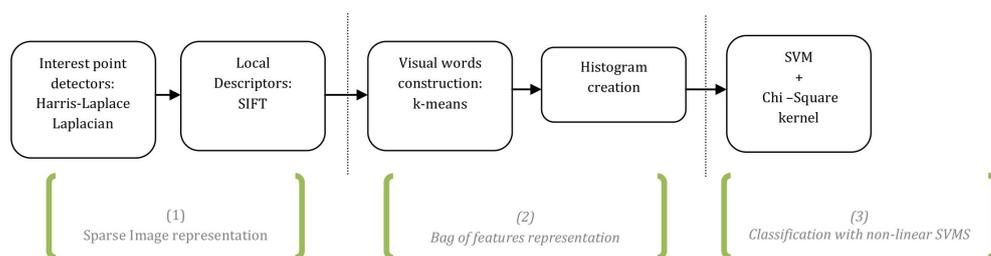


FIGURE 3.2.1 : Architecture de l'algorithme Zhang

pertinentes : le détecteur Harris-Laplace conçu pour détecter les contours, et le détecteur Laplacien destiné à détecter les blobs.

- Descripteur local : pour décrire visuellement les régions extraites, Zhang a utilisé le descripteur de SIFT (peu sensibles aux variations d'échelle et invariants aux changements d'illuminations).
2. Représentation par sac des mots visuels : Zhang calcule des vocabulaires visuels en quantifiant les descripteurs de l'ensemble d'apprentissage par une des méthodes de classification non supervisé (k-means). Ensuite, Chaque image est représentée par un histogramme des mots visuels déduits des vocabulaires visuels. Dans ce contexte, Zhang a extrait aléatoirement pour chaque classe un ensemble des descripteurs parmi ceux qui sont dédié à l'apprentissage. Ainsi, chaque descripteur dans une image est attribué au plus proche mot visuel, et un histogramme qui mesure la fréquence de chaque mot visuel est calculé pour chaque image.
 3. Classification : Zhang utilise un SVM non linéaire dans lequel une fonction de décision est estimée par l'équation suivante :

$$g(x) = \sum \alpha_i y_i k(x_i, x) - b$$

, où $k(x_i, x)$ est la valeur du fonction de noyau (khi-2) qui prend comme paramètre : x_i un élément de l'ensemble de l'apprentissage et x un élément de l'ensemble de test.

α_i est le coefficient de la pondération calculé lors de l'apprentissage de x_i .

b est le seuil déjà calculé.

En analysant les différentes étapes de cet algorithme, on observe que la première étape consiste à utiliser les détecteurs des points d'intérêt purement géométriques. Ces méthodes extraient un grand nombre des points d'intérêt, loin d'être tous utiles à la catégorisation des images.

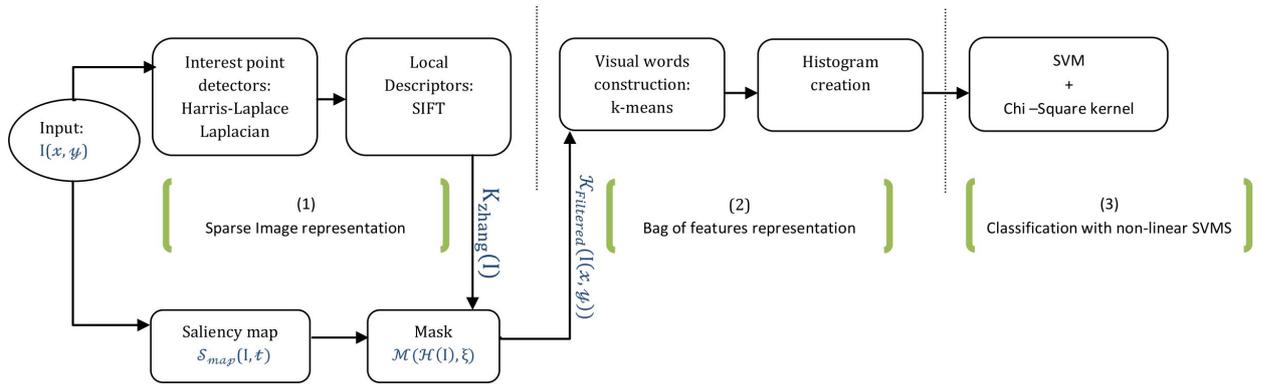


FIGURE 3.2.2 : Architecture de notre modèle

J'ai donc proposé d'utiliser notre système d'attention visuelle comme un filtre afin de sélectionner les points les plus saillants [Awad 12].

L'algorithme est le suivant. Pour une image $I(x, y)$, un ensemble des points d'intérêt $K_{Zhang}(I)$ est calculé ainsi qu'une carte de saillance $S(I, t)$. On peut noter que le calcul de carte de saillance peut se faire sur une image réduite dans un temps négligeable.

A partir de $S(I, t)$, on génère un masque $\mathcal{M}(\mathcal{H}(I), \xi)$ servant à filtrer les points d'intérêt selon un niveau minimal de saillance à considérer dans les images ξ . En général, ce masque peut être définir par :

$$\mathcal{M}(\mathcal{H}(I), \xi) = \begin{cases} 1 & \text{if } \mathcal{H}(x_h, y_h) > \xi \\ 0 & \text{otherwise} \end{cases} \quad (3.2.1)$$

Le processus de filtrage consiste à sélectionner l'ensemble $\mathcal{K}_{Filtered}(I)$ des points d'intérêt $\mathcal{K}_{Zhang}(I)$ pour lequel le masque $\mathcal{M}(\mathcal{H}(I), \xi)$ est égal à :

$$\mathcal{K}_{Filtered}(I(x, y)) = \{Key_j \in \mathcal{K}_{Zhang}(I(x_h, y_h)) \mid \mathcal{M}(\mathcal{H}(I), \xi) = 1\} \quad (3.2.2)$$

L'ensemble calculé $\mathcal{K}_{Filtered}(I)$ sera utilisé après comme une entrée pour les différents parties de l'algorithme de Zhang.

Nous avons mené une campagne de tests sur les bases VOC 2005 et VOC 2007. Les résultats, présentés dans [Awad 12], montrent que l'on peut maintenir la même performance en sélectionnant seulement 40% des points d'intérêts déjà calculé.

Ces résultats nous prouvent la pertinence de prendre en considération des informations attentionnelles dans le contexte de recherche d'informations dans de grandes bases de données.

L'étape suivante que nous avons explorée, réside dans une approche encore plus attentionnelle en remplaçant les descripteurs géométriques par des descripteurs attentionnels.

C'est l'objet de la section suivante.

3.2.2 « Attentive CBIR »

Le principal problème de la démarche présentée préalablement est qu'elle fait cohabiter des détecteurs et des descripteurs de natures différentes. Les premiers sont basés sur l'attention alors que les seconds sont basés sur des informations géométriques. Il a donc été naturel de se demander si nous pouvions proposer une chaîne complète basée sur l'attention. Notre démarche n'était pas d'obtenir de meilleurs résultats (nous n'avions cependant rien contre) mais d'avoir une approche cohérente du CBIR basée sur l'attention.

L'approche que nous avons mise au point, a été validée dans le cadre d'un article présenté lors d'ICPR 2014 à Stockholm [Awad 14].

Dans la suite de ce chapitre, nous présentons brièvement cette approche de description perceptuelle et sa validation. Elle consiste à décrire l'aspect fréquentiel de points d'intérêt de l'images et de ses composantes perceptuelles comme la couleur, l'intensité et l'orientation...

Cette proposition est motivée par des études psychophysiques qui ont montré que le cerveau humain fait un analyse fréquentielle de l'image [Campbell 68] postérieure à la détection de caractéristiques bas-niveau au niveau de la rétine. Ce descripteur présente l'avantage de fournir des vecteurs de dimension deux fois plus petit que ceux proposés dans l'état de l'art.

Dans notre approche, nous nous basons sur la méthode d'Unser [Unser 86], pour caractériser les régions d'intérêt dans des images naturelles. Unser a proposé une méthode computationnelle efficace et robuste pour calculer le vecteur de caractéristiques d'une région d'intérêt. Elle consiste en l'emploi d'opérateurs linéaires de filtrage dans un voisinage restreint pour extraire l'information de texture sous la forme de mesures statistiques compactes. Pour chaque point d'intérêt $x_{k,l}$, on calcul $y_{k,l}$ comme :

$$y_{k,l} = \sum_{l,k=0,..M-1} T_M \cdot x_{k,l} \quad (3.2.3)$$

où T_M représente la matrice de transformation $M \times M$ choisie.

Habituellement, on choisit la transformée de Fourier ; dans sa thèse, D.Awad a cependant décidé d'utiliser une transformée proche, réelle et computationnellement intéressante, la transformée de Hartley $H_{k,l}$ [Millane 94]. La matrice de

transformation de Hartley de taille $M \times M$ est calculée comme suit :

$$H_{k,l} = \sum_{l,k=0}^{M-1} \left(\cos \left(\frac{2\pi lk}{M} \right) + \sin \left(\frac{2\pi lk}{M} \right) \right) \quad (3.2.4)$$

où $k, l \in \{0, \dots, M - 1\}$.

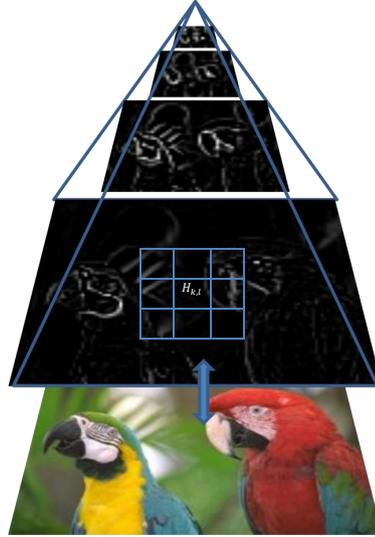


FIGURE 3.2.3 : Illustration de l'utilisation du filtrage par la matrice de transformation de Hartley

Cette transformée linéaire, séparable et orthogonale a l'avantage de posséder une excellent propriété de regroupement de l'énergie pour les régions des basses fréquences. Cette propriété permet de construire un vecteur de caractéristiques de dimension réduite par rapport aux descripteurs classiquement utilisés dans les méthodes de CBIR.

Pour résoudre le problème du voisinage, Unser a proposé une alternative intéressante. Cette alternative consiste à effectuer des analyses à des niveaux de résolution différents dont certains seront plus adaptés que d'autres à la structure analysée. Pour cette raison, nous avons décidé appliquer la méthode d'Unser sur les pyramides multi-résolutions des caractéristiques calculées par notre système d'attention visuelle présenté dans le chapitre précédent. Chaque point d'intérêt détecté est positionnée dans l'espace multi-résolutions de notre système attentionnel, et pour chaque pyramide le même voisinage de taille $M \times M$ est choisi $w_{x_{k,l}}$ et l'on y applique la transformation de Hartley.

Évaluation de notre méthode contre les résultats référence (SIFT)

Nous avons fait nos tests sur la base VOC2007 qui comprend plus de 10000 images annotées et réparties en 20 classes. Un classifieur *un-contre-tous* est appris pour chaque classe et la performance est évaluée par sa précision moyenne. Nous avons étudié la performance de chaque descripteur SIFT ou perceptuel sur plusieurs détecteurs (grille dense, Harris-Laplace et Laplacien, Système attentionnel), le reste de la chaîne restant identique.

Les résultats sont présentés ci-dessous :

Region selection methods	descriptor	aeroplane	bike	bird	boat	bottle	bus	car	cat	chair	cow
Dense grid	SIFT	69,58	56,28	40,63	64,17	24,56	60,63	75,06	56,17	49,19	38,90
Dense grid	our descriptor	68,02	44,96	36,90	58,99	21,98	46,57	63,84	43,67	41,53	27,07
Harris-Laplace, Laplacian	SIFT	61,45	48,12	37,76	51,27	21,34	47,30	65,60	46,16	41,45	31,62
Harris-Laplace, Laplacian	our descriptor	59,90	36,80	34,03	46,09	18,76	33,24	54,38	33,66	33,79	19,79
Perreira's system	SIFT	19,17	10,51	17,34	7,51	6,15	9,35	26,55	21,14	13,12	13,33
Perreira's system	our descriptor	44,72	22,28	21,68	14,23	10,99	18,66	44,41	25,62	19,59	8,37

Region selection methods	descriptor	dining table	dog	horse	motorbike	person	potted plant	sheep	sofa	train	tvmonitor
Dense grid	SIFT	50,54	36,79	75,69	63,59	81,51	26,55	45,21	46,78	74,36	50,10
Dense grid	our descriptor	51,37	34,26	70,13	54,41	81,29	35,08	46,44	47,93	68,23	45,91
Harris-Laplace, Laplacian	SIFT	35,85	28,37	67,62	50,95	73,06	15,21	29,74	22,51	62,95	37,12
Harris-Laplace, Laplacian	our descriptor	36,68	25,84	62,07	41,77	72,83	23,74	30,97	23,66	56,81	32,93
Perreira's system	SIFT	5,46	21,65	21,64	18,90	50,78	6,22	3,72	7,41	10,28	10,28
Perreira's system	our descriptor	12,96	24,96	36,38	23,72	63,32	20,03	11,03	12,94	32,72	21,36

FIGURE 3.2.4 : Impact du couplage détecteurs/descripteurs sur le résultat sur la tâche de reconnaissance sur la base VOC2007

Bien que la dimension de notre descripteur soit réduite de moitié, faisant ainsi gagner en temps d'exécution et d'occupation mémoire, la performance globale

de reconnaissance ne baisse seulement que de 5% en moyenne. Pour le détecteur basé sur notre système d'attention, le système basé sur notre détecteur perceptuel montre une progression de ses résultats de l'ordre de 9% par rapport à l'utilisation d'un descripteur SIFT.

Ces résultats valident notre hypothèse concernant la pertinence de coupler de façon cohérente détecteurs et descripteurs de points saillants.

Nous pouvons ainsi conclure que lorsque ce couplage est fait nous lions *saillance* et *intérêt*.

3.3 Applications en recherche d'information

Après un travail plus fondamental sur une recherche attentionnelle d'information, je présente dans cette partie deux exemples plus appliqués des recherches que j'ai menées en recherche attentionnelle d'information.

Dans le premier exemple, je présente l'annotation automatique de documents administratifs en se basant sur une recherche d'information sur les zones spatiales saillantes. Ces travaux sont menés dans le cadre d'une collaboration avec une entreprise reconnue mondialement en tant qu'éditeur de solutions de capture de documents.

Dans le second exemple, je présente des travaux que je co-encadre avec des chercheurs du laboratoire MICA d'Hanoï concernant la détection de gestes pour l'interaction homme-robot.

3.3.1 Annotation de documents basée saillance

Le contexte de ce premier travail est la valorisation de documents administratifs numérisés. Le L3i dispose d'une solide expérience dans le domaine de l'indexation d'informations et l'analyse de documents. Il est porteur ou acteur de nombreux projets nationaux et européens. À l'heure actuelle, il contribue au projet de PIXL, labellisé «Investissement d'Avenir» et est impliqué dans de nombreuses collaborations avec des entreprises. Le L3i est le leader d'une technopole liée à la numérisation de documents Valconum⁴, regroupant plusieurs sociétés (IBM, Docapost, ITESOFT, VEDANA, Sood ...). C'est dans le cadre d'un de ces partenariats que les travaux que je présente ici, ont été initiés en 2012. Ils sont réalisés en collaboration avec Benjamin Duthil et Mickael Coustaty, deux ingénieurs de recherche. Le cadre d'étude portait sur l'enrichissement sémantique de documents numérisés.

Cette thématique est d'ailleurs devenue le cadre d'un workshop à part entière dans le cadre de RFIA : l'Atelier sur l'Intégration de connaissances au sein des

4. <http://www.valconum.fr/>

processus d'analyse d'images (ICOPA)⁵. Il est traité dans cet atelier le problème de l'annotation d'images.

Mais revenons en à la problématique des documents. La numérisation de documents administratifs est un enjeu économique et écologique prioritaire dans le contexte sociétal actuel. La dématérialisation du document papier devient une tâche systématique dans les phases d'échange et de traçabilité de l'information. La dématérialisation massive de document n'est pas sans conséquence et soulève les problèmes d'organisation, de stockage et d'accès à l'information. Organiser, stocker et annoter ces documents devient une priorité pour les organisations. Le défi n'est donc plus la numérisation du document, mais l'extraction des informations qu'ils contiennent.

A la différence des images naturelles que j'ai évoquées jusqu'à maintenant, les documents administratifs sont produits par l'Homme et pour l'Homme. Cette propriété permet de relier informations utiles et zones saillantes (un total en rouge, un label fluo, un tampon en surimpression...). La saillance et la reconnaissance sont deux éléments essentiels pour la classification rapide de documents.

J'ai donc proposé une nouvelle approche d'annotation automatique de documents administratifs qui utilise une double saillance, une saillance visuelle et une saillance lexicale [Coustaty 13, Duthil 13b].

Cette section présente ces premiers travaux d'annotation automatique.

J'ai proposé d'utiliser le logo contenu dans les documents comme élément d'apprentissage. Le logo est en effet un élément graphique riche de sens [Duthil 13a] auquel il est possible de rattacher de multiples informations comme par exemple le nom de l'entreprise, son secteur d'activité ou encore sa géolocalisation. L'utilisation du logo comme élément d'annotation est une méthode graphique robuste qui permet par exemple de classer rapidement des documents.

L'approche d'annotation de documents administratifs que nous avons proposée est composée de quatre étapes présentées figure 3.3.1.

- La première étape consiste à extraire automatiquement les zones saillantes (vignettes) du document afin d'extraire et d'identifier le(s) logo(s) contenu(s) dans le document en utilisant l'approche présentée au chapitre précédent.
- La seconde étape permet la construction automatique d'un corpus d'apprentissage de documents web à partir des vignettes saillantes identifiées.
- La troisième étape sert d'apprentissage du vocabulaire (appelé descripteur) relatif au logo. Cette étape utilise le corpus d'apprentissage construit lors de l'étape précédente.
- La dernière étape correspond à l'annotation du document par le lexique de descripteurs appris lors de l'étape précédente.

La première étape consiste en l'application du modèle de saillance présenté au

5. <http://w3.mi.parisdescartes.fr/sip-lab/conferences/icopa2014/>

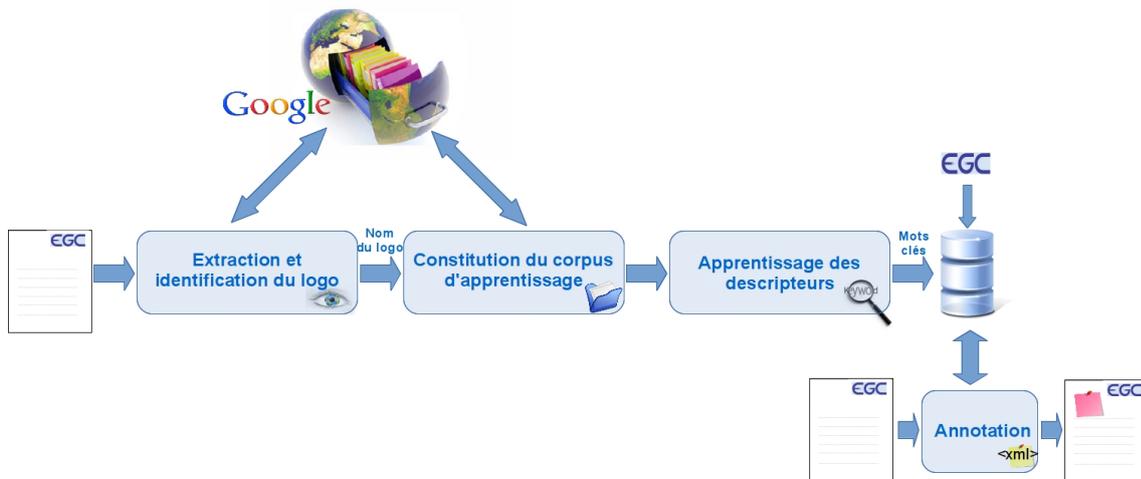


FIGURE 3.3.1 : Schéma global de l'approche

chapitre précédent. Les paramètres du modèles sont restés identiques tout au long du test, nous ne pouvions favoriser la couleur, la texture ou l'intensité, chacune de ces informations pouvant permettre de mettre en avant un logo. À partir de toutes les vignettes identifiées par le processus de saillance visuelle, un processus d'identification des logos est effectué. Chaque vignette fait l'objet d'une requête sous forme d'image sur un moteur de recherche web (*Google Images*) pour identifier le nom du logo. Si la vignette n'est pas reconnue comme un logo, elle ne sera plus considérée dans la suite du processus (fouille de texte), sinon, le nom du logo est conservé et il sera ensuite utilisé lors de la phase d'apprentissage des descripteurs sous forme de "mot germe".

Le choix du moteur de recherche Google est dû à différents constats :

- le web est une riche source d'information en perpétuelle évolution (apparition de nouveaux logos, etc),
- la diversité des documents disponibles (sites spécialisés, blogs, etc.) nous permet d'apprendre un large panel de descripteurs proches de la vignette requête (registres de langue différents, expressions, etc.)
- nous souhaitons montrer la pertinence du système avant de développer une recherche d'information plus spécialisée sur les logos.

Le processus de fouille de texte est composé des étapes 2 et 3. Afin d'être cohérent avec le positionnement de mes recherches, je souhaitais que notre système soit le plus autonome possible (expertise minimale). L'approche *Synopsis* proposée par Duthil et al. [Duthil 11, Duthil 12a] entre dans ce cadre. D'une part, *Synopsis* nous permet de construire automatiquement un corpus d'apprentissage à partir de "mots germes", et d'autre part, l'approche nous permet un apprentissage automatique des descripteurs. Dans notre cas, et uniquement pour la phase de construction

automatique du corpus d'apprentissage, "les mots germes" correspondent en réalité aux vignettes saillantes identifiées lors de l'étape 1, c'est pourquoi on parlera de "*germes visuels*". Ces "germes visuels" sont utilisés lors de la phase de construction du corpus d'apprentissage afin de rechercher des documents web contenant le germe visuel considéré (requête par l'image).

Concrètement, les mots germes sont les noms potentiels (noms de logo) des différentes vignettes qui ont été identifiées lors de l'étape 1. Le rôle des mots germes est de guider l'apprentissage des descripteurs textuels : lorsqu'un mot germe est rencontré dans un document, le système apprend les mots qui sont proches, au sens sémantique du terme, du mot germe. Autrement dit, lorsque l'on parle du mot germe, on parle systématiquement d'autres mots sémantiquement proches. Par exemple, lorsque l'on parle de voiture, on pourrait également parler de cylindrée ou de puissance DIN.

Je n'aborderai pas ici le détail des travaux de text-mining mis en œuvre dans cet algorithme grâce à Benjamin Duthil. En résumé, l'approche *Synopsis* est principalement basée sur deux éléments clés : la notion de fenêtre et la notion de classe/anti-classe. La fenêtre permet d'effectuer un apprentissage des descripteurs tout en assurant leur cohérence sémantique avec le mot germe [Duthil 12a]. La notion de classe/anti-classe permet de filtrer le bruit web et donc de gagner en précision sur le logo à caractériser. La figure 3.3.2 présente un exemple de résultats obtenus.

Logo	extrait du lexique	Logo	extrait du lexique	Logo	extrait du lexique
	AGPM Groupe AGPM contrat habitation assurance risque contrat assurance vie		EGC conférence extraction soumissions apprentissage connaissances		Banque Postale banque client services financement gestion conseiller

FIGURE 3.3.2 : Exemple de descripteurs

Nous avons évalué la qualité de l'apprentissage sur un corpus de 1766 documents administratifs dans un contexte de classification. Ce corpus est composé de 4 classes de documents : acte de mariage (A-M), certificat d'assurance (C-A), relevé d'identité bancaire (RIB) et certificat de naissance (C-N). Nous présentons la répartition de ces quatre classes (Nombre de logos différents dans la classe et nombre de documents de la classe). Cette base confidentielle provient d'un des leaders mondiaux de la dématérialisation documentaire. Chaque document contient un des 196 logos identifiés dans le corpus. Les documents sont scannés en 200 dpi

noir et blanc. Nous utilisons les indicateurs classiques de mesure pour évaluer la classification : *Précision*, *Rappel*.

La *précision* est calculée en considérant les erreurs d'identification du logo : le système identifie un logo qui n'est pas le bon, le lexique associé ne correspond donc pas au logo à identifier.

Le *rappel* est calculé en utilisant le nombre de logos correctement identifié par le système.

	nombre de logos	nombre de documents
acte de mariage	36	40
certificat de naissance	113	169
certificat de naissance	30	822
RIB	15	735

	C-A	A-M	C-N	RIB	Toutes classes
Rappel	95,5	22,5	60,4	71,7	80,6
Précision	100	100	100	100	100

FIGURE 3.3.3 : Répartition des classes et résultats de classification pour chaque classe : acte de mariage (A-M), certificat de naissance (C-N), certificat d'assurance (C-A), relevé d'identité bancaire (RIB)

Les résultats montrent la pertinence du système. Les différences de résultats entre chacune des classes s'expliquent par la qualité des documents. Par exemple, la qualité des "actes de mariage" est très mauvaise (cf. figure 3.3.4), ce qui explique le faible Rappel (22,5). Cependant, les résultats sont remarquables, nous obtenons un Rappel de 80,6 et une précision de 100 toutes classes confondues. La Précision (100) met en évidence la robustesse de l'approche. Ces travaux ont été publiés dans [Duthil 13b, Duthil 13a]. Nous avons essayé d'étendre ces résultats sur des images naturelles (cf. Figure 3.3.5).

Implémentation

Cet approche a donné lieu au développement d'un logiciel assez original puisque reposant sur une architecture à base de service web présenté dans [Duthil 14].

Nous présentons Figure 3.3.6 des captures de quelques cas d'utilisation.

3.3.2 Interprétation visuelle de gestes pour l'interaction homme machine

Comme je l'indique en introduction, ces travaux s'intègrent dans *la conception d'un système d'attention visuelle pour l'analyse contextualisée*. Le contexte peut

logo scanné	logo original
	
	
	

FIGURE 3.3.4 : Exemple de logos non-identifiés



Associated keywords
Senseo
Vittel
Blackberry

Exemple d'annotations générée

FIGURE 3.3.5 : Exemple d'annotations sur des images naturelles

3.3 Applications en recherche d'information

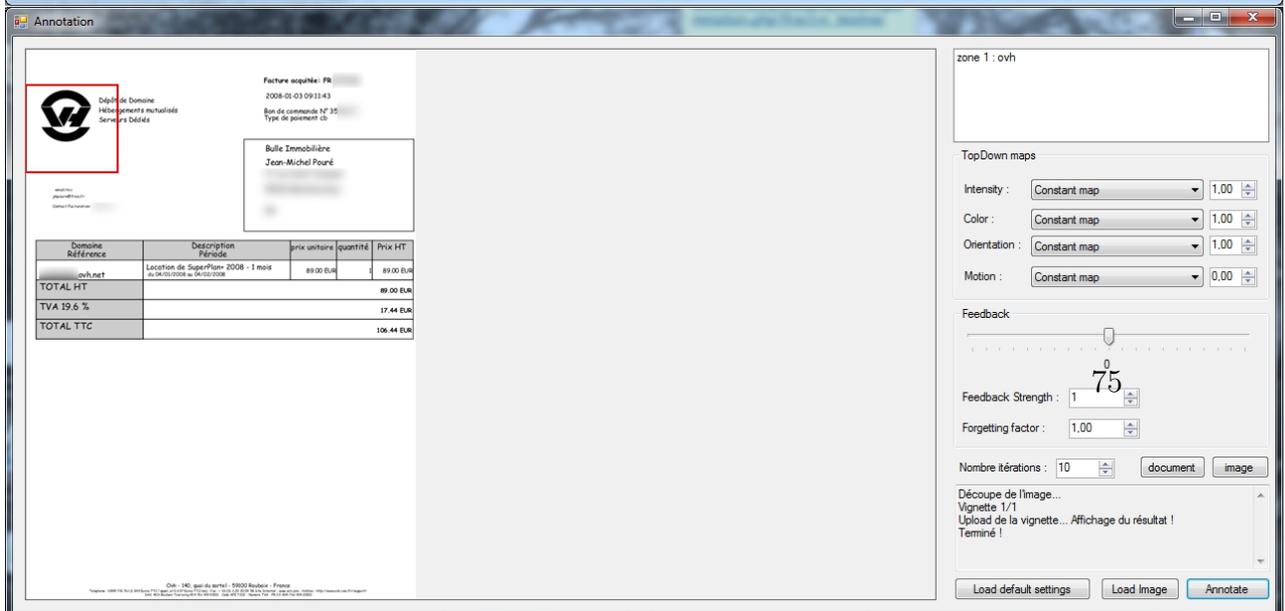
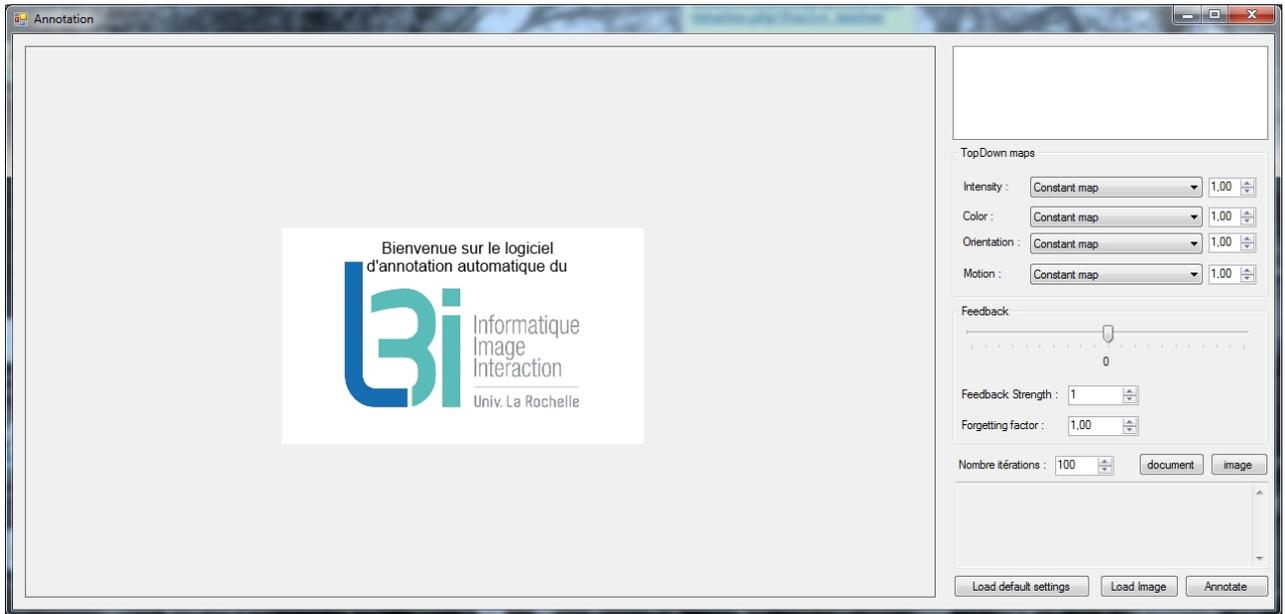


FIGURE 3.2.6 : Exemple de captures présentant notre logiciel d'annotation

bien évidemment être relatif à l'observé (*cf.* section précédente) mais également à l'utilisateur (*cf.* projet FT-AutiSTIC) comme nous allons le voir ici. Ce second exemple porte sur l'interprétation de gestes pour l'interaction. Avant de présenter ces travaux, je précise leur contexte. Le L3i collabore depuis de nombreuses années avec le Vietnam et l'Institut Francophone en Informatique et son laboratoire MSI (IRD), l'Institut Polytechnique d'Hanoï ainsi que le laboratoire MICA. L'institut MICA (Multimédia, Informations, Communications et Applications) est un institut de recherche franco-vietnamien sous tutelle de l'Institut Polytechnique de Hanoï, du CNRS et de Grenoble INP.

En relation avec R.Mullot et le laboratoire MICA, nous avons proposé une co-tutelle de thèse en 2010 autour de l'interprétation visuelle de geste pour l'interaction homme machine, thèse actuellement réalisée par Nguyen Van Toi.

Les gestes sont en effet un puissant outil de communication entre les êtres humains. La gestuelle est si profondément ancrée dans notre communication que les gens continuent souvent des gestes en parlant au téléphone. Les gestes constituent une modalité complémentaire distincte à la parole pour exprimer ses idées.

L'hypothèse de ces travaux consiste à considérer que les gestes sont un moyen intuitif et idéal d'échange d'informations avec d'autres personnes dans un espace virtuel, ou des robots (*cf.* figure 3.3.7).



FIGURE 3.3.7 : Intégration d'un robot capable d'interaction dans un musée (MICA)

Les premiers travaux relatifs à cette recherche concerne la reconnaissance de postures.

La démarche que nous avons proposé est une méthode de détection de la main sur la base de détecteur de Viola-Jones [Viola 02]. Nous avons introduit également deux nouveaux concepts que sont les *caractéristiques internes* et *caractéristiques pseudo-Haar (Haar-like features) internes*.

Définition 1 : *Caractéristique interne* : c'est une caractéristique extraite d'une région inscrite à l'objet d'intérêt sans arrière plan.

Définition 2 : *Caractéristique pseudo-Haar interne* : c'est une caractéristique pseudo-Haar extraite d'une région inscrite à l'objet d'intérêt sans arrière plan.

Nous avons utilisé un classifieur *AdaBoost* afin de classifier les poses de la main. Le résultat expérimental montre que cette méthode est prometteuse pour détecter la main en évitant la problématique de la prise en compte du fond [Nguyen 12].

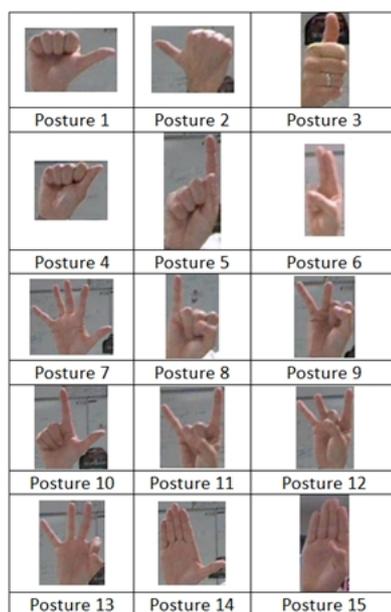


FIGURE 3.3.8 : Exemples d'un jeu de postures utilisé pour le détection de gestes.

L'objectif est maintenant d'inclure une détection préalable de la zone de la main basée saillance. Des chercheurs ayant déjà montré la faisabilité de cette approche [Ajallooeian 09], je souhaiterai appliquer ma méthode afin de compléter mes travaux dans le domaine de la robotique et d'inclure une carte haut-niveau dans le modèle de détection de saillance.

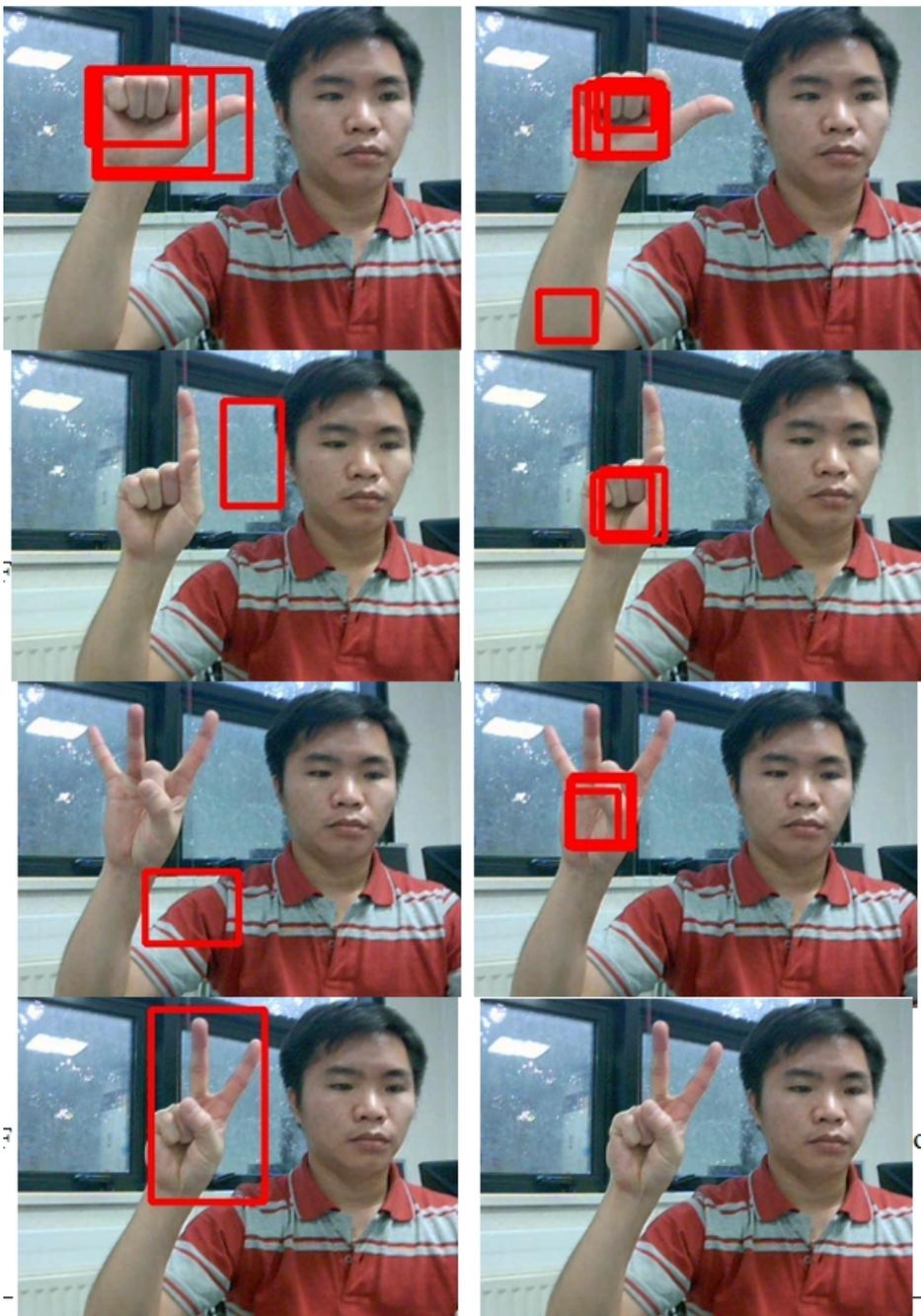


FIGURE 3.3.9 : A gauche, exemples de résultats avec un détecteur traditionnel type Viola-Jones ; à droite avec notre détecteur [ICCE2012]

3.4 Conclusion

Dans ce chapitre, j'ai présenté mes contributions autour de la problématique de la *recherche attentionnelle d'information* et de la *vision active avec a priori* de la diagonale informationnelle.

J'ai scindé ce chapitre en deux en présentant tout d'abord un apport concernant la recherche attentionnelle d'information puis des applications relevant de ce domaine.

Dans la première partie, j'ai proposé un filtrage perceptuel de points d'intérêts, puis une tentative de chaîne perceptuelle CBIR complète ainsi que son évaluation sur la base VOC2007. La conclusion principale de ces travaux réside dans la nécessité de faire des choix cohérents dans les détecteurs et descripteurs du CBIR.

Dans une seconde partie, j'ai exposé deux applications reposant sur une analyse attentionnelle :

- une annotation automatique d'images,
- une reconnaissance de gestes par un robot.

Les perspectives de ces travaux sont nombreuses. Nous pouvons, par exemple, aller plus loin qu'une annotation de documents. Dans le chapitre suivant, je présente un travail sur un enrichissement *sémantique* des documents.

Mais je souhaite me diriger vers un enrichissement conceptuel de documents. Je trouve utile de préciser ici la notion de *concept*.

Un concept est une représentation générale et abstraite de la réalité d'un objet, d'une situation ou d'un phénomène ; il n'est pas synonyme de notion car plus abstrait (par exemple, la notion de table, le concept de liberté). Concept vient du participe passé latin conceptus du verbe concipere, qui signifie « contenir entièrement », « former en soi ». Le concept se distingue donc aussi bien de la chose représentée par ce concept, que du mot, de la notion, ou de l'énoncé verbal, qui est le signifiant de ce concept⁶.

Dans ce cadre, j'ai obtenu le financement d'une thèse qui débute en 2014 sur ces thématiques.

Le lecteur attentif aura perçu l'absence de *bouclage* dans cette recherche attentionnelle d'information, malgré l'affichage du tableau en début de chapitre. Ce bouclage, proche d'un bouclage de pertinence, existe en filigrane du CBIR perceptuel, puisque basé sur un apprentissage. Néanmoins, ce bouclage fait bien évidemment partie de mes projets de recherche.

6. <http://fr.wikipedia.org/wiki/Concept>

Points clés

Positionnement

- ❑ Recherche concernant la case vision *proactive avec a priori* de la diagonale informationnelle
- ❑ Réflexions sur la recherche attentionnelle d'informations
- ❑ Développement d'applications et de démonstrateur

Contributions

- ❑ Validation du concept de recherche attentionnelle
- ❑ Application aux documents numérisés
- ❑ Des applications variées (annotation, robotique attentionnelle, art)

Publications

- Awad D., V.Courboulay, A.Revel “Attentive Content Based Image Retrieval” In From Human Attention to Computational Attention : A Multidisciplinary Approach Springer To be published
- Awad D., V.Courboulay, A.Revel. A new hybrid texture-perceptual descriptor :application CBIR, Paper accepted for the ICPR 2014.
- Duthil B. , V. Courboulay, Coustaty M. & Ogier j M. MAD : une plateforme mobile pour l'annotation de document vers la classification. In CORIA 2014.
- Duthil B., Coustaty M., Courboulay V., Ogier J.-M. Visual Saliency and Terminology Extraction for Document Annotation Dans 13th ACM Symposium on Document Engineering (DocEng), Italie (2013)
- Coustaty M., Duthil B., Courboulay V., Ogier J.-M. Adaptative Saliency for Logo Spotting Dans Home Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, États-Unis (2013)
- Duthil B., Coustaty M., Courboulay V., Ogier J.-M. Visual saliency and terminology extraction for document classification Dans Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, États-Unis (2013)
- Awad D., V.Courboulay, and A. Revel. Saliency Filtering of SIFT Detectors : Application to CBIR.In : ACIVS, Vol. 7517Springer (2012) , p. 290-300.
- NGuyen Van Toi, Le Thi Lan, Tran Thi Thanh Hai, Mullot Remy, Courboulay, Vincent A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier. The 4th International Conference on Communications and Electronics (ICCE 2012) Hue, Viet Nam - august 2012

Chapitre 4

Prégnance & Concept

4.1 Positionnement de mes travaux

Dans ce chapitre, je présente les travaux que j'ai réalisés et encadrés concernant ce que j'appelle, la recherche prégnante d'information.

	Vision passive	Vision réactive	Vision proactive	Interaction prégnance
Sans <i>a priori</i>	Restauration d'images isotrope & EPI [2002]			
Avec <i>a priori</i>	Restauration anisotrope d'images et EPI [2004]	Modélisation de la dynamique attentionnelle [2006-...]	VERS LA RECHERCHE D'INFORMATION [2009-...]	
Avec bouclage				Vers la recherche prégnante d'information [2012-...]
Intégration de l'utilisateur				

La définition de recherche prégnante d'information réside dans la nécessité de ne pas prendre en compte seulement les objectifs de l'utilisateur (ce que l'on aurait pu appeler recherche contextuelle d'information), mais également sa vie intérieure, ses émotions, ses connaissances. On voit donc rapidement l'extrême difficulté, voire l'impossibilité, à prévoir ses éléments, et tout l'intérêt d'accompagner l'utilisateur dans sa démarche pour petit à petit la co-construire avec lui. Cette *dernière* case

a été le cadre de plusieurs études. La première, que j'ai déjà présentée, a été la thèse de Nicolas Rempulski autour de la *synthèse dynamique de superviseur pour l'exécution adaptative interactive*. J'ai déjà présenté en partie ces travaux dans le premier chapitre de ce mémoire.

Il est à noter qu'il est très difficile d'avoir une approche holistique de ce domaine compte tenu de la *complexité* de l'approche. J'utilise volontairement le mot de *complexité* dans le sens qu'Edgar Morin lui a donné [Morin 05].

Le mot de complexité, lui, ne peut qu'exprimer notre embarras, notre confusion, notre incapacité à définir de façon simple, à nommer de façon claire, à ordonner nos idées. Sa définition première ne peut fournir aucune élucidation : est complexe ce qui ne peut se résumer en un maître mot, ce qui ne peut se ramener à une loi ni se réduire à une idée simple. La complexité est un mot problème et non un mot solution.

A la fin de ce chapitre, je présente quand même une tentative d'utilisation d'une approche plus formelle afin de résoudre un problème de recherche d'informations dans des documents anciens.

Malgré tout, ce chapitre sera principalement applicatif. Je présente ainsi, dans une première partie, l'intégration d'informations de plus haut niveau sémantique. La présentation se fait en suivant une approche incrémentale au sens de la complexité des informations mises en jeu. Je commence brièvement par les prémices de l'extension des travaux du chapitre précédent concernant l'enrichissement sémantique de documents, dans ce contexte la connaissance est issue du web.

Puis, je présente les travaux relatifs à l'émotion, domaine particulièrement délicat à aborder. Ne souhaitant pas développer ce travail seul, je me suis associé avec le laboratoire XLIM-SIC de Poitiers, dans le cadre de la thèse de Syntyche Gbehounou.

Enfin, je présenterai les travaux en cours de réalisation dans le cadre de l'ANR VISIIR. Dans les deux derniers cas, la connaissance est extraite d'une interaction avec l'utilisateur et entre donc complètement dans la dernière case de la diagonale informationnelle.

4.2 Vers la prise en compte d'informations de plus haut niveau...

4.2.1 issues du web

Pour rappel, le contexte de ces recherches est, non plus la numérisation de documents, mais l'extraction des informations qu'ils contiennent. Dans cette recherche initiée en 2003, l'objectif est de proposer une méthode applicable à la fois sur les

images en couleur et en noir et blanc, avec des documents de bonne qualité ou bruités, et capable d'intégrer un apprentissage incrémental. L'idée repose sur la combinaison de traitements d'image et de techniques de fouille de texte afin d'annoter un document par un ensemble de mots clés caractéristiques de la sémantique contenus dans le document.

Les premières étapes concernant l'extraction de zones saillantes et les requêtes web associées sont identiques à celles présentées précédemment. Le processus de fouille de texte utilise une construction automatique du corpus d'apprentissage et d'apprentissage automatique des descripteurs grâce à l'approche *Synopsis* proposée dans [Duthil 12b]. D'une part, *Synopsis* nous permet de construire automatiquement un corpus d'apprentissage à partir de "mots germes", et d'autre part, l'approche nous permet un apprentissage automatique des descripteurs.

Pour rappel, la construction du corpus d'apprentissage a pour objectif d'obtenir des documents web qui ont un contenu similaire à la vignette requête. À chaque vignette saillante est associé un corpus de documents. Plus formellement, à chaque vignette q , q variant de 1 à k , k étant le nombre de vignettes identifiées, un corpus Doc^q de n documents est associé tel que $Doc^q = doc_n^q, n = 1, \dots, n^q$.

L'objectif de l'apprentissage est de construire un lexique de descripteurs textuels (mots) décrivant sémantiquement chacune des vignettes saillantes. À chaque vignette est associé un lexique L_q .

Pour faire le lien entre les éléments graphiques contenus dans le document et les différents concepts auxquels ils font référence, l'approche *Synopsis* développée par B.Duthil est principalement basée sur deux éléments clés : la notion de *fenêtre* et la notion de *classe/anti-classe*. La fenêtre permet d'effectuer un apprentissage des descripteurs tout en assurant leur cohérence sémantique avec le mot germe. La notion de classe/anti-classe permet de filtrer le bruit web. Les noms communs et les noms propres sont les deux classes grammaticales apprises car elles sont reconnues comme porteuses de sens.

Plus formellement, une fenêtre de taille sz centrée sur un mot germe g pour un document doc est définie par $F(g, sz, doc) = \{m \in doc / d_{NC}^{doc}(g, m) \leq sz\}$ où $d_{NC}^{doc}(g, m)$ est la distance entre le mot germe g et le mot m .

Le principe général est de calculer la représentativité d'un mot M dans chacune des deux classes (fréquence d'apparition normalisée $\rho(M)$ dans la classe (mots présents dans les fenêtres) et dans l'anti-classe $\bar{\rho}(M)$ (mots en dehors des fenêtres). La représentativité d'un mot M dans chacune des classes est défini tel que :

$$\rho(M) = \sum_{doc} \sum_{\gamma \in \mathcal{O}(g, doc)} |\mathcal{O}(M, F(\gamma, sz, doc))| \quad \bar{\rho}(M) = \sum_{doc} |\mathcal{O}(M, \bigcap_{\gamma \in \mathcal{O}(g, doc)} \bar{F}(\gamma, sz, doc))|$$

À partir de la représentativité d'un descripteur dans chacune des classes, il devient possible de déterminer la proximité sémantique du descripteur M considéré en appliquant une formule de discrimination f tel que cela est proposé dans

[Duthil 12b]. Un score $Sc(M)$ est alors attribué à chaque descripteur. Chaque descripteur constitue une entrée du lexique L^q propre à la vignette q considérée. Le score d'un descripteur est calculé tel que : $Sc(M) = f^2(\rho(M), \bar{\rho}(M))$ où f est définie tel que : $f^2(x, y) = \frac{(x-y)^3}{(x+y)^2}$.

L'étape d'annotation sémantique consiste donc à rattacher à un document l'ensemble des concepts qu'il contient (lexique de mots précédemment construits).

Annoter sémantiquement un document au format image revient à rechercher et à identifier les éléments saillants qu'il contient afin de lui rattacher les lexiques associés (sémantique).

Annoter un document à partir de son contenu textuel (résultat d'OCR) consiste à identifier les concepts contenus dans le document. L'approche utilise une fenêtre glissante centrée sur les noms communs pour identifier les segments de textes pertinents. Cette méthode permet également de connaître l'intensité du discours [Duthil 12b]. À chaque document est associé un fichier xml qui contient un ensemble d'informations sémantiques : lexiques associés (identifiant du lexique correspondant au nom du logo), intensité du discours et l'importance du discours pour chacun des concepts (lexiques) qui ont été rattachés.

L'évaluation sémantique étant par définition difficile, nous avons réalisé les expérimentations sur une base d'images administratives, mais cette limitation m'a incité à creuser la notion d'évaluation subjective.

4.2.2 issues d'indices subjectifs

Et quoi de plus subjectif que l'émotion ressentie devant une image ?

C'est dans le cadre d'une assemblée générale MIREs¹ que j'ai pris connaissance des travaux de Syntyche Gbehounou autour de l'extraction de l'impact émotionnel des images à partir de descripteurs bas niveau.

Dans son exposé, elle présentait sa problématique et ses premiers résultats autour des images comme réels vecteurs émotionnels.

Mais essayer d'identifier l'émotion d'une image est une tâche ambitieuse. Différentes informations contenues dans une image (couleurs, contenu, textures, ...) peuvent être des vecteurs émotionnels, tout comme les *a priori* qui nous habitent tous. D'autres facteurs encore plus complexes que le contenu ou la couleur globale entrent en jeu dans l'interprétation émotionnelle d'une image. Il s'agit notamment de tous les facteurs culturels et personnels. Ces facteurs sont d'autant plus diffi-

1. La fédération MIREs - Mathématiques & leurs Interactions, Images & information numérique, Réseaux et Sécurité - est une fédération de laboratoires du PRES Limousin-Poitou-Charentes. Elle est née du rapprochement du Pôle de recherche PRIDES, regroupant des laboratoires de Poitou-Charentes, et de laboratoire de l'Université Limousin.

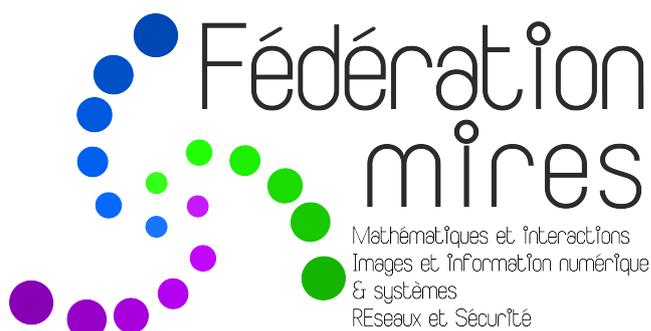


FIGURE 4.2.1 : Logo de la fédération MIRES

ciles à prendre en compte puisqu'ils ne peuvent être décrits par des descripteurs objectifs.

S.Gbehounou a également travaillé au développement d'une base d'images *SENSE* (Studies of Emotion on Natural image databaSE) annotée émotionnellement [Gbehounou 12].

Il existe en effet de nombreuses bases de données d'images pour les études des émotions ; la plus connue est le système *IAPS* de l'Université de Floride. En général, elles sont émotionnellement fortes (peu de nuances dans les émotions ressenties, je me refuse même à mettre des exemples dans ce manuscrit). Dans la base *SENSE*, les images ne choquent pas et ne forcent pas une forte réponse émotionnelle. Les images sont choisies afin de minimiser les interactions potentielles entre les émotions générées sur les images suivantes lors des évaluations subjectives. Enfin, et pour ne rien gâcher, la base *SENSE* est libre d'utilisation. Je présente Figure 4.2.2 quelques exemples d'images de la base *SENSE*.

Il s'agit d'un ensemble diversifié d'images qui contiennent paysages, animaux, nourriture et boisson, monuments touristiques...

La stratégie de test que les auteurs ont adoptée, est étroitement liée au modèle d'émotions qu'ils ont choisi. Ils ont travaillé avec un modèle d'émotion bidimensionnel :

- la nature de l'émotion,
- la puissance de l'émotion.

Cette modélisation d'émotion est équivalente à un modèle *valence/excitation* dans lequel, la valence permet de faire la distinction positive et négative et l'excitation qui varie de faible à élevé définit sa puissance.

La base *SENSE* a été constituée lors de deux séries de tests effectués sur le web² et séparément en temps. Au cours des différents tests, les observateurs ont évalué la nature et la puissance de l'émotion suscitée par l'image, comme le montre

2. <http://francois.lecellier.org/emotions/index.php>

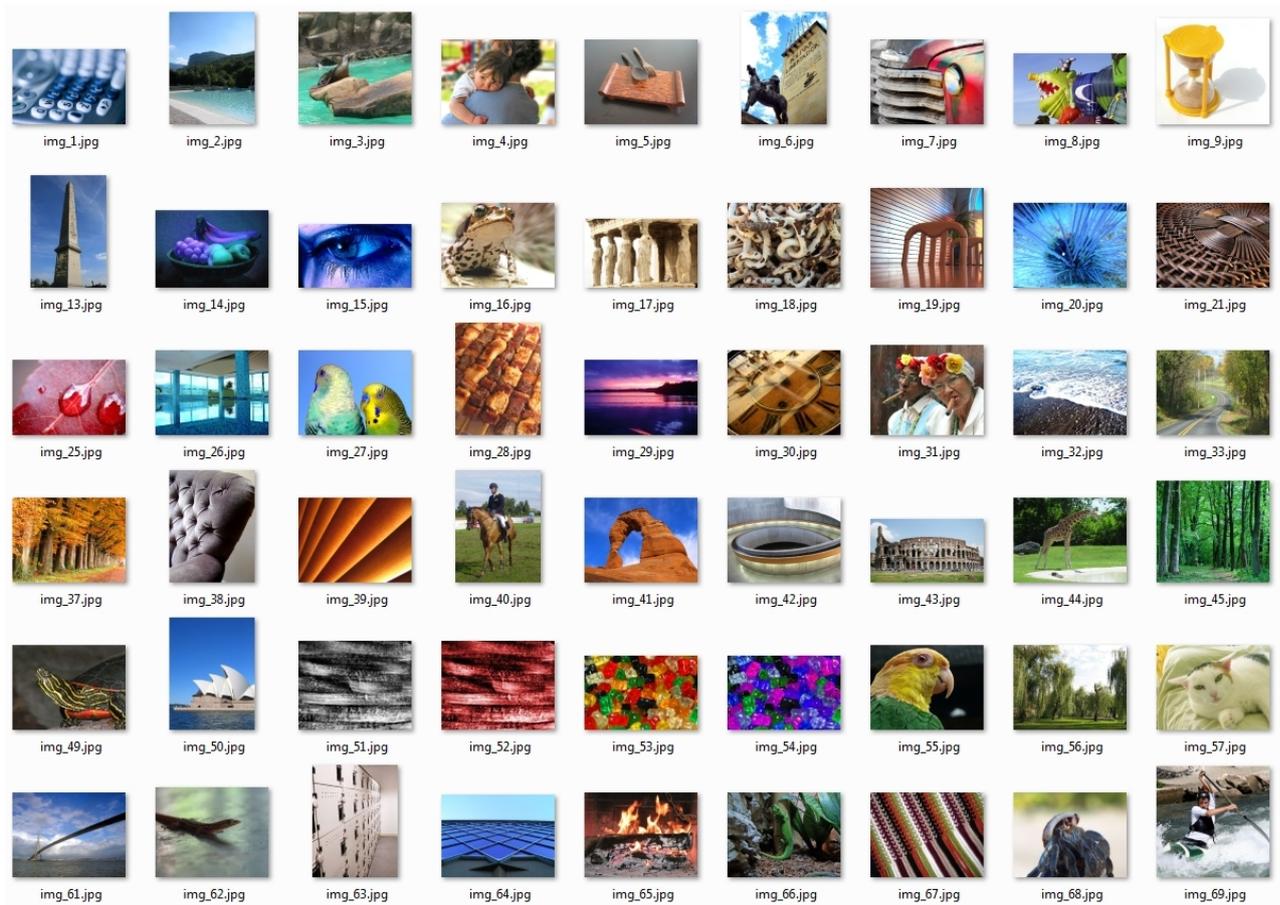


FIGURE 4.2.2 : Exemple d'images de la base SENSE.

la Figure 4.2.3, pour la nature de l'émotion, les participants ont eu choix entre "négatif", "neutre" et "positive" et pour la puissance de faible à élevée.

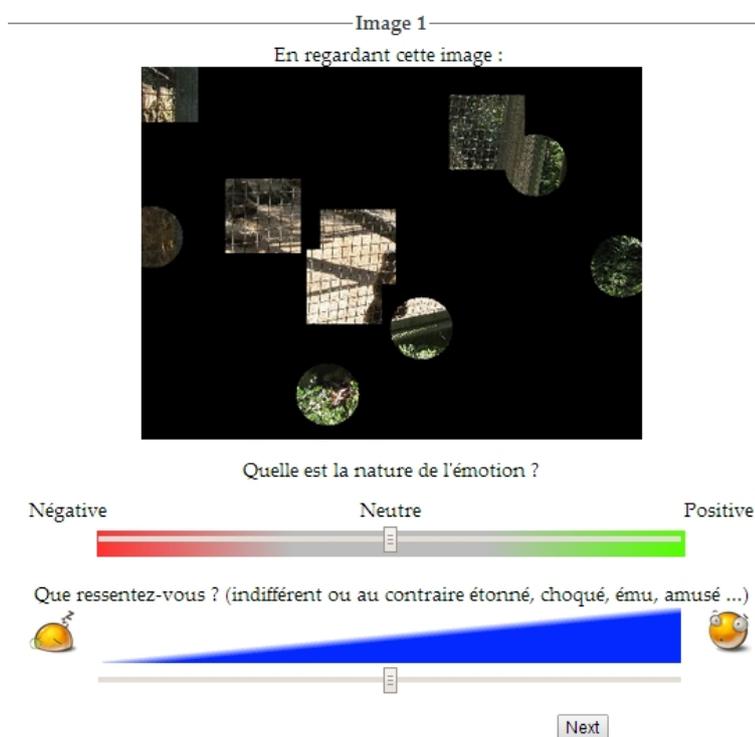


FIGURE 4.2.3 : Interface d'annotation émotionnelle.

La Figure 4.2.4 présente l'interface web indiquant les résultats obtenus.

Fort de cette base de données, et toujours dans l'optique d'extraire de l'information haut niveau issue d'indices subjectifs, j'ai voulu d'abord savoir si l'on percevait la même émotion lorsque l'on voyait l'image dans son ensemble et lorsque l'on ne voyait que la zone la plus saillante ?

En d'autres termes : pouvait-on localiser l'émotion ressentie dans la zone la plus saillante ?

La question est d'importance car j'ai montré l'intérêt et la faisabilité d'enrichir l'image en limitant son analyse aux régions saillantes ; si on pouvait aller jusqu'à une évaluation subjective, les applications n'en seraient que plus ouvertes.

Dans une étude que nous avons menée courant 2013, nous avons montré que la réponse aux questions précédentes étaient positives [Gbehounou 13]. Dans « *Can salient interest regions resume emotional impact of an image ?* », nous avons proposé de refaire une évaluation de la base SENSE non plus dans son intégralité, mais en ne faisant voir que la région la plus saillante.

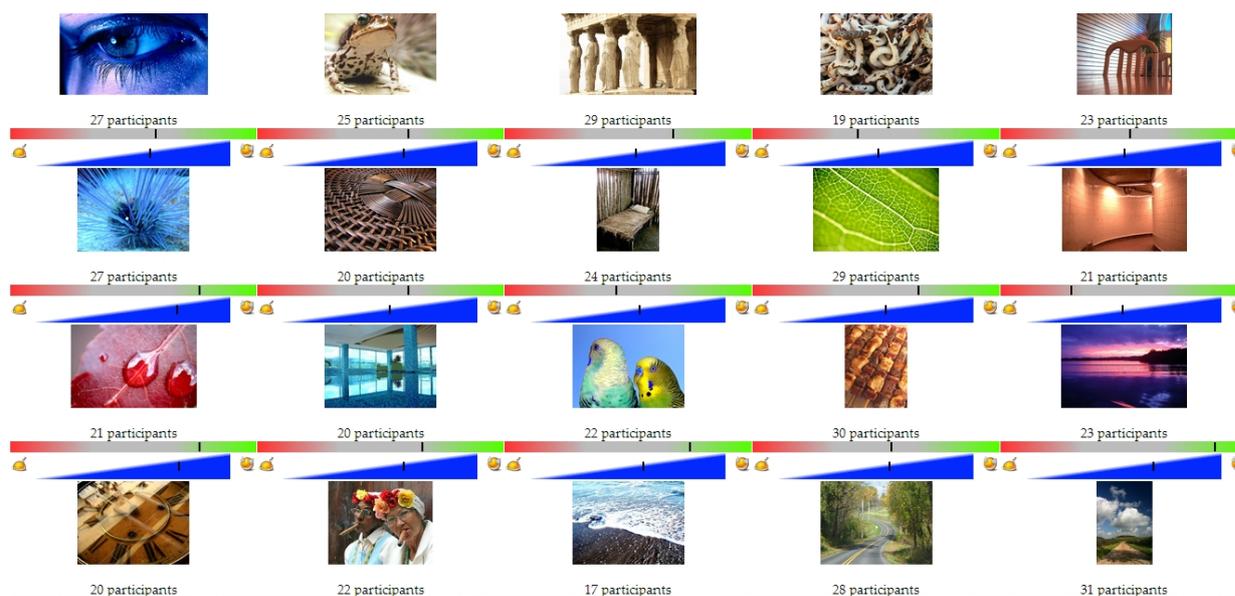


FIGURE 4.2.4 : Présentation web des résultats de l'évaluation

Afin d'éviter l'écueil d'un seuil fixe quant à la portion d'image affichée, nous avons laissé notre algorithme de saillance déterminer tout seul, la zone d'intérêt (voir 2.4.2).

Nous disposons ainsi de comparaisons entre l'émotion ressentie pour une image complète et pour la région la plus saillante ne couvrant que $p\%$ de l'image complète (cf. Figure 4.2.5).

Chaque image a été évaluée en moyenne plus de 100 fois.

La Figure montre le taux de corrélation entre l'évaluation des images réduites et des images complètes en fonction du taux d'affichage.

Une information beaucoup plus intéressante, fut le nombre d'images classées neutres lorsqu'on les voit complètement et désormais émotionnellement chargée lorsqu'on n'en voit qu'une partie.

Les résultats suggèrent ainsi que le modèle de l'attention visuelle bottom-up peut être très utile pour réduire la taille de l'image évaluée à des régions pertinentes afin de mieux permettre la reconnaissance des «émotions primaires». Nous envisageons de tester un autre modèle d'attention visuelle et de comparer les différents résultats avec ceux d'expérimentations subjectives avec un eye-tracker. Une autre perspective consiste en la description des régions saillantes par couleurs, textures, contours et positions caractéristiques afin de déterminer les liens qui existent entre indices bas niveau et émotions.

Je suis cependant conscient du caractère hautement complexe de ces études concernant les émotions, mais nous avons implémenté cette approche et nous pré-

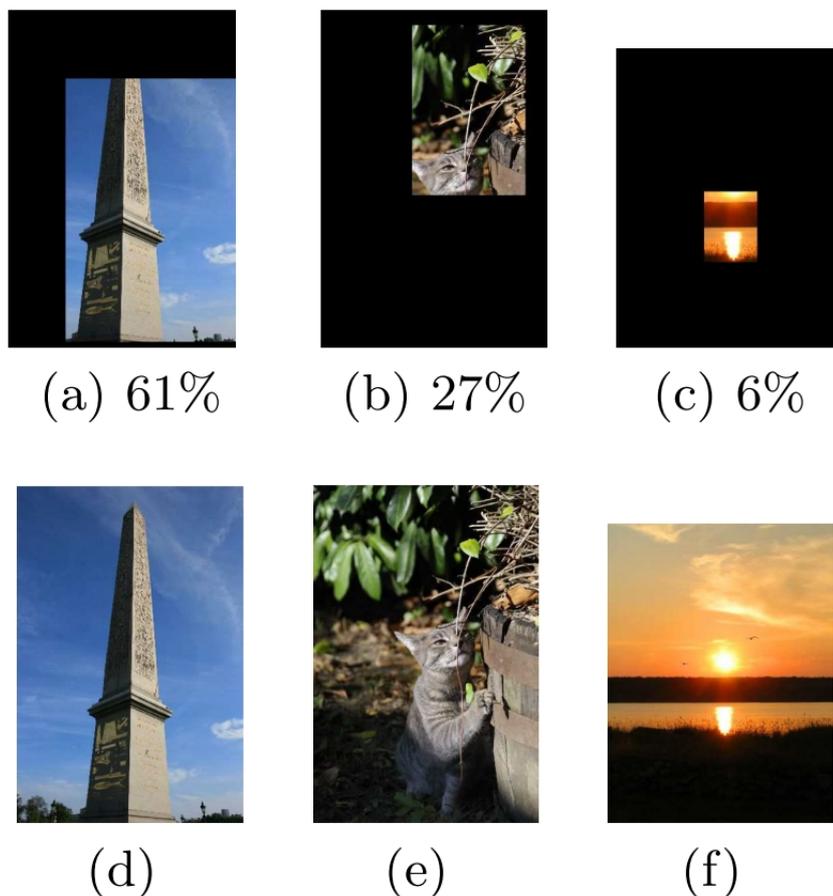


FIGURE 4.2.5 : (a)-(c) Exemples d'images évaluées lors d'une seconde campagne avec le pourcentage d'image présentée, avec en (d)-(f) les images complètes.

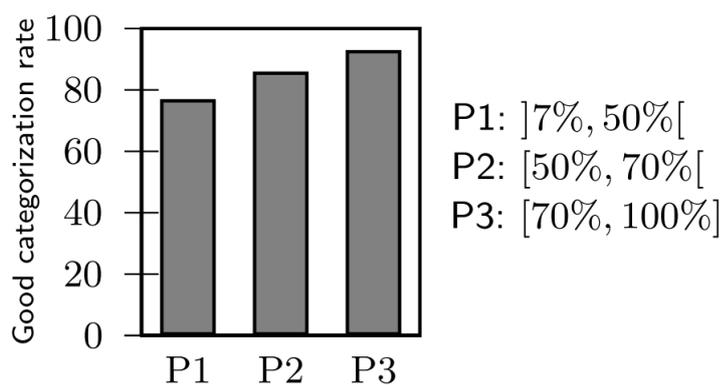


FIGURE 4.2.6 : Évaluation similaire entre les images réduites et complètes

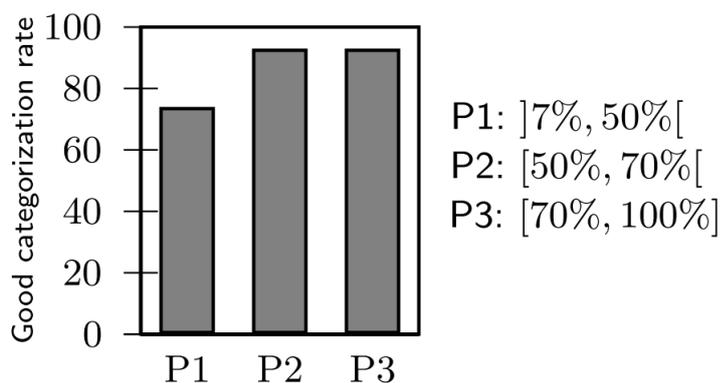


FIGURE 4.2.7 : Pourcentage d'images désormais classées positive ou négative en fonction du pourcentage d'images visualisées

sentons quelques résultats dans la figure 4.2.8.

4.2.3 issues d'indices objectifs

C'est en partie pour répondre à ces critiques que j'ai participé à un projet permettant d'utiliser des indices plus objectifs, en l'occurrence des trajectoires d'eye-tracking. Ce projet est le projet VISIIR³ porté par le LIP6 et financé par l'ANR sur la période 2014-2017. L'objectif de VISIIR est d'explorer de nouvelles méthodes pour l'annotation sémantique d'images. Cette thématique présentée au début de ce chapitre a été largement étudiée lors de la dernière décennie, du fait du très grand nombre d'applications qui en découlent, dans des domaines aussi variés que la recherche d'information, la vision artificielle, le traitement d'images ou encore l'intelligence artificielle. Les méthodes d'apprentissage supervisé nécessitent un grand nombre de données annotées pour être efficaces. En recherche d'information interactive (CBIR), l'utilisateur introduit une requête image, et la recherche s'effectue avec des boucles d'interaction pour affiner le concept sous-jacent recherché.

Le projet VISIIR a pour ambition d'explorer de nouvelles méthodes d'apprentissage interactif pour l'annotation sémantique. L'originalité de notre soumission se situe à trois niveaux :

- Mise au point d'un système de recherche et d'annotations d'images interactif exploitant un eye-tracker. Une spécificité majeure du projet est d'utiliser les dernières générations d'eye-trackers pour améliorer les modèles de vision et d'apprentissage.

3. Visual Seek for Interactive Image Retrieval

4.2 Vers la prise en compte d'informations de plus haut niveau...

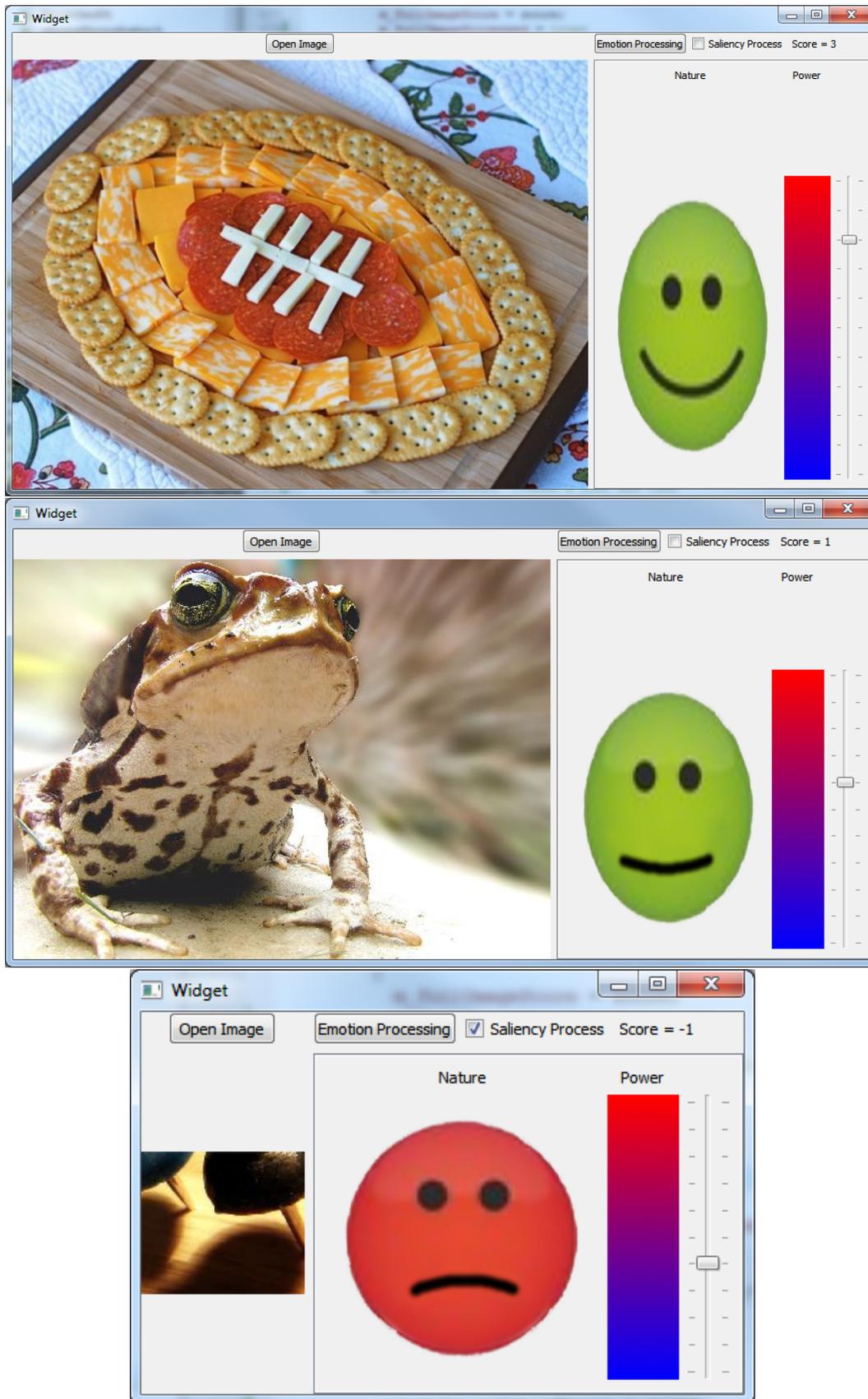


FIGURE 4.2.8 : Implémentation de l'approche information émotionnelle

- Nouveau paradigme pour l'apprentissage de représentations visuelles.
- Filtrage du web pour une application d'annotation de données images de recettes de cuisine. Les nouvelles méthodes développées dans VISIIR seront validées sur une application web de recherche de catégories d'illustration de plats cuisinés.

C'est dans le premier point que je suis porteur de la sous-tâche *New visual saliency models*.

Dans les chapitres précédents, j'ai montré que je pouvais obtenir un système de modélisation de l'attention visuelle robuste, temps réel fiable grâce à des indices bas-niveaux couplés à des EDPs. L'objectif de cette tâche est donc d'améliorer ce modèle en :

- utilisant la saillance visuelle comme descripteur d'image,
- combinant de cartes de bas niveau avec des informations de niveau supérieur,
- intégrant un retour de pertinence utilisant l'eye tracking.

Pour l'heure, nous sommes en phase de développement et nous n'avons que peu de résultats en partie parce que nous ne possédons pas encore d'eye tracking à disposition.

4.3 Une tentative d'approche formelle concernant l'intégration de connaissances *expert*

Ses travaux plus anciens, puisque datant de 2010, ont trait au document. Le document est une orientation majeure prise par le L3I au cours de ces dernières années.

J'ai orienté mes recherches dans ce domaine tout en essayant de rester dans ma diagonale de recherche. Ses travaux n'utilisent pas la saillance, mais s'appuient sur une démarche plus théorique et essaient de proposer une approche formelle d'une problématique de recherche d'information dans des documents anciens. Ces recherches ont été publiées dans plusieurs conférences, un chapitre de livre et une revue ([Coustaty 10a, Coustaty 10b, Coustaty 12]).

Dans ces travaux, nous présentons une méthode complexe pour la caractérisation et l'indexation d'images graphiques de documents anciens.

Le constat préalable à ces travaux étaient que de plus en plus de bibliothèques nationales, de projets nationaux, européens ou mondiaux (NaviDoMass (2010); Europeana; Passe-Partout; Google Livres; Open Content Alliance,...) cherchent à préserver leur patrimoine documentaire. Dans cette optique, de grandes campagnes de numérisation sont actuellement menées par ces différents projets qui cherchent à sauvegarder en masse, au format image, des copies de ces documents. Ces campagnes numérisent de grandes masses de documents et nécessitent de plus

en plus souvent des services de navigation pour permettre aux usagers de retrouver les documents.

Dans le cadre de ces recherches, nous nous sommes intéressés plus particulièrement à la recherche par le contenu d'images graphiques de documents du *XV^{ème}* et *XVI^{ème}* siècle et plus particulièrement aux lettrines qu'elles contiennent. Ces documents nous ont été fournis par le Centre d'Études Supérieures de la Renaissance⁴. Ces lettrines correspondent à des images très utilisées dans les ouvrages et très réutilisées au cours du temps. Une lettrine est une lettre ornementale qui débute un chapitre ou un paragraphe et peut-être vue comme une image binaire composée de traits. Les historiens que nous avons interrogés, décomposent les lettrines en quatre éléments. Elles peuvent être vues comme une superposition de trois couches (fond, motif et lettre) insérées dans un cadre (voir Figure 4.3.1).

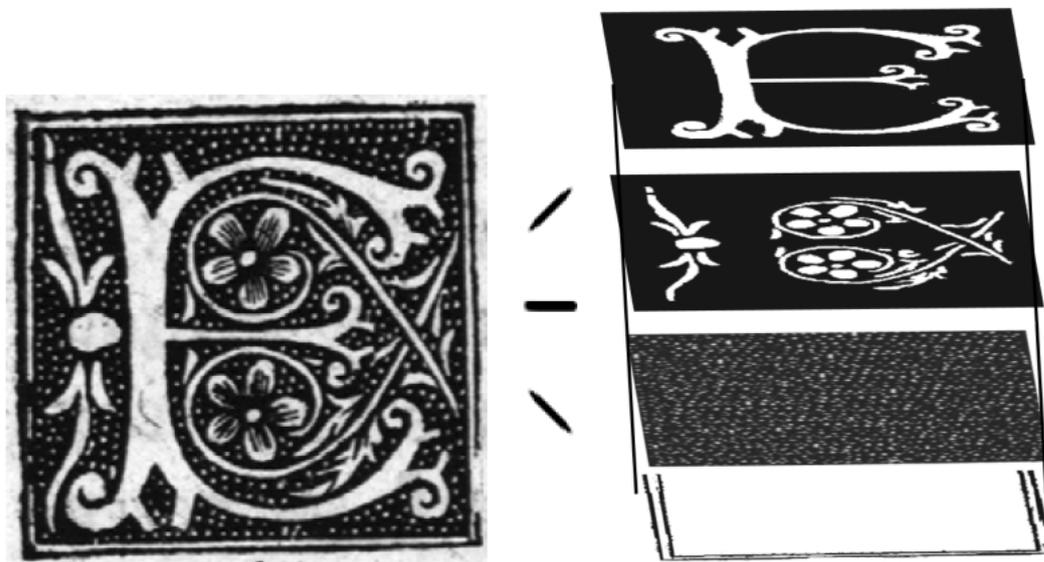


FIGURE 4.3.1 : Exemples de lettrine et sa décomposition

- Le fond : correspond à l'arrière plan et peut-être plein (noire ou blanc), hachuré ou criblé.
- Le motif : correspond aux ornements. Il peut être décoratif ou figuratif.
- La lettre : est un élément clé qui peut-être noir ou blanc et de différentes polices.
- Le cadre : composé d'aucun, un ou deux filets.

Les historiens cherchent à retrouver les images dans de grandes masses de données pour retracer l'histoire et la vie de l'époque. Ils extraient cette information en identifiant les changements qui apparaissent dans les images. Ces dégradations

4. <http://cesr.univ-tours.fr/>

sont intéressantes pour les historiens puisqu'elles permettent de déterminer un ordre chronologique entre les documents. Tous ces changements sont identifiés manuellement par les historiens à l'aide d'un thésaurus. Cette tâche est fastidieuse et non envisageable dans un contexte d'analyse automatique de grandes masses de données.

Notre objectif était donc de retrouver de l'information pilotée par des connaissances expert.

Dans ces travaux, j'ai proposé d'utiliser une approche *complexe*.

Pour reprendre encore un extrait d'Edgar Morin (1996) :

le principe de simplicité impose de disjoindre, le principe de complexité conjoint de relier, tout en distinguant.

Toutes les approches scientifiques jusqu'à la moitié du *XX^{ème}* siècle cherchaient à simplifier les problèmes compliqués pour les résoudre, or il faut dissocier les problèmes complexes des problèmes compliqués. La complexité par définition signifie ce qui est tissé ensemble, ce qui est relié, ce qui n'impose en rien d'être composé de problèmes compliqués. Au contraire, là où un problème compliqué va nécessiter une simplification sans chercher à respecter son environnement, un problème complexe nécessitera des allers-retours entre description certaine par simplification et description incertaine de son contexte.

Il m'a paru donc essentiel d'intégrer, dans un schéma d'ensemble, séparabilité et logique, avec une séparabilité de l'information pour gagner en certitude et un respect de la logique globale.

La complexité dans les documents s'articule autour d'une décomposition. Mais la complexité impose également d'ordonner et de remettre cette certitude dans son contexte, afin d'obtenir une description cohérente de l'image dans celui-ci. La description d'une lettrine d'un point de vue texture uniquement, sans respect de la lettre ne paraît pas cohérente. Elle correspond à une vision de traicteur d'images qui consiste à ne s'appuyer que sur des descripteurs bas niveau (de couleurs, de formes, de textures, ...).

L'objectif global est d'essayer de revenir à des informations utilisées dans la production des lettrines pour obtenir une esthétique visuelle. Dans ce cadre, l'idée est de regrouper par "catégories" les différentes couches visuelles (dans notre cas : textures, formes, contours, ...).

La complexité implique un entremêlement d'éléments hétérogènes. Pour décrire chacun de ces éléments, j'ai proposé de les séparer pour les étudier. Cette séparation repose sur une analyse en "Échelles et Unités Fonctionnelles", qui permettent de naviguer au sein de la description entre incertitude et imprécision. Cette analyse permet de mettre en avant des éléments particuliers de l'image auxquels sont associés des traitements particuliers et adéquats. Chacun de ces traitements sera intégré dans une unité fonctionnelle, rassemblées elles-mêmes au sein d'échelles

fonctionnelles. Chaque échelle permet d'avancer dans la description de l'image entre imprécision et incertitude sur la nature de son contenu. Le schéma de la Figure 4.3.2 présente cette démarche dans son ensemble.

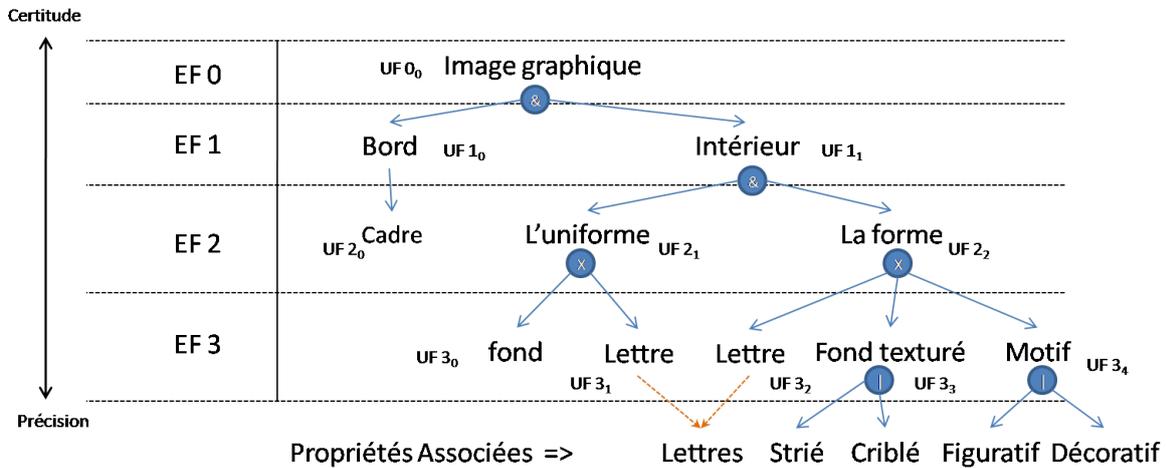


FIGURE 4.3.2 : Illustration du schéma de caractérisation des éléments constitutifs des lettrines

Nous appelons *unité fonctionnelle* une unité capable d'extraire et de caractériser un (ou des) élément(s) aux spécificités semblables. Ces unités peuvent être généralisées et/ou adaptées.

Nous appelons *échelle fonctionnelle* un niveau d'observation et/ou d'analyse du contenu de l'image réunissant des unités fonctionnelles cohérentes.

Les articles [Coustaty 10a, Coustaty 12] présentent en détails le fonctionnement des unités fonctionnelles UF_{2_2} , UF_{3_1} et UF_{3_2} ainsi que l'évaluation de l'approche. Les figures suivantes montrent des exemples de résultats obtenus.

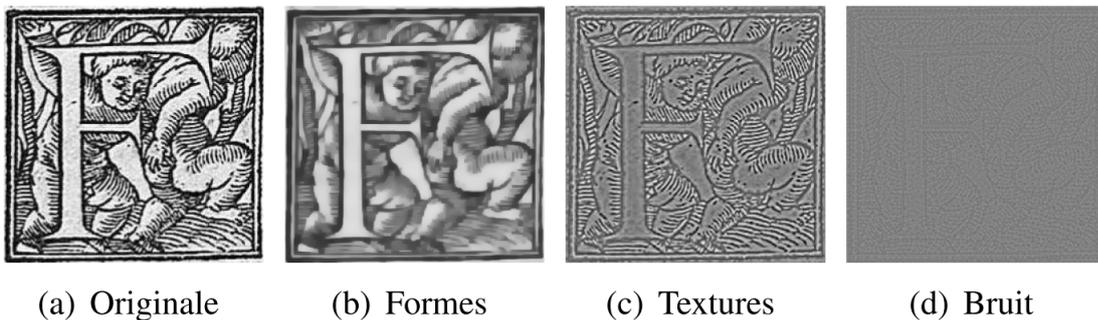


FIGURE 4.3.3 : Exemple de décomposition de lettrine



FIGURE 4.3.4 : Exemples de lettres extraites automatiquement à partir des lettrines

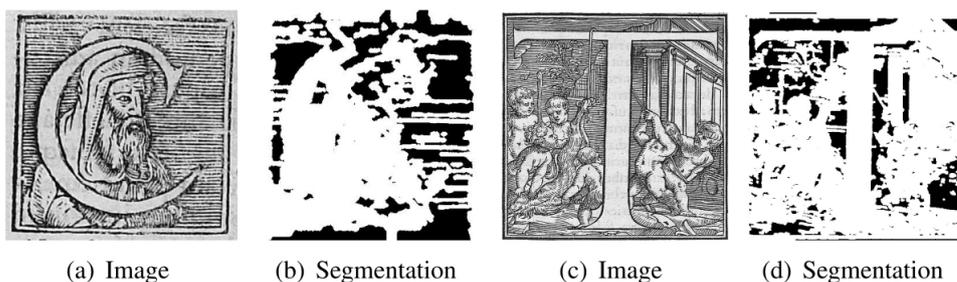


FIGURE 4.3.5 : Exemple de segmentation d'une texture dans une lettrine

4.4 Conclusion

Dans ce chapitre, j'ai présenté mes contributions autour de la problématique de la *recherche prégnante d'information* de la diagonale informationnelle.

J'ai scindé ce chapitre en deux, présentant tout d'abord mes travaux concernant la prise en compte d'informations de haut niveau issues :

- du web,
- d'indices subjectifs,
- d'indices objectifs.

Dans une seconde partie, j'ai exposé une approche plus formelle de la prise en compte d'informations haut niveau issues de connaissance expert.

Dans tous les cas, ces travaux sont des travaux charnières entre ma recherche entreprise depuis 10 ans et mon projet de recherche.

Le chapitre suivant va présenter mon projet de recherche.

Points clés

Positionnement

- ❑ Recherche concernant la *recherche prégnante d'information*
- ❑ Réflexions sur l'intégration d'informations haut niveau

Contributions

- ❑ Application aux documents numérisés
- ❑ Graduation de l'information haut niveau (web, subjective, objective)

Publications

- N. Rempulski, A. Prigent, P. Estrailier, V. Courboulay and M. Perreira Da Silva, Adaptive Storytelling based on model-checking approaches, IJIGS - International Journal of Intelligent Games and Simulation, 2009, Volume 5, Number 2, 33-42.
- Mickaël Coustaty, Vincent Courboulay, Jean-Marc Ogier, "Analyzing Old Documents Using a Complex Approach : Application to Lettrines Indexing " in Advances in Knowledge Discovery and Management. Volume 2 Series : Studies in Computational Intelligence, (2012) 155-172
- Gbehounou S., Lecellier F., Fernandez-Maloigne C., Courboulay V. Can salient interest regions resume emotional impact of an image ? Dans 15th international conference on computer analysis of images and patterns - CAIP 2013 - 15th international conference on computer analysis of images and patterns, Royaume-Uni (2013)
- Duthil B., Coustaty M., Courboulay V., Ogier J.-M. Visual Saliency and Terminology Extraction for Document Annotation Dans 13th ACM Symposium on Document Engineering (DocEng) - 13th ACM Symposium on Document Engineering (DocEng), Italie (2013)

Projet de Recherche

« La vraie nouveauté naît
toujours dans le retour aux
sources »"

(Edgar Morin)

Chapitre 5

De la valorisation à l'avalorisation - du big data au big dat@rt

Dans ce chapitre, je présente mon projet de recherche qui oriente mes choix depuis maintenant deux ans. C'est principalement cette orientation nouvelle qui m'a fait sentir le besoin de rédiger ce document.

Dans la première partie de ce chapitre, je présente le projet en lui même ainsi que le contexte sociétal et scientifique dans lequel il prend place, les travaux déjà entrepris et les collaborations initiées.

Dans la seconde partie, je positionne mon projet selon deux perspectives, une perspective spatiale et une autre temporelle. Je propose d'envisager mon projet selon une vision court, moyen et long terme, ainsi que selon sa portée locale, nationale et internationale.

5.1 Problématiques sociétales

5.1.1 De la valorisation ...

Mon projet de recherche trouve son origine dans l'apparition progressive et irréversible des données massives ou *big data* dans des écosystèmes de plus en plus complexes. Traditionnellement, le big data se caractérise par ses 3V :

- Volume : selon une étude Gartner, les données numériques créées dans le monde seraient passées de 1,2 zettaoctets par an en 2010 à 1,8 zettaoctets en 2011, puis 2,8 zettaoctets en 2012 et s'élèveront à 40 zettaoctets en 2020 (un zettaoctet équivaut à 10^{21} octets, cf. Figure 5.1.2)
- Variété : ces données sont brutes, semi-structurées voire non structurées, provenant du web (Web Mining), au format texte (Text Mining) et images (Image

Mining), elles peuvent être publiques (Open Data, Web des données) ou privées.

- Vitesse : c'est à la fois la fréquence à laquelle les données sont générées, capturées et partagées.

Certains rajoutent 2V supplémentaires :

- Valeur,
- Visualisation.

L'exploitation de ces données massives dont disposent les entreprises et les pouvoirs publics sont porteuses d'applications nouvelles et de gains de compétitivité considérables dans des domaines aussi variés que la santé, l'environnement, la culture, le tourisme, le marketing, l'éducation, les études historiques... Des projets émergents comme les *smart cities*¹ ou les *smart grids*² sont et seront fournisseurs et utilisateurs de données massives qu'il faudra traiter en temps réel.

La valorisation de ces données révèle un caractère fondamental³. La capacité pour les entreprises, les individus et les objets connectés à exploiter de façon pertinente ces énormes quantités d'informations est un enjeu d'autant plus important que des données issues de secteurs éloignés d'une entreprise peuvent être d'un intérêt primordial pour elle⁴.

Un chiffre : Mc Kinsey estime qu'en 2025, les Big Data représenteront 5 000 milliards de dollars par an.

Au titre de l'individu, la révolution du Big Data rencontre des tendances sociétales de fond.

Le citoyen consommateur souhaite avoir accès à de plus en plus d'informations pour décider par lui-même, il souhaite devenir *consom-acteur* (cf section et la notion de *spect-acteur*). Cette nouvelle position exige une information personnalisée. L'information extraite et présentée doit donc être individualisée pour répondre à son besoin.

En contrepoint de cette valorisation extrême, je revendique le besoin, voire la nécessité, d'une *avalorisation* salutaire.

5.1.2 ...à l'avalorisation des données

Ne cherchez pas dans un dictionnaire, *avalorisation des données* n'existe pas. Je propose de penser les données pour elles mêmes et pas pour leur utilisation, d'où l'avalorisation des données.

1. <https://www.youtube.com/watch?v=0iV4KpPzPRI&feature=plcp>

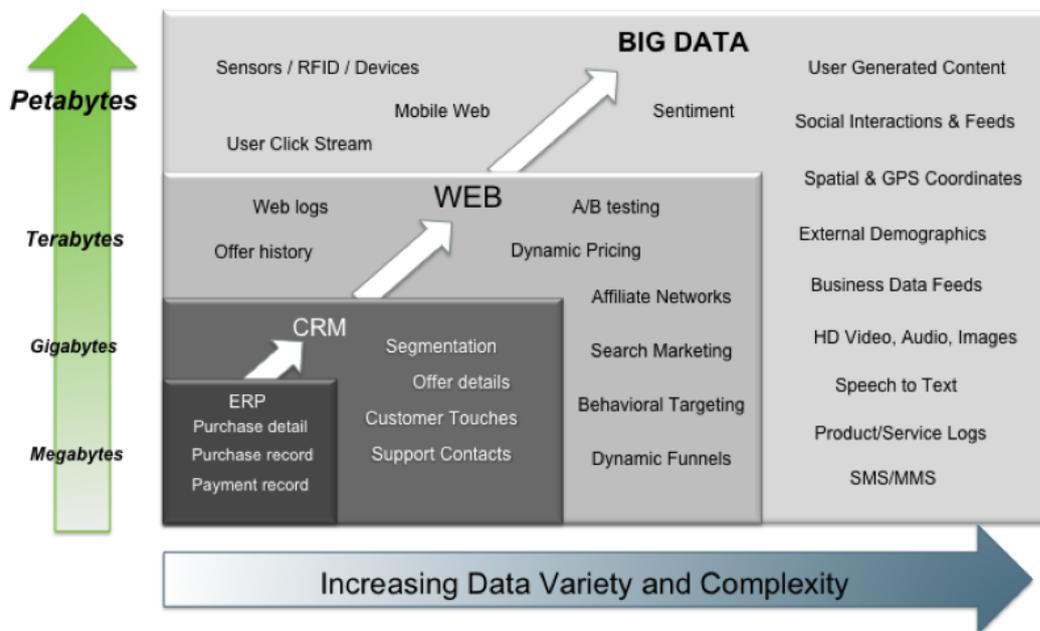
2. <http://www.smartgrids-cre.fr/>

3. je ne sais toujours pas ce qu'il y a au dessus de fondamental ? (cf première note de bas de page de la section 1)

4. <http://sciences.blogs.liberation.fr/home/2008/11/google-detecte-1.html>



Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

FIGURE 5.1.1 : Quelques images illustrant le Big Data

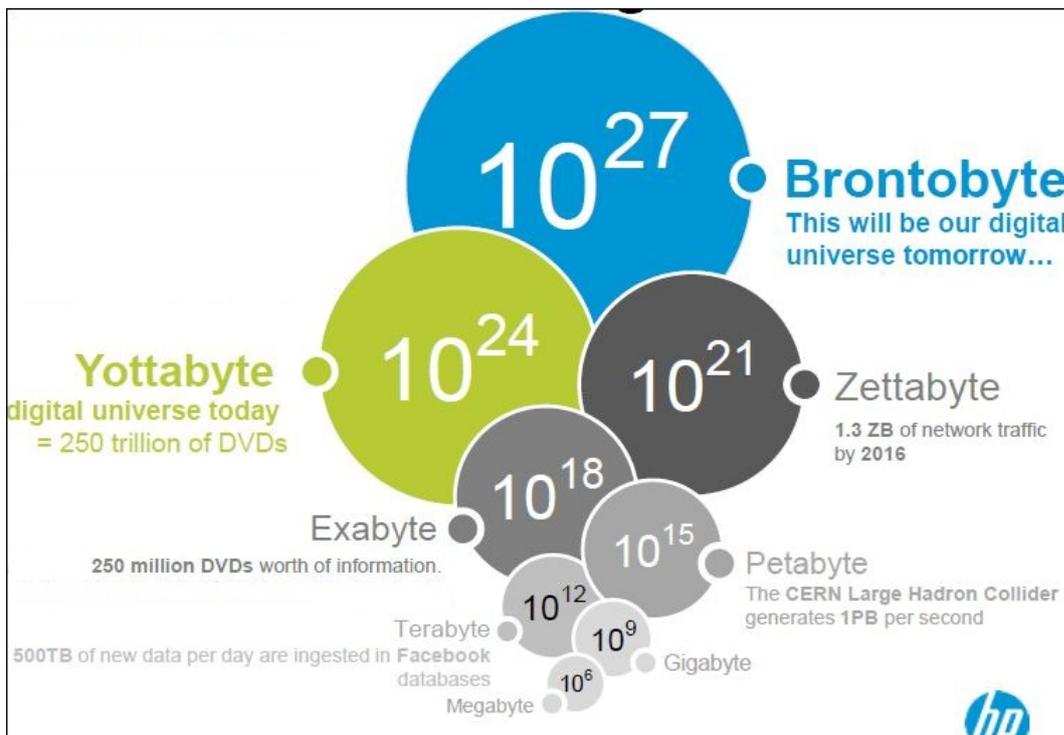


FIGURE 5.1.2 : Les échelles du Big Data (source HP)

La notion de Big Data est réellement abyssale (cf. Figure 5.1.2). Que représente donc un zettaoctet ? Peut-on envisager un brontobyte ?

Jusqu'à aujourd'hui les arts numériques étaient portés par la puissance des machines. Ce que je propose dans mon projet de recherche et d'inverser le paradigme habituel de traitement comme MapReduce et Hadoop ont inversé le paradigme de traitement des données.

Hadoop a proposé la délocalisation des scripts de traitements dans les données au lieu de l'importation des données vers les scripts de traitements.

Pourquoi ne pas faire porter l'art numérique par les données et non par la puissance de calculs des machines ? Les jeux vidéos bénéficient de la puissance phénoménale des consoles, mais l'onirisme d'un *Child of Light*⁵ ou d'un *Myst*⁶ éternel ne se mesurent pas à la quantité de GHz.

Considérons différemment les données de par leur volume, leur hétérogénéité, leur rapidité d'apparition ...

Je trouve nécessaire de proposer une vision *avalorisante* des usages des données numériques pour permettre au grand public de toucher du doigt l'immensité de l'Univers numérique et d'ouvrir des portes que des artistes empruntent déjà pour certains⁷.

5.2 Problématique et démarche scientifique

En adoptant la même démarche que celle présentée au premier chapitre, je propose ici une nouvelle problématique scientifique ainsi que la démarche associée.

Émergence de l'information dans un système complexe.

Cette problématique rencontre une réalité dont nous venons de parler, celle de la complexification des données (massives, hétérogènes, rapide...). L'ère du big data nous fait évoluer sur la vision que nous avons de l'information. Il est quasi impossible de parler d'extraction de l'information, cela supposerait qu'il suffit de *creuser* pour la *trouver*. Non, l'exploitation des données aujourd'hui requiert un raffinement, une association, voire une transformation pour enfin extraire ce que nous souhaitons.

C'est la raison pour laquelle je parle plutôt d'*émergence*.

L'émergence est un concept apparu au XIXe siècle et qui peut être résumé par « le tout est plus que la somme de ses parties ». Plus précisément, une propriété peut être qualifiée d'émergente si elle « résulte » de propriétés plus fondamentales tout en demeurant « nouvelle » ou « irréductible » à celles-ci.

5. <http://childoflight.ubi.com/>

6. http://www.dailymotion.com/video/xc3lbv_walkthrough-myst-partie-1_videogames

7. <http://bigdataweek.com/2014/04/28/big-data-art/>

Cette problématique soulève deux questions corolaires :

- De quel type d'émergence parle-t-on ?
- De quel type de complexité parle-t-on ?

5.2.1 Les différents niveaux d'émergence

Pour reprendre la présentation de Wikipédia, il est généralement admis qu'il en existe deux catégories principales⁸ :

- émergence « forte » : lorsque l'évolution des systèmes est non seulement imprévisible mais incompréhensible par ceux qui adhèrent encore aux états anciens de ces systèmes. L'existence de telles propriétés émergentes fortes est souvent discutée : l'impossibilité, dans tout système complexe, de voir le lien causal entre le niveau micro et le niveau macro pourrait résulter des limites de l'être humain. On peut citer deux exemples d'émergence qui semblent fortes : l'apparition de la vie à partir de l'inanimé, et l'émergence de la conscience.
- émergence « faible » : lorsque la dynamique causale du tout est entièrement déterminée par la dynamique causale des parties, et est subjectivement interprétée comme émergente par un observateur extérieur. Il n'y a pas de propriété *réellement* nouvelle, si ce n'est celle qu'un humain peut observer, et qui peut surprendre ce dernier de premier abord. Encore faut-il être capable de les observer et de les interpréter, c'est-à-dire être déjà en partie engagé dans le nouvel état émergent. Le domaine de la robotique est un exemple particulièrement pertinent⁹.

5.2.2 Les différents niveaux de complexité

La complexité est une notion utilisée en philosophie, épistémologie, en physique, en biologie (avec les travaux d'Henri Atlan) ou en écologie.

Pour Edgar Morin, la complexité est un concept qui exprime une forme de pensée acceptant les imbrications de chaque domaine de la pensée et la transdisciplinarité. Le terme de complexité est pris au sens de son étymologie « complexus » qui signifie « ce qui est tissé ensemble » dans un enchevêtrement d'entrelacements (plexus) [Morin 05].

Ce qui est complexe est donc ce qui est composé d'éléments qui entretiennent des rapports nombreux, diversifiés, difficiles à saisir par l'esprit, et présentant souvent des aspects différents.

Il ne faut pas confondre compliqué et complexe, tout comme il ne faut pas confondre ce qui est simple de ce qui est simplexe [Berthoz 09].

8. <http://fr.wikipedia.org/wiki/Emergence>

9. <http://www.automatesintelligents.com/edito/2005/mai/edito2.html>

Le premier à avoir appréhendé et modélisé la complexité en différents niveaux fut Kenneth E. Boulding en 1956 dans un article du *Management Science*, dénommé *General Systems theory - The skeleton of science* [Boulding 56]. Il propose un modèle en neuf niveaux auquel il ajouta deux niveaux supplémentaires en 1985 (cf Tableau 5.1).

Niveau	Modèle de 1956	Modèle de 1985
1	Canevas	Système mécanique
2	Horloge	Système cinématique
3	Thermostat	Système à rétroaction positive
4	Cellule	Système créodique
5	Plante	Système reproductif
6	Animal	Système démographique
7	Humain	Système écologique
8	Organisation sociale	Système évolutionnaire
9	Système transcendantal	Système humain
10		Système social
11		Système transcendantal

TABLE 5.1 : Modèles de Boulding

Ludwig von Bertalanffy, connu comme le fondateur de la Théorie systémique, a repris le modèle de Boulding et en a fait une nouvelle présentation en 1968 dans *eneral System Theory* [Bertalanffy 02] (cf Tableau 5.2).

Enfin, le dernier « modélisateur » de la complexité est Jean Louis Le Moigne. Ce spécialiste français de la systémique a pour domaines de recherche privilégiés les sciences des systèmes, de l'ingénierie, de l'intelligence artificielle.

Dans son livre [Moigne 77], il définit un modèle en 9 couches (cf Tableau 5.3).

On peut voir que le sujet de la complexité est vaste et ... complexe. J'ai volontairement simplifié ces modèles en couches pour n'en conserver que trois dans mon projet de recherche. Ces trois niveaux correspondent à trois types de connaissances intégrées au système.

- Faible : depuis le système passif, jusqu'au système informé. Ce niveau s'étend, dans le modèle de von Bertalanffy, aux structures à systèmes ouverts.
- Moyenne : niveau correspondant au organisme de bas niveau dans le modèle de von Bertalanffy, et aux systèmes décideurs.
- Forte : ce niveau intègre tous les niveaux supérieurs, de l'animal à l'homme jusqu'aux systèmes artistiques.

Niveau	Théorie et modèles
Structures statiques	Formules structurales de la chimie
Mouvements d'horlogerie	Physique conventionnelle
Mécanismes d'autorégulation	Cybernétique
Systèmes ouverts	Extension à la théorie physique à des systèmes qui se maintiennent eux-mêmes
Organismes de bas niveau	La théorie et les modèles ont tendance à manquer
Animaux	Débuts de la théorie des automates, rétroaction, comportement autonome
Homme	Théorie naissante du symbolisme
Systèmes socio-culturels	Lois statistiques et peut-être dynamique de la dynamique des populations, sociologie, économie, peut-être histoire. Début de la théorie des systèmes culturels.
Systèmes symboliques	Algorithmes symboliques (mathématiques, grammaires), "règles du jeu" comme dans les arts visuels, musique

TABLE 5.2 : Hiérarchie des systèmes de von Bertalanffy (source Wikipedia)



FIGURE 5.2.1 : Jean-Louis Le Moigne (à droite) avec ... Edgar Morin

Niveau	Système
1	Système passif
2	Système actif
3	Système régulé
4	Système informé
5	Système décideur
6	Système mémorisant
7	Système coordonné
8	Système auto-organisé
9	Système auto-finalisé

TABLE 5.3 : Modèle de Le Moigne

5.2.3 Démarche scientifique

Une fois ces classifications établies, je propose, comme dans la section 1.3, **une matrice support de mon projet de recherche.**

Complexité Émergence	Faible	Moyenne	Forte
Faible			
Forte			

FIGURE 5.2.2 : Matrice support de mon projet de recherche

5.3 Intégration des travaux en cours

La Figure 5.3.1 représente l'intégration des travaux initiés depuis moins de deux ans dans la matrice support à mon projet de recherche. Ces travaux s'appuient sur un réseau de collaborations que j'ai impulsé d'abord de manière informelle puis, de par les travaux entrepris pour créer l'Institut de Valorisation des Contenus Numériques, plus officiellement.

	Complexité	Faible	Moyenne	Forte
Émergence		Requêtes	Règles	Interaction
Faible	Documents	Annotation sémantique des documents administratifs	Valorisation des usages des documents juridique	Émotions et images
	Robotique	Implémentation de systèmes attentionnels	Développement d'un système attentionnel	Émergence d'intérêt
Forte	Art	Datavisualisation (projet BigDatArt)		Emergence émotionnelle

FIGURE 5.3.1 : Positionnement des travaux en cours et des perspectives identifiées.

L'ensemble des perspectives évoquées dans les chapitres précédents se retrouvent par conséquent dans ce tableau.

5.4 Exemples de projets

5.4.1 Projet *BigDatArt*

Un premier projet auquel je souhaite donner corps consiste en une étude informelle autour du *BigDatArt*.

Le projet soumis consiste à mélanger la recherche attentionnelle d'information visuelle et textuelle avec une interaction possible entre des personnes à La Rochelle et Mons.

Mon idée serait de profiter du fait que Mons sera la Capitale européenne de la Culture en 2015¹⁰. Dans cette optique, j'ai soumis un projet *Tournesol*¹¹ avec Polytech Mons, et obtenu le soutien de la maison de la culture de l'université de la rochelle.

10. <http://www.mons2015.eu/fr/>

11. Tournesol est le Partenariat Hubert Curien (PHC) mis en place avec la Fédération Wallonie-Bruxelles.

Le projet *BigDatArt* nécessite deux installations simples et identiques à Mons et à La Rochelle¹² : une tv grand écran couplée avec une Kinect.

Sur cet écran plusieurs zones. Une partie image et une autre nuage de mot, l'avatar de celui qui est en face peut également être présent. J'envisage aussi un système de sonorisation ainsi qu'un rendu tactile par tablette braille. L'ensemble de la pièce peut être couverte par le rendu (mur et plafond) Toutes ces parties sont évidemment reliées par un lien masqué au spectateur (ou spect-acteur si il interagit avec l'installation). Je souhaite le faire interroger sur les liens existants entre les images et les données, le faire cheminer d'un détail d'une image à une image complète en compagnie de chansons ou textes choisis en fonction de mots clés générés par un moteur de recherche. Cheminer sur les ressources sans fin du web avec pour guide un moteur impersonnel. L'idée étant d'agir sur le plus de sens de l'observateur.

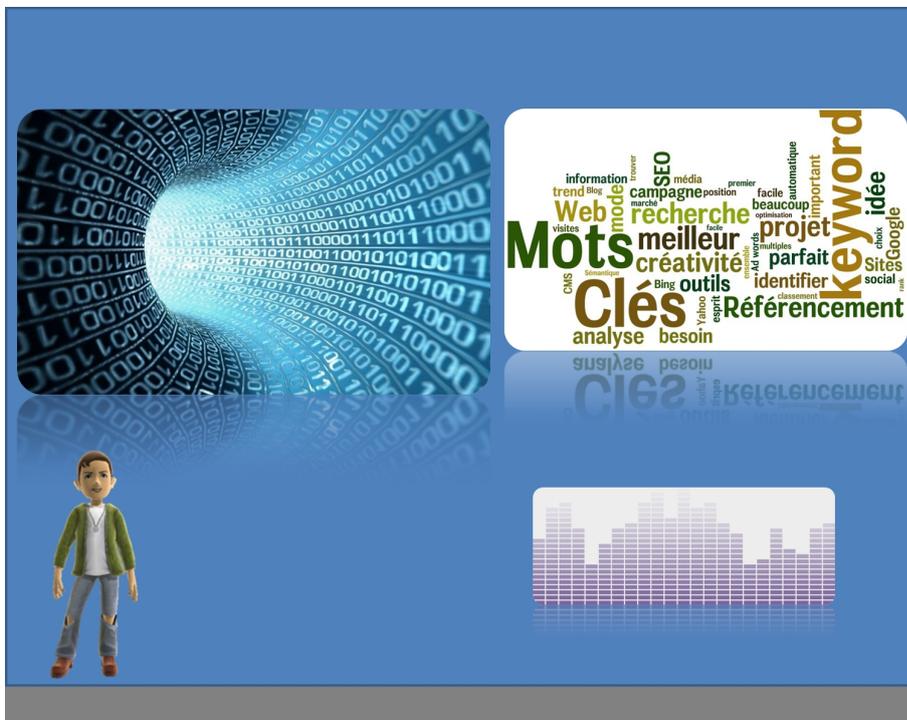


FIGURE 5.4.1 : Proposition d'interface du projet *BigDatArt*

Le système a trois modes de fonctionnement :

- autonome, *du hasard à l'état pur* : une image est tirée au sort, on fait un zoom sur la partie la plus saillante. Cette vignette est envoyée à Google qui renvoie une image similaire (sur des critères textes ou images (cf notre papier

12. Je travaille aussi à son installation à Sarajevo, cf section 5.4.2

- d'annotation automatique). Cette image est présentée puis on reboucle au début du process sans fin ;
- semi interactif, *de la sérendipité à l'état pur* : une personne prend place devant une des deux installations et peut interagir avec le système soit en changeant la zone à zoomer soit en jouant avec le nuage de mot pour choisir celui qui l'intéresse ;
 - totalement interactif : sorte de *battle* multimédia entre deux "joueurs" l'un à Mons l'autre la Rochelle.
 - Un jeu sans règle
 - un gameplay à inventer
 - analogie en image du jeu où on passe d'un mot à un autre en ne changeant qu'une lettre. On affiche deux images, une début, une de fin la personne choisit un élément de l'image on interroge le web pour une série de mots clés ses mots clés sont la requête d'une image on affiche l'image on reboucle jusqu'à temps que la distance des mots clés de l'image sont inférieurs à un seuil.

Nous envisageons également de jouer sur deux autres types de saillance. Une saillance sonore : les mots clés de la vignette sont envoyés à une base internationale de parole de chansons et la chanson la plus pertinente est jouée pendant la phase de zoom. L'autre saillance est tactile : une plage braille écrira les paroles de la chanson sous les doigts des observateurs.

Enfin, l'utilisation de lunettes permettant un suivi oculaire permettra une nouvelle sorte d'interaction.

J'ai d'ailleurs eu l'occasion d'essayer les Google Glass lors de ma visite à l'Université de Bratislava, l'expérience est ... surprenante, tu es *OK GOOGLE?* (cf Figure 5.4.2).

Pour finir, ce projet pourra, je l'espère, être déployé sur d'autres lieux, il intéresse déjà l'artiste plasticienne et chorégraphe Emmanuelle Grangier.

5.4.2 Projet de valorisation des archives pénales internationales

Le L3I est un partenaire fondateur de la Chaire d'excellence consacrée à « La gestion du conflit et de l'après-conflit ». La responsabilité pénale des personnes et des entreprises est désormais un enjeu juridique et économique de la société internationale en voie d'organisation. La justice pénale internationale s'est jusqu'à présent focalisée sur la répression des crimes de guerre, des crimes contre l'humanité et des génocides. D'autres champs se développent aujourd'hui : la justice transitionnelle et la gestion mémorielle des conflits. D'autre part le monde des affaires prend conscience de sa responsabilité en zone de conflit ainsi qu'en matière environnementale. L'aspect qui m'intéresse le plus en tant que correspondant de l'université pour cette chaire, réside dans création d'un *centre international de*

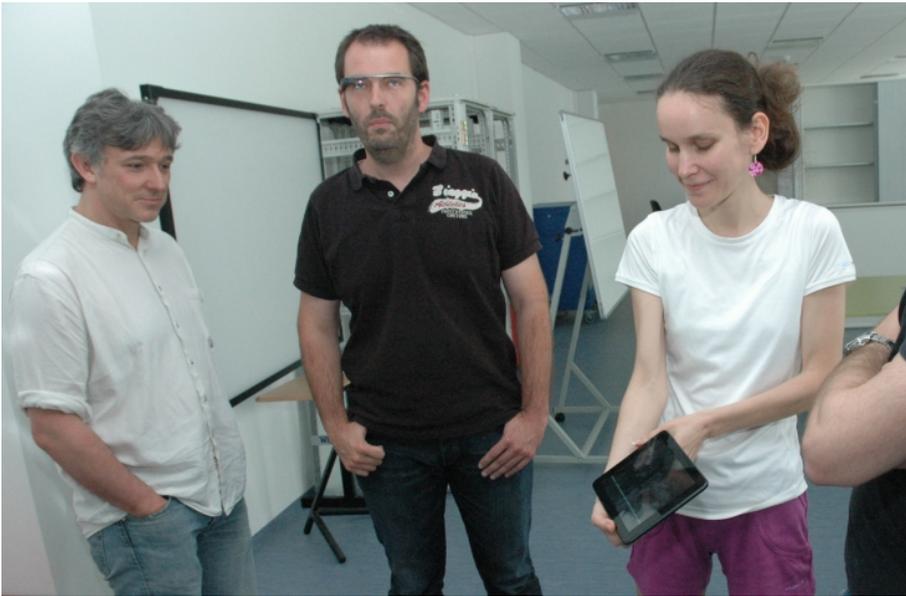


FIGURE 5.4.2 : Google Glass et moi à l'Université de Bratislava (Juin 2014). Attention l'abus de Google Glass nuit sérieusement à votre image.

documentation, de recherche et d'expertise sur la conflictualité.

Aujourd'hui, malgré l'incroyable explosion de la numérisation de notre patrimoine, il est encore très difficile de classer automatiquement, de trier ou de rechercher de l'information à l'intérieur de ces connaissances. Une des caractéristiques principales de ces bases réside dans l'hétérogénéité des données (texte, vidéo, images, ...) On peut aussi trouver de l'information dans les métadonnées associées aux différents supports de données. Toutes ces informations peuvent nous permettre d'enrichir ces données en annotant et en indexant plus efficacement ces données hétérogènes.

L'objectif final étant de pouvoir accéder rapidement et précisément à l'information contenue dans ces différents supports. Mais ce n'est qu'un premier pas vers la conceptualisation de ces données hétérogènes, à savoir la représentation mentale que se fait l'esprit humain d'un objet pensé, concret ou abstrait, et qui lui permet de rattacher à ce même objet les diverses perceptions qu'il en a afin d'en organiser les connaissances.

L'extraction des différents concepts contenus dans une entité (image, texte, vidéo, etc.) aurait pour objectif de caractériser sur une échelle commune ces différents supports afin de les traiter dans un haut niveau sémantique commun.

Ce projet m'a permis la mise en place d'une mission de numérisation à Sarajevo d'un l'album photo d'une association d'anciens prisonniers de camps Serbes.

Ces photos représentent plus de 650 témoignages visuels de lieux d'exactions

commises pendant la guerre des Balkans. Ce trésor patrimonial se devait d'être sauvé.

5.4.3 Institut de Valorisation des Contenus Numériques

Cet institut en création s'appuie sur un groupe de recherche interdisciplinaire Humanités Droit Management Gestion Informatique (HDMI) regroupant des membres des quatre laboratoires L3i, CEJEP¹³, CEREGE¹⁴, CRHIA¹⁵ afin d'identifier des problèmes pouvant être traités selon une approche pluri ou interdisciplinaire. En effet, le besoin d'avoir une approche interdisciplinaire se fait de plus en plus sentir dans les laboratoires de recherche. Plusieurs initiatives ont ainsi été prises par les chercheurs des différents laboratoires. Certaines d'entre elles ont permis de mettre en place plusieurs projets. La chaire d'excellence : la gestion du conflit et de l'après-conflit évoquée précédemment en est un parfait exemple. On trouve aussi la chaire d'excellence Pierre Deffontaines¹⁶ destinée à accueillir à la fois des chercheurs en sciences sociales développant une approche comparée sur les sociétés américaines et des chercheurs en sciences et technologies de l'information et de la communication (STIC).

5.5 Mise en perspective

En début de chapitre, je proposais de positionner mon projet selon une double perspectives, spatiale et temporelle. A l'heure d'une mutation profonde du système universitaire national, il me paraît essentiel de positionner mon projet dans ce nouveau paysage encore changeant. Dans cette optique je propose encore de positionner ce projet, dans une vision matricielle à court, moyen et long terme, ainsi qu'à une échelle territoriale. La Figure 5.5.1 présente cette mise en perspective.

5.5.1 Vision à court terme

Dans la première colonne, nous retrouvons certaines missions ou projet dont j'ai déjà parlé, comme les chaires d'excellence *Gestion du conflit* et *Pierre Deffontaines*. Ces projets sont le socle d'une demande CPER / FEDER 2014-2020.

Ainsi, dans le cadre du FEDER, les 4 actions de l'axe Méthodologie, modèles et outils pour le numérique (projet ECONUM : Innovation et économie numérique du Programme n°1 : Excellence numérique pour une société innovante) ont été

13. Centre d'Études Juridiques et Politiques, <http://cejep.univ-larochelle.fr/>

14. Centre de REcherche en GEstion, <http://cerege-lr-mos.univ-larochelle.fr/>

15. Centre de Recherches en Histoire Internationale et Atlantique, <http://www.crhia.fr/>

16. <http://flash.univ-larochelle.fr//14-mars-Inauguration-de-la-Chaire.html>

vision	Court terme	Moyen terme	Long terme
Nationale & international	<ul style="list-style-type: none"> • ANR (shades, CartoDialect, audinm (chine)) • Collab UMONS • Reconnaissance par un réseau de recherche 	stratégie de recherche pour le numérique	Innovation 2030 (http://innovation-2030.dgcis.gouv.fr/fr/défi 7) lien rapport
Trans-régionale	Chaires d'excellence (Limoges & La Rochelle)	COMUE & ISITE	
Locale	Institut de valorisation des contenus numériques		
Industrielle	Transferts de technologie Valconum		

FIGURE 5.5.1 : Mise en perspective temporelle et territoriale de mon projet de recherche.

conçues grâce à la collaboration des laboratoires L3i, CEREGE-LRMOS, CEJEP et CRHIA. En ce qui concerne le CPER, il s'agira d'adresser la question de la valorisation des contenus numériques sur quelques cas d'usage, quelle que soit leur nature, sous un angle considérant cette vision globale et systémique, afin de créer les conditions de passage à l'échelle et de permettre le réel transfert vers la société.

Nous retrouvons également des projets ANR dont je fais partie. Je peux citer par exemple le projet SHADES (Hachage sémantique pour la signature électronique avancée de document), ou CartoDialect (Géomatique et Extraction Automatique de Contenus Cartographiques pour la Dialectologie), et surtout Audinm (Analyse et Interprétation d'images de documents sur les réseaux sociaux).

5.5.2 Vision à moyen terme

A moyen terme, mon projet de recherche s'intègre à deux stratégies.

Une stratégie nationale de recherche

En vue de l'élaboration de la stratégie nationale de recherche (S.N.R.), l'état a piloté le travail de près de 400 experts, issus notamment de la communauté universitaire et scientifique. Ils ont travaillé au sein de 10 ateliers afin d'identifier

des priorités scientifiques et technologiques et, à partir de ces priorités, de faire émerger trois ou quatre grandes priorités transversales et communes. La S.N.R. sera finalisée à l'été 2014, mais dès maintenant nous pouvons nous appuyer sur le travail de l'atelier 7 concernant la Société de l'Information et de la Communication. La lecture de ce rapport¹⁷, permet d'appréhender les défis relevés par ce panel d'experts dont la composante transverse interdisciplinaire majeure inclue la sociologie, l'économie et les humanités numériques.

La Figure 5.5.2 présente cette vision avec la criticité des enjeux.

Défis sociétaux	Technologies et services du numérique									
	Composants et micro-nano technologies	Logiciel et programmation	Infrastructures numériques	Robots et objets numériques de communication et de calculs	Simulation et calcul avancés	Données massives pour comprendre, prévoir et décider	Techniques d'interaction	Technologies des contenus et services	Sécurité, sûreté, résilience	
Supervision de notre environnement	B	B		B	A	A				
Intégration des énergies renouvelables et efficacité énergétique	A		B	B	A	B			B	
Usine numérique du futur, fabrication agile et instrumentation		A		A	A	B	A			
Production de données massives dans tous les domaines de la santé	B	A	B	B	A	A	B	B	B	
Modélisation et simulation des écosystèmes complexes			B	B	A	A	B			
Mobilité et ville numériques		B	A	B	B	B	A	A	B	
Données, information et décision pour l'innovation sociale, éducative et culturelle			B				A	A	A	
Le numérique dans l'espace	A	B		A		B				
Cybersécurité et risques	B	A	A						A	

FIGURE 5.5.2 : Principaux verrous numériques des défis sociétaux relevés par l'atelier 7 (A : critique, B : essentiel).

L'analyse des verrous numériques des défis sociétaux a permis de lever deux priorités, le calcul des futurs, mais surtout la *mutation numérique*. Dans la Figure 5.4, j'ai placé certains verrous numériques levés par ce rapport avec en correspondance certains de mes travaux.

17. [http://cache.media.enseignementsup-recherche.gouv.fr/file/Strategie_ Recherche/28/1/Rapport_atelier_7_314281.pdf](http://cache.media.enseignementsup-recherche.gouv.fr/file/Strategie_Recherche/28/1/Rapport_atelier_7_314281.pdf)

Verrous	Recherche
Aide à la décision multi-acteurs	Valorisation des contenus numériques (Chaire d'excellence gestion du conflit)
Hétérogénéité sémantique	Valorisation des contenus numériques (Chaire d'excellence gestion du conflit)
Robots autonomes	Robots attentionnels
Interaction Homme-robot	Interprétation visuelle de geste pour l'interaction homme machine
Sémantique des contenus	Enrichissement <i>sémantique</i> des documents

TABLE 5.4 : Adéquation verrous / recherche.

Je souhaite ici reproduire un passage de ce rapport qui me paraît particulièrement intéressant.

Les liens interdisciplinaires à créer ou renforcer couvrent tout le spectre des disciplines du numérique avec les disciplines des sciences humaines et sociales, celles des sciences exactes ou de l'ingénierie. L'ouverture interdisciplinaire comporte deux volets, (i) la mise en place effective de structures et de moyens et (ii) la facilitation thématique des modèles interdisciplinaires et de convergence.

Les lieux interdisciplinaires pour la créativité dans tous les projets scientifiques et techniques sont les conditions nécessaires mais non suffisantes pour qu'il en sorte de réelles innovations. Cet effort doit impliquer non seulement les acteurs de la chaîne de l'innovation, notamment les pôles de compétitivité, mais aussi les acteurs de la recherche académique et de l'enseignement supérieur, à l'instar de ce qui se fait dans des centres de recherche comme le Culture Lab de l'Université de Newcastle, le Media Lab du MIT et le Human-Computer Interaction Institute de CMU.

Une de mes ambitions seraient d'accompagner la création d'une telle structure sur le territoire Rochelais, voire transrégional.

Une stratégie transrégionale de recherche¹⁸

L'enseignement supérieur et la recherche fonctionnent aujourd'hui dans un système mondialisé très concurrentiel et contribuent de manière déterminante au rayonnement ainsi qu'au développement des territoires. Ainsi, dans le contexte national de regroupement des universités et la constitution de pôles régionaux

18. Compte tenu de la volatilité des décisions politique concernant les nouvelles régions, il se peut que cette section soit caduque au moment où vous lirez ces lignes.

d'enseignement supérieur et de recherche d'envergure, les 5 universités de La Rochelle, Limoges, Orléans, Poitiers et Tours ont décidé de se réunir au sein d'une COMUE (Communautés d'universités et d'établissements) Centre, Limousin et Poitou-Charentes. Cette solution apporte une « masse critique » suffisante en nombre d'étudiants (environ 85 000) et constitue un pôle d'enseignement supérieur et de recherche fort.

Ce regroupement permet :

- une offre de formation diversifiée, complémentaire et coordonnée à l'échelle interrégionale ;
- une véritable synergie des activités de recherche et de technologie nécessaires au développement économique et social des régions.

Cette synergie compte bénéficier de la création d'un ISITE (initiative sciences innovation territoire économie) lors de l'appel à projet PIA2¹⁹. L'ISITE est limité à quelques thématiques de reconnaissance internationale avec un ancrage territorial et des relations partenariales avec le monde socioéconomiques. Un de ses axes seraient *l'usage du numérique et la dématérialisation pour le patrimoine, l'administration et l'éducation*. L'aspect patrimonial et administratif de l'ISITE est en adéquation parfaite avec mon projet de recherche.

5.5.3 Vision à long terme

Le 18 avril 2013, le Président de la République a mis en place la *Commission Innovation 2030*²⁰. L'objectif de cette commission est de s'approprier les principaux enjeux du monde de 2030 et d'identifier un nombre limité d'opportunités majeures au potentiel particulièrement fort pour l'économie française. A l'issue de ces travaux, sept *ambitions* ont vu le jour :

- le stockage de l'énergie,
- le recyclage des matières : métaux rares,
- la valorisation des richesses marines : métaux et dessalement de l'eau de mer,
- les protéines végétales et la chimie du végétal,
- la médecine individualisée,
- la silver économie, l'innovation au service de la longévité,
- la valorisation de données massives (Big Data).

C'est bien évidemment ce dernier point qui m'intéresse ici et rencontre la problématique sociétale à laquelle je m'intéresse.

19. <http://investissement-avenir.gouvernement.fr/>

20. <http://innovation-2030.dgcis.gouv.fr/fr/>

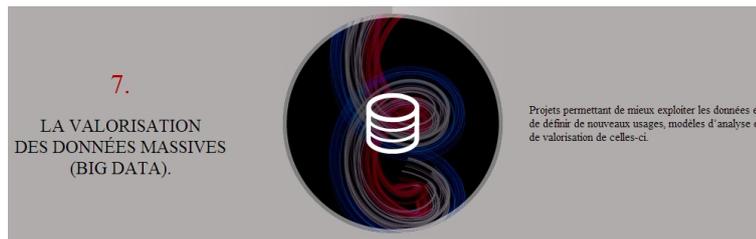


FIGURE 5.5.3 : Ambition « Valorisation Big Data » Commission *Innovation 2030*.

5.5.4 Transfert de technologie

Enfin, que ce soit au niveau régional, national ou européen, toutes les instances ou dispositifs militent pour que les résultats de la recherche soient valorisés et profitent aux entreprises et à la société. Dans ce contexte, les laboratoires vont devoir s'adapter à ces nouvelles exigences de nos tutelles afin de concilier Recherche Fondamentale et Recherche Appliquée.

L'Université de La Rochelle, et plus particulièrement le laboratoire L3i, a déjà mis en place une telle démarche depuis plusieurs années. Je suis personnellement impliqué dans deux collaborations.

Itesoft

Cet acteur mondial de la dématérialisation de documents et des processus propose une offre logicielle permettant d'accélérer, de fluidifier et de sécuriser les processus métier des entreprises et des administrations.

Nous travaillons avec cette société spécialisée dans la dématérialisation de documents administratifs. Nos recherches ont pour but de développer de nouveaux outils que l'entreprise pourra intégrer dans ses solutions. Je co-encadre par exemple la thèse d'Élodie Carel dont le sujet est *l'extraction des différentes couches couleurs à partir du document*. Nous souhaitons exploiter ces couches dans le processus traditionnel tout en améliorant la segmentation et en y ajoutant de l'information sémantique sur la couleur [Carel 13].



Valconum

L'objectif de Valconum (Centre Européen de Valorisation des Contenus) est de permettre la création d'innovations et le développement de l'économie de la dématérialisation et de la valorisation des contenus numériques. Le projet repose sur la collaboration entre une vaste communauté scientifique européenne reconnue au niveau mondial et des entreprises privées qui font de le R&D un facteur clé de leur développement, de leur compétitivité et de leur réussite. Ce centre regroupe des laboratoires de recherche et des entreprises. Il a pour objectif d'aider les entreprises à résoudre leurs problèmes de dématérialisation et de valorisation des contenus. Ce thème de valorisation des contenus est d'ailleurs le nouvel axe stratégique du laboratoire.



Valconum a organisé LA SEMAINE DU NUMÉRIQUE ET DE L'INNOVATION²¹, et m'a invité pour intervenir lors de la journée *Cultures Numériques – «Nouveaux enjeux au croisement des disciplines»*; Au cours de cette journée, je suis intervenu lors de la table ronde consacrée à la dématérialisation des supports.

21. <http://sni2014.valconum.fr/>

Points clés

Positionnement

- Définition de la problématique
- Intégration de la problématique dans une vision matricielle

Contributions

- Vision intégrée multi-niveaux de l'émergence de l'information
- Proposition d'une approche de la complexité selon trois niveaux

Conclusion

L'information est un concept dont la portée n'a d'égale que la difficulté qu'il y a à l'évoquer en termes généraux.

Emmanuel Dion [Dion 97]

A travers ce manuscrit, j'espère avoir démontré suffisamment d'autonomie, de maturité, de vision stratégique et de capacité à l'encadrement.

J'ai proposé, dans une première partie, les travaux que j'ai réalisés et dans une seconde ceux récemment initiés. J'ai proposé une double vision matricielle permettant une meilleure compréhension de mes travaux. J'ai déroulé en fil rouge la thématique de la recherche de l'information.

Mais finalement à quel *besoin* cette recherche répond ?

De façon classique, *la pyramide des besoins* propose une représentation pyramidale de la hiérarchie des besoins, une théorie de la motivation élaborée à partir des observations réalisées dans les années 1940 par le psychologue Abraham Maslow (cf. Figure 5.5.4). Elle postule que tant qu'un besoin n'est pas satisfait il constitue une source de motivation. A partir du moment où il est satisfait, c'est le besoin de niveau supérieur qui apparaîtra comme une nouvelle source de motivation.

Mais quel est donc ce *besoin d'information* ? Ce besoin qui a même sa journée mondiale le 17 mai. Le besoin d'information est, à la différence des besoins naturels, de nature sociale. Il est cependant considéré par certains comme un besoin fondamental puisque son absence place l'individu face à des problèmes ne lui permettant pas d'accéder à des niveaux supérieurs de la pyramide de Maslow. De plus, à la différence des besoins naturels pour lesquels la même réponse suffit à les satisfaire (l'eau pour la soif), ce besoin est extensif. L'information appelle l'information dans la mesure où une nouvelle information suscite de nouvelles questions auxquelles il faut apporter une réponse. La loi de *Moers* postule cependant que *la recherche d'information s'interrompt dès l'instant où il est plus pénible de rechercher cette information que de s'en priver*. Nous sommes donc en face d'un volume informationnel littéralement inimaginable et d'un besoin fondamental pour notre espèce et pour les robots [Oudeyer 07].

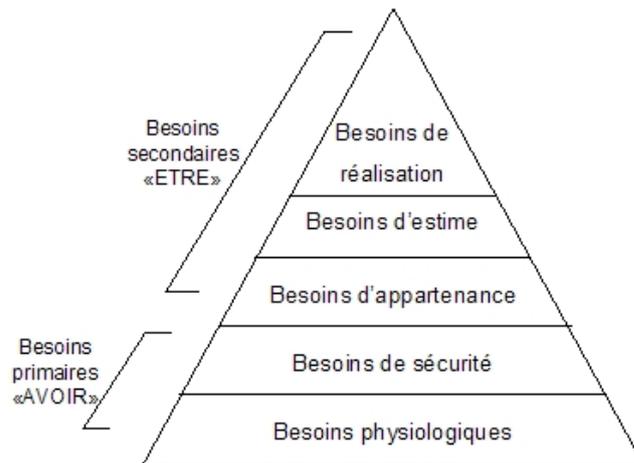


FIGURE 5.5.4 : La pyramide des besoins de Maslow (1954)

Pyramide de Maslow 2.0

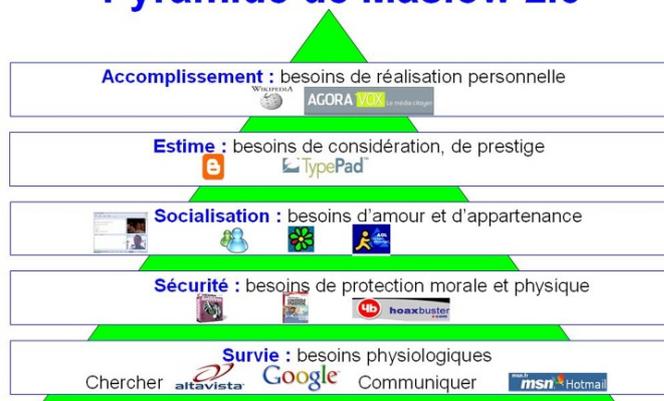


FIGURE 5.5.5 : La pyramide des besoins de Maslow 2.0

C'est dans la conciliation de ces deux situations que se trouve la source de ma recherche. De façon plus moderne, il existe *la pyramide de Maslow 2.0*²² (cf figure 5.5.5).

22. <http://blog.cozic.fr/la-pyramide-des-besoins-de-maslow-illustree-20>

Et pendant ce temps à l'Université de l'Invisible ...

Il se passait des tas de choses à l'Université de l'Invisible et, malheureusement, l'enseignement en faisait forcément partie. La faculté s'était depuis longtemps attaquée à la question et avait mis au point diverses techniques pour y échapper. Mais ce n'était pas un problème parce que, pour être juste, les étudiants avaient agi de même.

Le système fonctionnait plutôt bien et, comme il arrive souvent dans ces cas là, avait pris valeur de tradition. Les cours avaient manifestement lieu puisqu'ils figuraient noir sur blanc dans l'emploi du temps. Le fait que personne n'y assistait était un détail secondaire. On soutenait à l'occasion que les cours n'existaient pas, mais personne ne s'y rendait jamais pour le vérifier. De toute façon, avait on avancé (en l'occurrence le lecteur de pensée embrouillée), les cours avaient lieu en essence, tout était donc pour le mieux.*

En matière d'instruction, l'Université pratiquait donc l'ancestrale méthode qui consistait à mettre des tas de jeunes gens dans le voisinage d'un tas de livres en espérant que quelque chose passerait des uns aux autres, tandis que les jeunes gens se mettaient dans le voisinage d'auberges et de tavernes pour exactement les mêmes raisons. On était en milieu d'après-midi. Le titulaire de la chaire des études indéfinies donnait un cours dans la salle 3B, donc sa présence endormie devant le feu de la salle Peu Commune était un détail technique sur lequel le tact interdisait d'émettre le moindre commentaire.

Ridculle lui flanqua un coup de pieds dans les tibias.

*"Ouille! -Pardon de vous interrompre, titulaire, fit Ridculle pour la forme. Que les dieux me viennent en aide, faut que je réunisse le conseil des mages. Où sont-ils tous?" Le titulaire des études indéfinies se massa la jambe. "Je sais que l'assistant des runes modernes donne un cours dans la salle 3B**", dit-il. Mais je ne sais pas où il est réellement..."*

**Ce qui est comme la logique floue, mais en moins bien.*

***Tous les cours virtuels avaient lieu dans la salle 3B, une salle non seulement introuvable sur les plans de l'Université mais aussi, pensait-on, aux dimensions infinies.*

Terry Pratchett, *les annales du disque monde.*

Bibliographie

- [Ajallooeian 09] M. Ajallooeian, A. Borji, B.N. Araabi, M.N. Ahmadabadi & H. Moradi. *Fast Hand gesture recognition based on saliency maps : An application to interactive robotic marionette playing*. In The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009, pages 841–847, September 2009.
- [Allen 89] P M Allen & M J Lesser. *Evolution : Travelling in an Imaginary Landscape*. In Friedhelm Mündemann Jörg D. Becker, Ignaz Eisele, editeur, Parallelism, Learning, Evolution, Workshop on Evolutionary Models and Strategies, pages 420–444, Bedford, U.K., 1989. International Ecolotechnology Research Centre, Springer-Verlag.
- [Amini 13] Massih-Reza Amini, eric Gaussier & Gregoire Pean. *Recherche d’information : applications, modeles et algorithmes*. Eyrolles, Paris, 2013.
- [Awad 12] Dounia Awad, Vincent Courboulay & Arnaud Revel. *Saliency Filtering of SIFT Detectors : Application to CBIR*. In Advanced Concepts for Intelligent Vision Systems, Brno, Tchèque, République, September 2012.
- [Awad 14] D Awad, V. Courboulay & A. Revel. *A new hybrid texture-perceptual descriptor : application CBIR*. In 22st International Conference on Pattern Recognition - ICPR, page to be published, August 2014.
- [Baroni 07] Raphael Baroni. *La tension narrative : suspense, curiosite et surprise*. Editions du Seuil, Paris, 2007.
- [Bertalanffy 02] Ludwig von Bertalanffy. *Theorie generale des systemes*. Dunod, 2e edition edition, September 2002.
- [Berthoz 09] Alain Berthoz. *La simplicité ?* O. Jacob, Paris, 2009.

- [Boulding 56] Kenneth E. Boulding. *General Systems Theory—The Skeleton of Science*. Management Science, vol. 2, no. 3, pages 197–208, April 1956. [www](#)
- [Breazeal 04] Cynthia L. Breazeal. *Designing sociable robots*. MIT Press, 2004.
- [Bruce 09] N D B Bruce & J K Tsotsos. *Saliency, attention, and visual search : An information theoretic approach*. Journal of Vision, vol. 9, no. 3, page 5, 2009.
- [Buttcher 10] Stefan Buttcher, Charles L. A Clarke & Gordon V Cormack. *Information retrieval : implementing and evaluating search engines*. MIT Press, Cambridge, Mass., 2010.
- [Campbell 68] F W Campbell & J G Robson. *Application of Fourier analysis to the visibility of gratings*. J. Physiol. (Lond.), vol. 197, no. 3, pages 551–566, August 1968.
- [Carel 13] Élodie Carel, Vincent Courboulay, Jean-Christophe Burie & Jean-Marc Ogier. *Dominant Color Segmentation of Administrative Document Images by Hierarchical Clustering*. In 13th ACM Symposium on Document Engineering (DocEng), page to be published, Florence, Italie, August 2013. [www](#)
- [Courboulay 02] Vincent Courboulay. *Une nouvelle approche variationnelle du traitement d’images : application à la coopération détection-reconstruction*. La Rochelle, January 2002. [www](#)
- [Courboulay 14] V Courboulay. *Recherche de l’information dans les images RGB-D Premier pas vers une modélisation de dynamique attentionnelle 3D*. GDR ISIS-Analyse conjointe des images RGB-Depth, Février 2014.
- [Coustaty 10a] Mickaël Coustaty, Vincent Courboulay & Jean-Marc Ogier. *Analyzing Old Documents Using a Complex Approach : Application to Lettrines Indexing*. In EGC (best of volume), pages 155–171, 2010.
- [Coustaty 10b] Mickaël Coustaty, Nguyen Giap, Vincent Courboulay & Jean-Marc Ogier. *Approche complexe de l’analyse de documents anciens*. In EGC’2010 Extraction et gestion des connaissances, Tunisie, 2010.
- [Coustaty 12] Mickael Coustaty, Vincent Courboulay & Jean-Marc Ogier. *Analyzing Old Documents Using a Complex Approach : Application to Lettrines Indexing*. In Fabrice

- Guillet, Gilbert Ritschard & Djamel Abdelkader Zighed, éditeurs, *Advances in Knowledge Discovery and Management*, numéro 398 in *Studies in Computational Intelligence*, pages 155–171. Springer Berlin Heidelberg, January 2012. [www](#)
- [Coustaty 13] Mickaël Coustaty, Benjamin Duthil, Vincent Courboulay & Jean-Marc Ogier. *Adaptative Saliency for Logo Spotting*. In Home Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, page to be published, États-Unis, August 2013. [www](#)
- [Csikszentmih 97] Mihaly Csikszentmih. *Creativity flow and the psychology of discovery and invention*. Harper & Row, London, 1997.
- [Dautenhahn 02] K. Dautenhahn & A. Billard. *Games children with autism can play with robots, a humanoid robotic doll*. In in In Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology, page 179–190. Springer-Verlag, 2002.
- [Dave 12] Akshat Dave, Rachit Dubey & Bernard Ghanem. *Do humans fixate on interest points ?* In ICPR, pages 2784–2787, 2012.
- [Descartes 00] Rene Descartes. *Discours de la methode*. GF Flammarion, Paris, 2000.
- [Dion 97] Emmanuel Dion. *Invitation à la théorie de l'information*. Seuil, Paris, March 1997.
- [Duthil 11] Benjamin Duthil, François Troussel, Mathieu Roche, Gérard Dray, Michel Plantié, Jacky Montmain & Pascal Poncelet. *Towards an automatic characterization of criteria, DEXA '11*. In Proceedings of the 22nd International Conference on Database and Expert Systems Applications DEXA 2011, page 457, 2011.
- [Duthil 12a] Benjamin Duthil. *De l'extraction des connaissances à la recommandation*. PhD thesis, Université de Montpellier II, France, 2012.
- [Duthil 12b] Benjamin Duthil, François Troussel, Gérard Dray, Pascal Poncelet & Jacky Montmain. *Extraction d'opinions appliquée à des critères*. In Yves Lechevallier, Guy Melançon & Bruno Pinaud, éditeurs, EGC, volume RNTI-E-23 of *Revue des Nouvelles Technologies de l'Information*, pages 483–488. Hermann-Éditions, 2012. [www](#)

- [Duthil 13a] Benjamin Duthil, Mickaël Coustaty, Vincent Courboulay & Jean-Marc Ogier. *Visual Saliency and Terminology Extraction for Document Annotation*. In 13th ACM Symposium on Document Engineering (DocEng), page to be published, Italie, September 2013. [www](#)
- [Duthil 13b] Benjamin Duthil, Mickaël Coustaty, Vincent Courboulay & Jean-Marc Ogier. *Visual saliency and terminology extraction for document classification*. In Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, page to be published, États-Unis, August 2013. [www](#)
- [Duthil 14] B. Duthil, V. Courboulay, Coustaty M. & Ogier j M. *MAD : une plateforme mobile pour l'annotation de document vers la classification (CORIA)*. In CONference en Recherche d'Information et Applications 2014 (CORIA), 2014.
- [Everingham 10] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn & Andrew Zisserman. *The Pascal Visual Object Classes (VOC) Challenge*. Int J Comput Vis, vol. 88, no. 2, pages 303–338, June 2010. [www](#)
- [Fabri 08] Marc Fabri & Ruth Aylett. *Animating expressive characters for social interaction*. John Benjamins Publishing, 2008.
- [Fath 03] B D Fath, H Cabezas & C W Pawlowski. *Regime changes in ecological systems : an information theory approach*. J. Theor. Biol., vol. 222, pages 517–530, 2003.
- [Fath 04] B Fath. *Exergy and Fisher Information as ecological indices*. Ecological Modelling, vol. 174, no. 1-2, pages 25–35, 2004. [www](#)
- [Fox 07] Michael D Fox, Abraham Z Snyder, Justin L Vincent & Marcus E Raichle. *Intrinsic Fluctuations within Cortical Systems Account for Intertrial Variability in Human Behavior*. Neuron, vol. 56, no. 1, pages 171–184, 2007.
- [Frieden 98] B. Roy Frieden. *Physics from Fisher information : a unification*. 1998.
- [Frieden 04] B. Roy Frieden. *Science from Fisher Information : A Unification*. Cambridge University Press, cambridge edition, 2004. [www](#)
- [Gbehounou 12] Syntyche Gbehounou, François Lecellier & Christine Fernandez-Maloigne. *Extraction of emotional impact in*

- colour images*. In 6th European Conference on Colour in Graphics, Imaging, and Vision, pages 314–319, Pays-Bas, May 2012. [www](#)
- [Gbehounou 13] Syntyche Gbehounou, François Lecellier, Christine Fernandez-Maloigne & Vincent Courboulay. *Can salient interest regions resume emotional impact of an image?* In 15th international conference on computer analysis of images and patterns, page to be published, York, Royaume-Uni, August 2013. [www](#)
- [Grynszpan 07] O. Grynszpan, J.-C. Martin & J. Nadel. *What influences human computer interaction in autism?* In IEEE 6th International Conference on Development and Learning, 2007. ICDL 2007, pages 53–58, July 2007.
- [Guillaume 37] Paul Guillaume. *La Psychologie de la forme*. Flammarion, Paris, 1937.
- [Itti 98] Laurent Itti, Christof Koch, E Niebur & Others. *A model of saliency-based visual attention for rapid scene analysis*. IEEE Transactions on pattern analysis and machine intelligence, vol. 20, no. 11, pages 1254–1259, 1998.
- [Itti 00] Laurent Itti. *Models of bottom-up and top-down visual attention*. phd, California Institute of Technology, 2000. [www](#)
- [Itti 01] Laurent Itti & Christof Koch. *Computational modelling of visual attention*. Nature Reviews Neuroscience, vol. 2, no. 3, pages 194–204, 2001.
- [James 50] William James. *The principles of psychology*. Dover-Publ., New York, NY, 1950.
- [Koch 85] C Koch & S Ullman. *Shifts in selective visual attention : towards the underlying neural circuitry*. Hum Neurobiology, vol. 4, no. 4, pages 219–227, 1985.
- [Koch 06] Kristin Koch, Judith McLean, Ronen Segev, Michael A. Freed, Michael J. Berry, Vijay Balasubramanian & Peter Sterling. *How Much the Eye Tells the Brain*. Curr Biol, vol. 16, no. 14, pages 1428–1434, July 2006. PMID : 16860742 PMCID : PMC1564115. [www](#)
- [Le Meur 06a] O Le Meur, Xavier Castellan, Patrick Le Callet & Dominique Barba. *Efficient saliency-based repurposing method*. In IEEE International Conference on Image Processing, pages 421–424, Atlanta, USA, 2006.

- [Le Meur 06b] O Le Meur, Patrick Le Callet, Dominique Barba & D Thoreau. *A coherent computational approach to model bottom-up visual attention*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pages 802–817, 2006.
- [Lesser 98] Mike Lesser & Murray Dinah. *Mind as a dynamical system : Implications for autism*. In In Psychobiology of autism : current research & practice, 1998.
- [Lloyd 01] S Lloyd. *Measures of complexity : a nonexhaustive list*. IEEE Control Systems Magazine, vol. 21, no. 4, pages 7–8, 2001. [www](#)
- [Millane 94] R.P. Millane. *Analytic properties of the Hartley transform and their implications*. Proceedings of the IEEE, vol. 82, no. 3, pages 413–428, 1994.
- [Mohamed 06] A. Ould Mohamed, V. Courboulay, K. Sehaba & M. Menard. *Attention Analysis in Interactive Software for Children with Autism*. In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, Assets '06, page 133–140, New York, NY, USA, 2006. ACM. [www](#)
- [Moigne 77] Jean-Louis Le Moigne. *La theorie du systeme general : Theorie de la modelisation*. Presses universitaires de France, 1977.
- [Morin 81] Edgar Morin. *La methode t. 1, la nature de la nature*. Ed. du Seuil, Paris, 1981.
- [Morin 05] Edgar Morin. *Introduction a la pensee complexe*. Seuil, Paris, 2005.
- [Murray 03] J D Murray. *Mathematical Biology : Spatial models and biomedical applications*. Springer Verlag, 2003.
- [Murray 05] Dinah Murray, Mike Lesser & Wendy Lawson. *Attention , monotropism and the diagnostic criteria for. Autism*, vol. 9, no. 2, pages 139–156, 2005.
- [Nguyen 12] Van-Toi Nguyen, Thi-Lan Le, Thi-Thanh-Hai Tran, Rémy Mullot & Vincent Courboulay. *A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier*. In Proceedings of The Fourth International Conference on Communications and Electronics (ICCE 2012), pages 608–613, Hue Royal, Viet Nam, August 2012. [www](#)

- [Oudeyer 07] Pierre-Yves Oudeyer & Frederic Kaplan. *What is Intrinsic Motivation? A Typology of Computational Approaches*. Front Neurobotics, vol. 1, November 2007. [www](#)
- [Ould Mohamed 07a] Abdalahi Ould Mohamed, Matthieu Perreira Da Silva, Vincent Courboulay & Michel Ménard. *Intégration d'un modèle d'inattention dans une plateforme de rééducation adaptative pour enfants autistes*. In ASSISTH'2007 : Pour une meilleure insertion dans la société, Toulouse, France, 2007. [www](#)
- [Ould Mohamed 07b] Abdallah Ould Mohamed, Matthieu Perreira Da Silva, Vincent Courboulay & Michel Menard. *Intégration d'un modèle d'inattention dans une plateforme de rééducation adaptative pour enfants autistes*. In ASSISTH'2007 : Pour une meilleure insertion dans la société, France, 2007. [www](#)
- [Perreira Da Silva 07] Matthieu Perreira Da Silva, Abdalahi Ould Mohamed & Vincent Courboulay. *Gaze tracking, attention and interactive applications*. Rapport technique, Laboratoire Informatique Image Interaction, La Rochelle, France, 2007. [www](#)
- [Perreira Da Silva 08] Matthieu Perreira Da Silva, Vincent Courboulay, Armelle Prigent & Pascal Estrailier. *Real-time face tracking for attention aware adaptive games*. In Computer Vision Systems, volume 5008 of *Lecture Notes in Computer Science*, pages 99–108, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. [www](#)
- [Perreira Da Silva 09] Matthieu Perreira Da Silva, Vincent Courboulay, Armelle Prigent & Pascal Estrailier. *Fast, low resource, head detection and tracking for interactive applications*. PsychNology Journal, vol. 7, no. 3, pages 243–264, 2009. [www](#)
- [Perreira da Silva 10a] Matthieu Perreira da Silva. *Modèle computationnel d'attention pour la vision adaptative*. La Rochelle, December 2010. [www](#)
- [Perreira Da Silva 10b] Matthieu Perreira Da Silva, V Courboulay, A Prigent & P Estrailier. *Evaluation of preys / predators systems for visual attention simulation*. In VISAPP 2010 - International Conference on Computer Vision Theory and Applications, pages 275–282, Angers, 2010. INSTICC. [www](#)
- [Perreira Da Silva 11a] Matthieu Perreira Da Silva, Vincent Courboulay & Pascal Estrailier. *Image complexity measure based on visual*

- attention*. In Ieee International Conference On Image Processing, pages 3342–3345, Brussels, 2011.
- [Perreira Da Silva 11b] Matthieu Perreira Da Silva, Vincent Courboulay & Pascal Estraillier. *Objective validation of a dynamical and plausible computational model of visual attention*. In 3rd European workshop on Visual Information Processing (EU-VIP), 2011.
- [Perreira Da Silva 12a] Matthieu Perreira Da Silva & Vincent Courboulay. *Implementation and evaluation of a computational model of attention for computer vision*. In Developing and Applying Biologically-Inspired Vision Systems : Interdisciplinary Concepts, pages 273–306. Hershey, Pennsylvania : IGI Global., August 2012. [www](#)
- [Perreira Da Silva 12b] Matthieu Perreira Da Silva, Vincent Courboulay & Patrick Le Callet. *Real Time Dynamic Image Re-targeting Based on a Dynamic Visual Attention Model*. In Proceedings of 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pages pp.653–658, Melbourne, Australie, July 2012. [www](#)
- [Picard 08] David Picard, Matthieu Cord & Arnaud Revel. *Image retrieval over networks : active learning using ant algorithm*. Multimedia, IEEE Transactions on, vol. 10, no. 7, pages 1356–1365, 2008.
- [Posner 78] Posner, Nissen M. & W M., Ogden. *Attended and unattended processing modes : the role of set for spatial location*, pages 137–157. Laurence Erlbaum, Hillsdale, New Jersey, 1978.
- [Rempulski 09] Nicolas Rempulski, Armelle Prigent, Pascal Estraillier, Vincent Courboulay & Matthieu Perreira Da Silva. *Adaptive Storytelling Based On Model-Checking Approaches*. International Journal of Intelligent Games and Simulation (IJIGS), vol. 5, no. 2, pages 33—41, 2009. [www](#)
- [Rempulski 13] Nicolas Rempulski. *Synthese dynamique de superviseur pour l'execution adaptative d'applications interactives*. PhD thesis, Universite de La Rochelle, 2013.
- [Sehaba 06] Karim Sehaba, Vincent Courboulay & Pascal Estraillier. *Observation and analysis of behaviour of autistic children using an interactive system*. Technology and Disability, vol. 18, no. 4, pages 181–188, January 2006. [www](#)

- [Thériault 13] Christian Thériault, Nicolas Thome & Matthieu Cord. *Extended coding and pooling in the HMAX model*. IEEE Transactions on Image Processing, vol. 22, no. 2, pages 764–777, 2013.
- [Thom 88] Rene Thom. *Esquisse d’une semiophysique*. InterEditions, Paris, 1988.
- [Thom 09] Noel emile Chenciner Alain Thom Rene. *Predire n’est pas expliquer*. Flammarion, Paris, 2009.
- [Treisman 80] Anne Treisman & Garry Gelade. *A Feature-Integration Theory of Attention*. Cognitive Psychology, vol. 136, no. 12, pages 97–136, 1980.
- [Unser 86] M Unser. *Local Linear Transforms for Texture Measurements*. Signal Process., vol. 11, no. 1, page 61–79, July 1986. [www](#)
- [Viola 02] Paul Viola & Michael Jones. *Robust real-time object detection*. International Journal of Computer Vision, vol. 57, no. 2, pages 137–154, 2002.
- [Volterra 28] Vito Volterra. *Variations and fluctuations of the number of individuals in animal species living together*. ICES Journal of Marine Science, vol. 3, no. 1, pages 3–51, 1928.
- [Yarbus 67] Alfred L Yarbus. *Eye Movements and Vision*. Plenum Press, new York, 1967.

CV détaillé

« "Un CV doit être comme la robe d'une jolie femme. Assez long pour couvrir l'essentiel, assez court pour rester intéressant" mais ... "Un Cv c'est comme une pierre tombale, je n'en ai jamais vu qui dénigre celui qui est dessous" »"

(Pierre Dac)

> Identité

Prénom Nom	Vincent Courboulay	
Statut	Maitre de Conférences	
Fonction	Responsable du Coursus Mastère en Ingénierie Informatique	
Correspondant C2i		
Numéro de Téléphone	+33 5 46 45 83 71	
Courrier électronique	vincent.courboulay@univ-lr.fr	
Page Web Personnelle	http://perso.univ-lr.fr/vcourbou	
Date arrivée dans le laboratoire	Septembre 1999	
Date de titularisation	Septembre 2004	

> Thèmes de recherche

Mes activités scientifiques s'organisent autour de l'extraction d'informations dans les images et les séquences d'images et cela sur des niveaux distincts, présentant chacun une définition particulière autour de l'information. En d'autres termes, comment sont reliées images et information ? Comment extraire l'information contenue dans une image ? Comment acquérir de l'information grâce à l'image ?

J'ai orienté ma recherche dans une démarche cohérente me permettant d'avoir une vision large spectre de la recherche de l'information

Je m'intéresse aussi à l'attention et sa modélisation biologiquement inspirée.

- Extraction de l'information dans les documents
- Modélisation de l'attention
- Utilisation de la saillance dans l'analyse de l'attention
- Interactivité logicielle par analyse du comportement

Des collaborations nationales et internationales attestées par des articles en cours de publication avec, entre autres, l'ircyn (Institut de Recherche en Communications et Cybernétique de Nantes), l'institut de recherche international MICA du Viêt Nam et la Faculté Polytechnique de Mons (FPMs).

Ces collaborations sont également attestées par des séjours de recherche en Belgique et une invitation de 5 mois d'une chercheuse Vietnamiennne autour du document.

J'ai encadré ou encadre 6 thèses, plus un nombre équivalent de masters :

J'ai également été sollicité pour être membre de plusieurs jurys (Poitiers, Grenoble) ainsi que du comité de suivi de thèse (ircyn).

Je suis membre du comité éditorial de la revue *EAI Endorsed Transactions on Creative Technologies*.

Je suis élu au conseil de laboratoire du L3I depuis 2010.

> Points forts de mes activités de recherche

Théorisation de l'incertitude imprécision Modélisation de l'attention Utilisation de la saillance dans l'analyse de l'attention Interactivité logicielle par analyse du comportement Extraction de l'information dans les documents

> Rayonnement scientifique national et international

- Elu au conseil de laboratoire du L3I depuis 2010.
- Expert-évaluateur auprès du FNRS (Fonds de la Recherche Scientifique Belge).
- Expert auprès du conseil régional d'Aquitaine.
- Membre du comité éditorial de la revue *EAI Endorsed Transactions on Creative Technologies*.
- Membre du comité de programme de la conférence BICT (*International Conference on Bio-inspired Information and Communications Technologies*).
- Organisateur de la journée transverse GDR ISIS et GDR VISION *La saillance visuelle : de son exploitation à son évaluation – Round 2* (19/6/14).
- Correspondant de la chaire d'excellence « *gestion du conflit et de l'après conflit* » de Limoges.

Collaboration nationale (attestée par des conférences, revues et des séjours de recherche)

- IRCYN Nantes
- SIC-XLIM Poitiers
- LISA Angers
- Gipsa-Lab Grenoble
- Université de Limoges (Correspondant de la chaire d'excellence « *gestion du conflit et de l'après conflit* »)

Collaboration internationale (attestée par des conférences, revues et des séjours de recherche)

- MICA Viet Nam
- Université de Mons, Numediart Research Institute Belgique.

Echanges scientifiques

- 2014 : Séjour d'un mois en Belgique (Faculté Polytechnique de Mons (FPMs))
- Réception d'une chercheuse invitée du Viet Nam autour de la problématique de détection de logos
- Réception d'une chercheuse grenobloise autour de l'analyse de trajectoires de regard

Relecteur auprès de revues majeures

- Computer Vision and Image Understanding, Elsevier
- Signal Processing: Image Communication, Elsevier
- The Visual computer, Springer.
- Journal on Image and Video Processing, EURASIP
- Journal of Electronic Imaging, SPIE
- Journal of Multimedia Processing and Technologies, IAPR-ICDAR (International Conference on Document Analysis and Recognition), IASTED-Assistive Technology (ancien membre du comité scientifique), IEEE-CGAMES (Conference on Computer Games)
- ...

Membre des GDR STIC-Santé et ISIS

> Coopération industrielles et valorisation :

- **Représentant pour l'Université de La Rochelle de la chaire d'excellence « *gestion du conflit et de l'après conflit* »**
- **Responsable d'une tâche de l'ANR VISIIR Visual Seek for Interactive Image Retrieval (2014-2017)**
- **Dépôt d'un logiciel à l'APP - Agence pour la Protection des Programmes**
- Responsable du contrat *FT-AutiSTIC (Fondation Orange)*
- Membre de l'ANR *RobAutiSTIC*
- Membre de l'ERT *Interactivités Numériques*

> Encadrements scientifiques :

Thèses

- Axel Jean-Caurant (Septembre 2014)

Sujet : Analyse haut niveau de données hétérogènes massives : de l'enrichissement vers la conceptualisation. Applications à la valorisation d'archives pénales

- Elodie Carel depuis Janvier 2011

Sujet : Segmentation d'images couleurs de documents

- Dounia Awad depuis 2010

Sujet : Conception d'un système d'attention visuelle pour l'analyse de comportements dans le cadre d'activités interactives

- Nguyễn Văn Tới depuis 2010. *Co-Encadrement avec le MICA d'Hanoi au Vietnam*

Sujet : Interprétation visuelle des gestes pour l'interaction homme-machine

- Encadrant scientifique de la thèse de Nicolas Rempulski 2008-2012,

Sujet : Synthèse dynamique de superviseur pour la scénarisation adaptative

- Matthieu Ferreira Da Silva 2007-2010.

Sujet : «Étude du comportement d'un utilisateur et de son environnement dans une application interactive»

Post-doc

- Patrice Denis 2014-2015

Sujet : «Nouveaux descripteurs d'images basés sur une approche bio inspirée». Financement ANR

Masters et apprentis (depuis 2010)

- Capture de justificatifs pour aide à la saisie de contrats d'assurance sur table Microsoft Surface **Logica** Nantes **Anas BOUTADGHART** Master 2 – ICONE 2011 (6mois)
- Intégration de la Kinect dans un système d'attention visuelle L3I La Rochelle **Geoffrey Lacombe** Master 1 – ICONE 2012 (3mois)
- Réalisation d'un portail intranet et étude d'un outil de gestion électronique de documents (GED) EADS Sogerma **Frédéric Coutantin** Apprenti ICONE 2010/2012
- Réalisation d'un portail intranet, numérisation et gestion d'archives à la mairie de Rochefort. Thibault Maurice. Apprenti ICONE 2014/2015 (http://www.numerique.culture.fr/pub-fr/document.html?base=dcollection&id=FR-DC-AC017299_caa17b93)

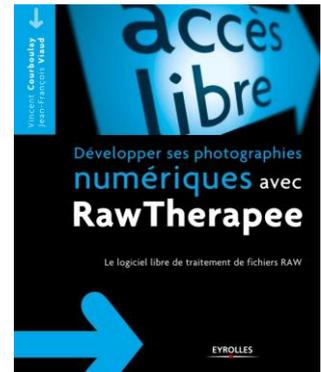
> Administratif :

- Représentant officiel du L3I et de l'Université de La Rochelle dans la chaire d'excellence « Gestion du conflit et de l'après conflit » portée par Limoges.
- Représentant de l'Université dans l'incubateur de Charente Maritime
- Responsable du Cursus Mastère en Ingénierie parcours Informatique depuis 2013 (<http://www.reseau-figure.fr/>).
- Directeur du service d'informatique transversale 2007-2013 (2 personnels, 30 intervenants, 2000 étudiants)
- Membre élu du conseil de laboratoire depuis 2010
- Membre du conseil de département depuis 2007
- Membre du CARI depuis 2007

> Diffusion scientifique

- Rédaction d'un livre autour de l'image et la photographie numérique. Les premiers chapitres sont une introduction à la couleur et aux images numériques. (V.Courboulay, J.F.Viaud, *RawTherapee, une chambre claire et ouverte*. Collection : *Accès libre* - avril 2014).

- Intervention auprès du Muséum d'Histoire Naturelle de La Rochelle. Présentation de mes travaux avec pour illustration une des pièces du muséum montrée à titre exceptionnel (décembre 2014).
- Organisation d'une rencontre autour de la robotique à des enfants d'une école maternelle et élémentaire du département (Intervention avec Mr A.Revel et Nao) (Mai 2014)

**> Activités d'enseignement :**

Mes enseignements de spécialité s'articulent autour de différents axes :

- Enseignements web (PHP, HTML/CSS, AJAX, CMS...)
- Eco-Informatique (informatique et développement durable, Certification C2I métiers de l'environnement et de l'aménagement durable)
- Programmation et traitement d'images (vision industrielle, traitement d'images, algorithmique et prog...)

J'accorde un soin particulier à la diffusion scientifique au travers de formation comme l'informatique et développement durable, méthodologie universitaire et informatique et nouvelles technologies, mais aussi au travers des formations GRETA et du personnel de l'Université.

De par ma fonction de directeur de l'informatique transversale entre 2005 et 2014 (structure gérant plus de 2000 étudiants/an et l'ensemble des certifications C2I) j'ai été responsable de la formation C2I métiers de l'environnement et j'interviens dans l'ensemble des autres C2I couverts à l'Université de La Rochelle.

Enfin, depuis Octobre 2013, je suis responsable du Coursus Mastère en Ingénierie Informatique.

Le CMI Informatique est une formation licence, master, cohérente sur 5 années débouchant sur des fonctions d'ingénieur. C'est une formation universitaire progressive et exigeante bénéficiant d'un label national sur le modèle international du "master of engineering".

> Principales publications (4 dernières années)**Publications dans des revues internationales avec comité de lectures**

- V.Courboulay, M.Mancas, *CuriousMind photographer: inciting robots to find out remarkable scenes*, EAI Endorsed Transactions on Creative Technologies, *to be published*.
- Perreira Da Silva M., Courboulay V., Prigent A., Estraillier P. *Fast, low resource, head detection and tracking for interactive applications*. PsychNology Journal 7, 3 (2009) 243-264
- N. Rempulski, A. Prigent, P. Estraillier, V. Courboulay and M.Perreira Da Silva, *Adaptive Storytelling based on model-checking approaches*, IJIGS - International Journal of Intelligent Games and Simulation, 2009, Volume 5, Number 2, 33-42.
- K. Sehaba, V. Courboulay, and P. Estraillier. *Interactive system by observation and analysis of behavior for children with autism*. Technology and Disability Journal, 2006 Volume 18, Number 4 pp 181 – 188.
- **Publications dans des revues nationales avec comité de lectures**
- Perreira Da Silva M., Courboulay V. Implémentation et évaluation d'un modèle d'attention pour la vision adaptative, *Traitement du Signal* vol 28, 6 (2012) p611-641
- Histace, C. Cavaro-Ménard, V. Courboulay, M. Ménard, Traitement et analyse quantitative de séquences IRM cardiaques marquées, *Traitement du Signal* 2006 vol. 23, n°2 p. 125-143

Livre

V.Courboulay, J.F.Viaud, *RawTherapee, une chambre claire et ouverte*. Collection : Accès libre - avril 2014. <http://izibook.eyrolles.com/produit/3149/9782212260175>

Chapitre de livre

- Mickaël Coustaty, Vincent Courboulay, Jean-Marc Ogier, “Analyzing Old Documents Using a Complex Approach: Application to Lettrines Indexing “ in *Advances in Knowledge Discovery and Management. Volume 2 Series: Studies in Computational Intelligence*, (2012) 155-172
- M. Perreira Da Silva, V. Courboulay “Implementation and evaluation of a computational model of attention for computer vision” In *Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts*. IGI Global (2012) pp 273-306
- D.Awad, V.Courboulay, A.Revel “Attentive Content Based Image Retrieval” In *From Human Attention to Computational Attention: A Multidisciplinary Approach* Springer *To be published*

Publications dans des conférences internationales avec comité de lectures depuis 2010

D.Awad, V.Courboulay, A.Revel. *A new hybrid texture-perceptual descriptor: application CBIR*, Paper accepted for the ICPR 2014.

B. Duthil, V. Courboulay, Coustaty M. & Ogier j M. *MAD : une plateforme mobile pour l'annotation de document vers la classification*. In CORIA 2014.

- Duthil B., Coustaty M., Courboulay V., Ogier J.-M. *Visual Saliency and Terminology Extraction for Document Annotation* Dans 13th ACM Symposium on Document Engineering (DocEng) - 13th ACM Symposium on Document Engineering (DocEng), Italie (2013)
- Carel E., Courboulay V., Burie J.-C., Ogier J.-M. *Dominant Color Segmentation of Administrative Document Images by Hierarchical Clustering* Dans 13th ACM Symposium on Document Engineering (DocEng) - 13th ACM Symposium on Document Engineering (DocEng), Italie (2013)
- Gbehounou S., Lecellier F., Fernandez-Maloigne C., Courboulay V. *Can salient interest regions resume emotional impact of an image?* Dans 15th international conference on computer analysis of images and patterns - CAIP 2013 - 15th international conference on computer analysis of images and patterns, Royaume-Uni (2013)
- Coustaty M., Duthil B., Courboulay V., Ogier J.-M. *Adaptative Saliency for Logo Spotting* Dans Home Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, États-Unis (2013)
- Duthil B., Coustaty M., Courboulay V., Ogier J.-M. *Visual saliency and terminology extraction for document classification* Dans Tenth IAPR International Workshop on Graphics Recognition - GREC 2013, États-Unis (2013)
- D.Awad, V.Courboulay, and A. Revel. *Saliency Filtering of SIFT Detectors: Application to CBIR*.In: ACIVS, Vol. 7517Springer (2012) , p. 290-300.
- NGuyen Van Toi, Le Thi Lan, Tran Thi Thanh Hai, Mullot Remy, Courboulay, Vincent A *method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier*. The 4th International Conference on Communications and Electronics (ICCE 2012) Hue, Viet Nam - august 2012
- Perreira Da Silva, M., Courboulay, V., & Le Callet, P. (2012). Real Time Dynamic Image Re-targeting Based on a Dynamic Visual Attention Model. In 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (pp. pp.653–658).
- Courboulay, V., Perreira Da Silva, M., & Courboulay, V. (2012). Real-time computational attention model for dynamic scenes analysis: from implementation to evaluation. In SPIE (Ed.), SPIE Optics, Photonics and Digital Technologies for Multimedia Applications - Visual Attention
- M. Perreira Da Silva, V. Courboulay, P. Estrailier, «Image complexity mesure based on visual attention », ICIP 2011.
- M. Perreira Da Silva, V. Courboulay, « Une nouvelle mesure de complexité pour les images basée sur l'attention visuelle », GRETSI 2011.
- M. Perreira Da Silva, V. Courboulay, P. Estrailier, « Objective validation of a dynamical and plausible computational model of attention », EUVIP 2011.
- M. Perreira Da Silva, V. Courboulay, A. Prigent, P. Estrailier, « Evaluation of preys / predators systems for visual attention simulation », VISAPP 2010.